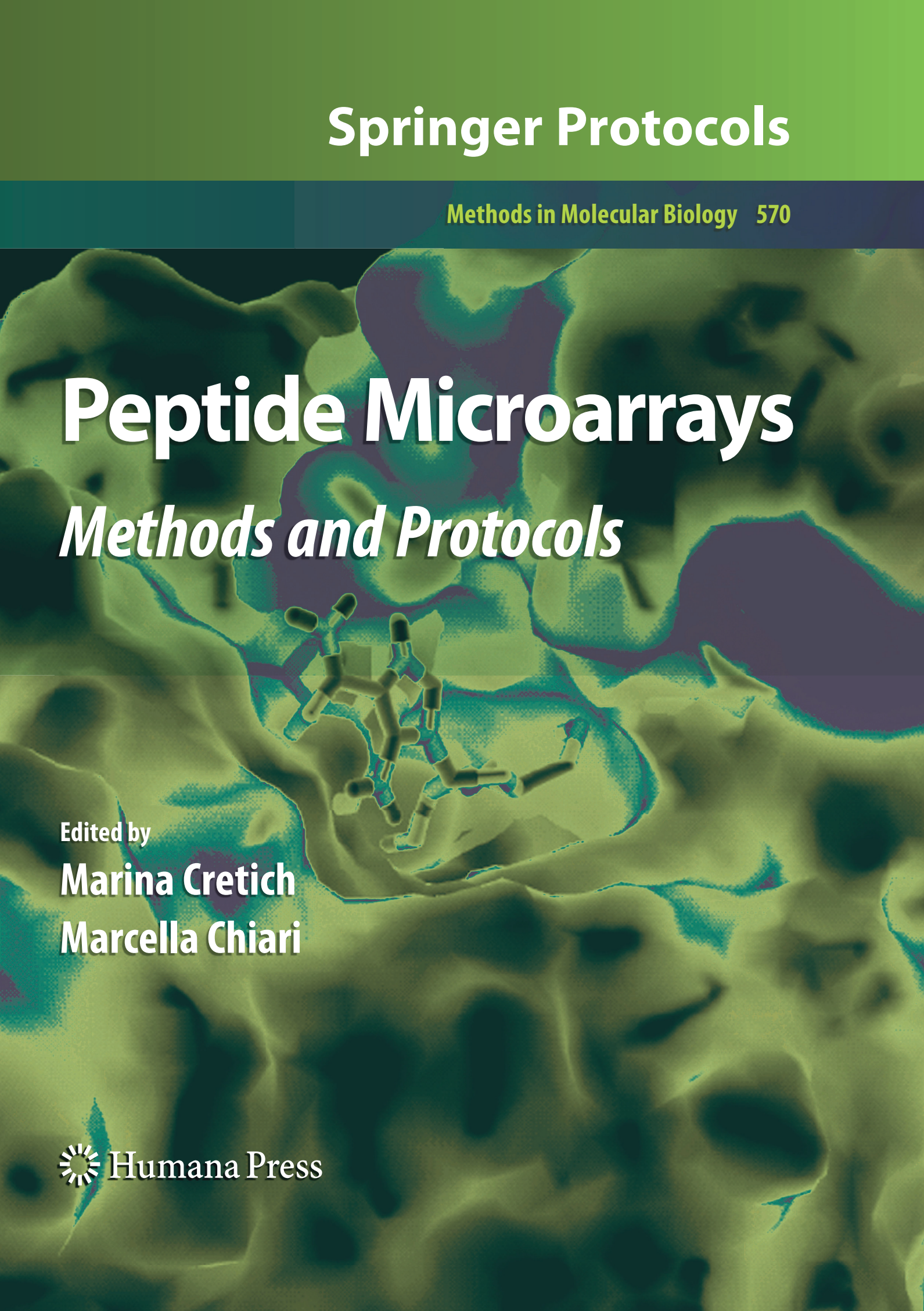
Springer Protocols

Methods in Molecular Biology 570

Peptide Microarrays

Methods and Protocols



Edited by
Marina Cretich
Marcella Chiari

Humana Press

METHODS IN MOLECULAR BIOLOGY™

Series Editor
John M. Walker
School of Life Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For other titles published in this series, go to
www.springer.com/series/7651

Peptide Microarrays

Methods and Protocols

Edited by

Marina Cretich

ICRM-C.N.R., Milano, Italy

Marcella Chiari

ICRM-C.N.R., Milano, Italy

Humana Press

Editors
Marina Cretich
CNR
Inst. Chimica del Riconoscimento
Molecolare (ICRM)
Via Mario Bianco, 9
20131 Milano
Italy
marina.cretich@icrm.cnr.it

Marcella Chiari
CNR
Inst. Chimica del Riconoscimento
Molecolare (ICRM)
Via Mario Bianco, 9
20131 Milano
Italy
marcella.chiari@icrm.cnr.it

Series Editors
John M. Walker
University of Hertfordshire
Hatfield, Herts
UK

ISSN 1064-3745 e-ISSN 1940-6029
ISBN 978-1-60327-393-0 e-ISBN 978-1-60327-394-7
DOI 10.1007/978-1-60327-394-7
Springer Dordrecht Heidelburg London New York

Library of Congress Control Number: 2009931197

Preface

The ability to directly interrogate protein interactions in a high-throughput format provides an unprecedented opportunity to dissect the complex molecular architectures of living systems. Traditional molecular biology techniques provide valuable information on the expression, structure, and function of proteins; nonetheless, these methods are unable to provide the massively parallel analysis capacity which is essential to map an entire proteome or to accomplish the present-day drug discovery programs. Parallel sensing using arrayed systems has proved to be successful in genomic research where DNA microarrays are widely used for large-scale analysis of gene expression. However, the protein equivalent of the DNA microarrays poses a more difficult challenge, especially in the identification of suitable high-affinity capture ligands, which retain their specificity and functionality following immobilization on the arrayed sensor substrate. Synthetic peptides have some very interesting features as capture ligands in microarray experiments: they are easy to synthesize and manipulate, highly stable, and inexpensive. More importantly, since peptide ligands can be modeled to act as a binding site for almost any target structure of the proteome, they can mimic biological activities of proteins and provide a straightforward analytical approach in a variety of applications. Beyond their initial utility in protease profiling, researchers are adopting peptide microarrays for the comparative screening of many different classes of enzymes and proteins for the study of complex biological matrices and even living cells. The peptides on the arrays may serve to sense protein activity (as substrates) or act as small molecule ligands (for potential therapeutic leads) in profiling, detection or diagnostic applications. Due to these developments, along with the diminishing costs of library synthesis and the growing of commercial support, peptide microarrays will no longer remain just a research tool but also a versatile and powerful platform to be harnessed for wider drug discovery and point-of-care applications.

Peptide Microarrays: Methods and Protocols is a view on the peptide array technology, on its applications and technical issues. The book is divided into four sections:

i) The **Introduction** comprises four review chapters aimed at giving to the readers the broadest view of how peptides can characterize proteins and clarify, at the amino acid level, the molecular recognition events in which they are involved.

ii) **Section II** is dedicated to several practical applications of peptide arrays: to their production and use for enzyme and binding motifs characterization, epitope mapping and diagnostics.

iii) **Section III** is devoted to new technological advancements that can improve the state-of-the-art of peptide microarray methodologies and even pose the way toward new concepts in parallel analysis.

iv) **Section IV** is dedicated to software and Web tools for the design of peptide arrays and for the analysis of output data.

We hope that readers will take advantage of the unique insights and the novel solutions in peptide arrays technology offered by each chapter of this book.

We thank all the authors of this Volume for their valuable contributions and for their willingness in sharing their experience and knowledge. We thank Prof. John Walker for continuous support and editorial help and the staff of Humana Press for their professionalism.

Finally, we wish to express gratitude to the members of our group and to all the colleagues and friends who, despite the difficulties, stay beside us.

Marina Cretich
Marcella Chiari

Contents

Contributors

Costin Apostol • *CIB, Génopole de Lille, Universités de Lille 1 et 2, LIFL, Bâtiment M3, Cité scientifique, Villeneuve d'Ascq cedex, France*

Bernard Aÿ • *Institut für Medizinische Immunologie, Charité – Universitätsmedizin Berlin, Berlin, Germany*

Mario Beyer • *Department of Chip-Based Peptide Libraries, German Cancer Research Center, Heidelberg, Germany*

Ralf F. Bischoff • *Department of Chip-Based Peptide Libraries, German Cancer Research Center, Heidelberg, Germany*

Ines Block • *Department of Chip-Based Peptide Libraries, German Cancer Research Center, Heidelberg, Germany*

Alex Bossers • *Central Veterinary Institute of Wageningen UR, Lelystad, The Netherlands*

Carine Brakha • *Inserm, U823, IAPC, Grenoble, France; University J. Fourier, Grenoble, France*

Ole Brandt • *Centre for Microbial Diseases and Immunity Research, University of British Columbia, Vancouver, BC, Canada*

Frank Breitling • *Department of Chip-Based Peptide Libraries, German Cancer Research Center, Heidelberg, Germany*

Douglas J. Briant • *Centre for Systems Biology, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON, Canada*

Roberta Carbone • *Tethis s.r.l., Milano, Italy*

Gang Chen • *Department of Neuroscience, University of Minnesota, Minneapolis, MN, USA*

Marcella Chiari • *Istituto di Chimica del Riconoscimento Molecolare (ICRM) – C.N.R., Milano, Italy*

Giorgio Colombo • *Istituto di Chimica del Riconoscimento Molecolare (ICRM) – C.N.R., Milano, Italy*

Sandra Cortès • *Inserm, U823, IAPC, Grenoble, France; University J. Fourier, Grenoble, France*

Angelo Corti • *Immunobiotechnology Unit, Department of Biological and Technological Research, San Raffaele H Scientific Institute, Milano, Italy*

Marina Cretich • *Istituto di Chimica del Riconoscimento Molecolare (ICRM) – C.N.R. Milano, Italy*

Sara D'Angelo • *Department of Medical Sciences and Interdisciplinary Research Center for Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara, Italy*

Francesco Damin • *Istituto di Chimica del Riconoscimento Molecolare (ICRM) – C.N.R., Milano, Italy*

Francois Debaene • *Institut de Science et Ingénierie Supramoléculaires Université Louis Pasteur – CNRS UMR 70068, Strasbourg, France*

Rémi Desmet • *UMR8161, CNRS, Universités de Lille 1 et 2, Institut Pasteur de Lille, IFR 142 Institut de Biologie de Lille, Lille, France*
Clarisse Dhaenens • *CIB, Génopole de Lille, Universités de Lille 1 et 2, LIFL, Bâtiment M3, Cité scientifique, Villeneuve d'Ascq cedex, France*
Gabriele Di Carlo • *Istituto di Chimica del Riconoscimento Molecolare (ICRM) – C.N.R., Milano, Italy*
Roberto Di Niro • *Department of Life Sciences, University of Trieste, Trieste, Italy*
Gael Even • *CIB, Génopole de Lille, Universités de Lille 1 et 2,LIFL, Bâtiment M3, Cité scientifique, Villeneuve d'Ascq cedex, France*
Vronique Fafeur • *UMR 8161, CNRS, Universités de Lille 1 et 2, Institut Pasteur de Lille, IFR 142 Institut de Biologie de Lille, Lille, France*
Thomas Felgenhauer • *Department of Chip-Based Peptide Libraries, German Cancer Research Center, Heidelberg, Germany*
Simon Fernandez • *Department of Chip-Based Peptide Libraries, German Cancer Research Center, Heidelberg, Germany*
Gunter Fischer • *Max Planck Research Unit for Enzymology of Protein Folding, Halle, Germany*
Xiaolian Gao • *Department of Biology and Biochemistry, University of Houston, Houston, TX, USA*
Robert E.W. Hancock • *Centre for Microbial Diseases and Immunity Research, University of British Columbia, Vancouver, BC, Canada*
Kai Hilpert • *Centre for Microbial Diseases and Immunity Research, University of British Columbia, Vancouver, BC, Canada*
Ailing Hong • *Atactic Technologies, Houston, TX, USA*
Kai König • *Department of Chip-Based Peptide Libraries, German Cancer Research Center, Heidelberg, Germany*
Nadezhda V. Kulagina • *US Naval Research Laboratory, Washington, DC, USA*
Jan P.M. Langeveld • *Central Veterinary Institute of Wageningen UR, Lelystad, The Netherlands.*
Klaus Leibe • *Department of Chip-Based Peptide Libraries, German Cancer Research Center, Heidelberg, Germany*
Genie C. Leung • *Therapeutics Group, The Campbell Family Institute for Breast Cancer Research, Toronto, ON, Canada*
Shawn S.-C. Li • *Department of Biochemistry and the Siebens–Drake Medical Research Institute, Schulich School of Medicine and Dentistry, University of Western Ontario, London, ON, Canada*
Tongbin Li • *Department of Neuroscience, University of Minnesota, Minneapolis, MN, USA*
Thierry Livache • *University J. Fourier, Grenoble, France; UMR SPrAM 5819 (CEA, CNRS), INAC, Grenoble, France*
Renato Longhi • *Istituto di Chimica del Riconoscimento Molecolare (ICRM) – C.N.R., Milano, Italy*
Patrice Marche • *Inserm, U823, IAPC, Grenoble, France; University J. Fourier, Grenoble, France*
Roberto Marzari • *Department of Life Sciences, University of Trieste, Trieste, Italy*

Massimiliano Meli • *Istituto di Chimica del Riconoscimento Molecolare (ICRM) – C.N.R., Milano, Italy*
Oleg Melnyk • *UMR CNRS 8161 Institut de Biologie de Lille 1, Lille Cedex, France*
Hisakazu Mihara • *Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Midori, Yokohama, Japan*
James M. Murphy • *Division of Molecular Medicine, Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia*
Alexander Nesterov • *Department of Chip-Based Peptide Libraries, German Cancer Research Center, Heidelberg, Germany*
Stella H. North • *US Naval Research Laboratory, Washington, DC, USA*
Fabien Pamelard • *CIB, Génopole de Lille, Universités de Lille 1 et 2, LIFL, Bâtiment M3, Cité scientifique, Villeneuve d'Ascq cedex, France*
Cristian Preda • *CIB, Génopole de Lille, Universités de Lille 1 et 2, LIFL, Bâtiment M3, Cité scientifique, Villeneuve d'Ascq cedex, France*
Jan Priem • *Central Veterinary Institute of Wageningen UR, Lelystad, The Netherlands*
Marie Reilly • *Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden*
Alan Rigter • *Central Veterinary Institute of Wageningen UR, Lelystad, The Netherlands*
André Roget • *University J. Fourier, Grenoble, France; UMR SPrAM 5819 (CEA, CNRS), INAC, Grenoble, France*
Clément Roux • *UMR CNRS 8161 Institut de Biologie de Lille 1, Lille Cedex, France*
Claudio Santoro • *Department of Medical Sciences and Interdisciplinary Research Center for Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara, Italy*
Daniele Sblattero • *Department of Medical Sciences and Interdisciplinary Research Center for Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara, Italy*
Christopher Schirwitz • *Department of Chip-based Peptide Libraries, German Cancer Research Center, Heidelberg, Germany*
Mike Schutkowski • *JPT Peptide Technologies GmbH, Berlin, Germany*
Paola Secco • *Department of Medical Sciences and Interdisciplinary Research Center for Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara, Italy*
Valentina Sedini • *Istituto di Chimica del Riconoscimento Molecolare (ICRM) – C.N.R., Milano, Italy*
Frank Sicheri • *Centre for Systems Biology, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON, Canada*
Vianney Souplet • *UMR CNRS 8161 Institut de Biologie de Lille 1, Lille Cedex, France*
Volker Stadler • *Department of Chip-based Peptide Libraries, German Cancer Research Center, Heidelberg, Germany*
Chris R. Taitt • *US Naval Research Laboratory, Washington, DC, USA*
Victor E. Tapia • *Institut für Medizinische Immunologie, Charité-Universitätsmedizin Berlin, Berlin, Germany*
Alexandra Thiele • *Max Planck Research Unit for Enzymology of Protein Folding, Halle, Germany*

Drophatie Timmers-Parohi • *Pepscan Systems BV, Lelystad, The Netherlands*
Kin-ya Tomizaki • *Innovative Materials and Processing Research Center and Department of Materials Chemistry, Ryukoku University, Seta, Otsu 520–2194, Japan*
Kenji Usui • *Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Midori, Yokohama, Japan*
Davide Valentini • *Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden*
Marie-Bernadette Villiers • *Inserm, U823, IAPC, Grenoble, France; University J. Fourier, Grenoble, France*
Rudolph Volkmer • *Institut für Medizinische Immunologie, Charité-Universitätsmedizin Berlin, Berlin, Germany*
Zhenxin Wang • *State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, China*
Matthias Weiwad • *Max Planck Research Unit for Enzymology of Protein Folding, Halle, Germany*
Dirk F.H. Winkler • *Centre for Microbial Diseases and Immunity Research, University of British Columbia, Vancouver, BC, Canada*
Nicolas Winssinger • *Institut de Science et Ingénierie Supramoléculaires Université Louis Pasteur – CNRS UMR 70068, Strasbourg, France*
Chenggang Wu • *Department of Biochemistry and the Siebens-Drake Medical Research Institute, Schulich School of Medicine and Dentistry, University of Western Ontario, London, ON, Canada*
Johannes Zerweck • *JPT Peptide Technologies GmbH, Berlin, Germany*
Xiaochuan Zhou • *LC Sciences, Houston, TX, USA*
Qi Zhu • *Department of Biology and Biochemistry, University of Houston, Houston, TX, USA*
Zhixiang Zuo • *Department of Neuroscience, University of Minnesota, Minneapolis, MN, USA*

Section I

Introduction

Chapter 1

Exploring and Profiling Protein Function with Peptide Arrays

Victor E. Tapia, Bernhard Ay, and Rudolf Volkmer

Abstract

Development of array technologies started in the late 1980s and was first extensively applied to DNA arrays especially in the genomic field. Today this technique has become a powerful tool for high-throughput approaches in biology and chemistry. Progresses were mainly driven by the human genome project and were associated with the development of several new technologies, which led to the onset of additional "omic" topics like proteomics, glycomics, antibodyomics or lipidomics. The main characteristics of the array technology are (i) spatially addressable immobilization of a huge number of different capture molecules; (ii) probing the array in a simultaneous and highly parallel manner with a biological sample; (iii) tendency towards miniaturization of the arrays; and (iv) software-supported read-out and data analysis. We review some general concepts about peptide arrays on planar supports and point out technical aspects concerning the generation of peptide microarrays. Finally, we discuss recent applications by describing relevant literature.

Key words: Peptide microarray, high-throughput screening, kinase activity, fluorescence-based detection, native chemical ligation, immobilization, binding assay.

1. Introduction

Peptide arrays are high-throughput devices for binding assays. The array format refers to the spatially addressable presentation of peptide probes immobilized to discrete areas of a support surface. This principle has been extensively and fruitfully applied to RNA hybridization experiments using cDNA microarray devices (1–3). Development in this area has led to the establishment of advanced technical aspects with respect to immobilization chemistries, industrialization of dedicated equipment and tailor-made bioinformatics. Current aims to globally analyse cellular behaviour relying in cDNA capture molecules include mutation and polymorphism analysis, determination of clinically predictive genes, as well as relating expression profiles with cellular states.

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_1,

However, distribution of cellular labour between DNA and protein structures may render measurements of genetic change insufficient to explain emergent phenotypes (4) involved in normal or pathological cellular behaviour. Situations in which this point cannot be overseen are the intricate balance between promiscuity and selectivity of protein interactions as well as the regulation of enzyme activity. Here, assays at the protein level are required to point out gene function at the level of the phenotype. Moreover, in pragmatic terms, RNA levels do not correlate with protein abundance (5–7) and proteins, rather than genes, are the immediate causative agents of cellular function.

Several technologies aim to bridge this causative gap and are coined under the research field of proteomics. Rudely stated, their object of study is the proteome, originally defined as "the entire complement of proteins expressed by a genome, cell, tissue or organism" (8). Advanced protein science is characteristic of the field, and it is dedicated to satisfy the original expectations of functional genomics to globally analyse cellular organization. Just as genomics has reached an endpoint in gene discovery, namely the sequencing of the human genome, an equivalent ambition to discover the human proteome is borne by protein science (9, 10).

Protein arrays are predestined to support proteomic research. Due to their structural and functional complexities, proteins are highly informative systems and well suited to investigate protein interactions occurring over extensive complementary surfaces. However, several critical factors such as stability, native folding or activity of the immobilized proteins pose stringent demands on production, storage and experimental conditions. Due to that, it is not amazing that the first reports on protein arrays came just in time for the new millennium (11–13).

Peptides, on the contrary, are chemically quite resistant and have more modest dynamics to achieve their active conformations. In general, peptides retain partial aspects of protein function and, thus, peptide arrays are suitable to support proteomic research in a more reductionist approach. Not surprisingly, peptides paved the way for array technologies: in 1991 two techniques for the preparation of peptide arrays were published.

First, Frank et al. presented the SPOT technique – a stepwise synthesis of peptides on filter paper (14). Second, Fodor et al. reported the concept of light-directed spatially addressable chemical synthesis (15). Both these techniques are milestones in the advancement of the peptide array technology, but the former has found a wider spread use as preparative source of peptides for varied biological applications and for binding assays carried out in situ. We review some general concepts about peptide arrays on planar supports and point out technical aspects concerning the generation of peptide microarrays. Finally, we discuss recent applications by describing relevant literature but without the pretension to be extensive.

2. Advancement of Peptide Arrays

The technological advancement of the peptide arrays can roughly be reviewed by focusing (at least) on three milestones of stepwise peptide synthesis. The development of solid-phase peptide synthesis (SPPS) by Merrifield (16, 17) and adaptions of this procedure (18) set the chemical ground for innovative technologies to follow. The development of the "Pin" method by Geysen (19) introduces the array format to peptide synthesis. Definitive establishment of peptide arrays came along with the development of the SPOT synthesis by Frank (20, 21) which simplified chemical synthesis of peptide arrays to the addressable deposition of reagents on a cellulose sheet. Moreover, chemical synthesis allows incorporation of non-natural building blocks, preparation of branched and cyclic structures and labelling with chromophores. Modern peptide synthesis approaches and molecular biology make peptides accessible in a high degree of structural diversity. The two greatest drawbacks of synthetic peptide arrays are peptide length, with a quality threshold between 30 and 50 amino acids, as well as the restriction to linear motives, since the mimicry of nonlinear motives with linear peptide constructs is still under development (22).

Peptide arrays also present a milestone in the advancement of analytical binding assay systems. Since the 1990s, a major aspect of development to achieve the required sensitivities to analyse biological samples has been the miniaturization of analytical devices (23). It is important to note that miniaturization is not only a matter of high throughput and economy. Miniaturization is an essential factor that should provide saturation of binding sites under low analyte concentrations without significantly altering its bulk (or ambient) concentration upon capturing (24–27). In this sense, the first application of a peptide microarray device in 1991, anticipating even the application of cDNA arrays, achieved already the impressive feature density of about 1024 peptides in 1.6 cm^2 by means of in situ light-directed parallel synthesis (15). The several methods available to generate peptide arrays on planar solid surfaces offer a range between 16 peptides per cm^2, in the case of SPOT macroarrays (28, 29) and 2000–4000 peptides in 1.5 cm^2, in the case of microarrays generated by digital photolithography (30–33).

In order to support synthesis, planar materials have to fulfil several requirements including stability towards solvent and reagent deposition. The functional groups on the surface must also be biochemically accessible for chemical derivatization. Furthermore, upon solid-phase binding assay, generated peptides must be functionally displayed to

allow molecular recognition with a binding partner in the solution phase. In particular, non-specific interactions should be ruled out. Flexible porous supports such as cellulose (34, 35), cotton (36, 37) or membranes (38–40) are preferentially used for peptide array generation. Rigid, non-porous materials such as glass (41), gold films (42–44) or silicon (15, 33) have also been used for in situ synthesis but are much more technically demanding. On the other side, rigid materials have a number of advantages over porous supports for functional display of molecules. Impermeability and smooth two dimensionality of the material do not limit diffusion of the binding partner and lead to more accurate kinetics of recognition events. Finally, the flatness and transparency of glass improve image acquisition and simplifies the use of fluorescence dyes for the read-out process. In some cases, assembled 3D structure on a non-porous surface could be a fruitful approach. Several techniques for coherent surface modifications are described over the past 20 years in the literature. For a comparative overview on this field, we refer to articles dedicated to the peptide and protein array technologies (40, 45–52).

3. Preparation of Peptide Arrays

3.1. In Situ Peptide Synthesis

The SPOT technique consists of the stepwise synthesis of peptides on planar supports, such as amino-functionalized cellulose membranes or aminated polypropylene (20, 21, 28, 40, 53–58). The basic principle involves the positionally addressed delivery of small volumes of activated amino acid solutions directly onto a coherent membrane sheet. The areas wet by the resulting droplets can be considered as microreactors, provided a non-volatile solvent system is used. The functional groups fixed on the membrane surface react with the pipetted reagents as in a conventional solid-phase synthesis. The physical properties of the membrane surface, the solvent system and the applied volumes define the size of the resulting spots.

Peptide arrays prepared by the SPOT technique became popular tools for studying numerous aspects of molecular recognition. Recent developments such as new linker and cleavage strategies (59–62) as well as the fully automated synthesis achieved by the MultiPep SPOT-robot (INTAVIS Bioanalytical Instruments AG) have increased the spreading of the technique. The new linker and cleavage strategies, especially, open up the opportunity to use the SPOT technology as a highly parallel synthesis method for the preparation of a huge number of small amounts of soluble peptides (50–100 nmol/peptide). Although quantitative assays based

on this technology suffer from lack of precision and accuracy, achievements in the quantification of measurements (63, 64) and its application to analyse the affinity and specificity of PDZ protein binding domains have been reported (65, 66).

Other in situ technologies for peptide synthesis permit extreme miniaturization of peptide array formats. Addressable peptide arrays have been generated by using lithographic masks and photodeprotection (15, 67). The synthesis involves different masks for each amino acid type at each coupling step, light irradiation through the mask to remove photo-labile protecting groups and subsequent addition of the amino acid type required at the deprotected site. Although promising for high-density microarrays, the process has its limitations because production of photomasks is time consuming and expensive and requires a high-grade clean-room environment. Furthermore, the applied photochemistry is less efficient than conventional Boc and Fmoc chemistry. This results in a poor yield per step, reduces the quality of the surface-bound peptides and limits their sequence length. As a consequence, higher rates of false-positive and false-negative measurements are expected upon the screening assay. An interesting alternative is the use of a novel photogenerated reagent chemistry combined with standard Boc chemistry (31–33). It is based on an addressable maskless photogenerated acid formation using digital photolithography. The generated acid deprotects the t-Boc group of the growing peptide chain and releases the amino-terminus which is then amidated by an activated amino acid. Up to 4,000 features in a 1.5 cm^2 area have been synthesized with this method (31).

3.2. Immobilization of Pre-synthesized Peptides

Whenever sample economy, length of peptides and probe replicas are critical factors, immobilization of pre-synthesized peptides to a suitable display platform may be more provident. The solid-phase peptide synthesis (SPPS) and, specially, SPOT technology are viable preparative methods. The latter enables the parallel synthesis of up to 2,000 peptides in a very short time (1 week) and at amounts of approximately 50–100 nmol of cleaved material. Once cleaved from the preparative substrate, pre-synthesized peptides can be analysed and purified prior to immobilization. Thus, a quality control may be carried out at this key step of the manufacturing process.

A further advantage is the possibility to expand and redistribute the original peptide probes from the preparative substrate to numerous replicas in tailor-made array designs over one or more display platforms. The deposition of sub-nanolitre volumes of prepared peptides in printing buffers is assisted by contact or non-contact microdispensing robots. In general, contact methods work best for dispensing numerous different peptide probes to a small number of platforms, while non-contact methods work best

for dispensing to many platforms with feature replicas (68). Printing buffers may contribute essentially to coupling efficiency, spot morphology and the extent of array miniaturization. Buffer agents like sodium acetate and sodium chloride may enhance coupling efficiency, though crystallization may affect the functionality of attached peptides or signal-to-noise ratio upon fluorescence detection. The use of hygroscopic reagents, such as DMSO and glycerol, is also a conflictive affair: they reduce the surface tension of deposited peptide probes and, thus, generate more uniform feature morphology but may considerably compromise array quality in an ambient with even light changing air humidity. It has been recommended to consider the use of commercially available optimized printing buffers and to relegate printing devices to small rooms with reduced traffic (68). If buffers contain hygroscopic reagents, then the arrayer should be sealed off and equipped with a humidity controlling set at 60–70% relative humidity (69).

Several commercial products are available which offer immobilization platforms coated with sorptive films or silane modified with varied chemical functions. Protein immobilization to such platforms has profit from many basic principles established for DNA microarrays and has been extensively reviewed (10, 26, 27, 70–72). Perhaps with the exception of some methods for the functional immobilization of antibodies, the protocols described there are not applicable to the immobilization of peptides. Although helpful for allowing the adoption of enzyme-linked detection schemes, sorptive attachment fails to assure an oriented display of peptide probes. Due to the shorter length of peptides, the loss of residues essential for molecular recognition will be more hardly compensated as upon the sorptive attachment of proteins may be the case. Similarly, immobilization methods adapted from DNA microarrays and bioconjugation protocols may be cross-reactive to nucleophile side-chain residual functions essential for molecular recognition. These unspecific attachment mechanisms render a heterogeneous efficiency, which is strongly dependent on the amino acid sequence of peptide probes.

An additional problem with these immobilization strategies is that blocking is strongly recommended to avoid unspecific capturing of protein analyte to the surface. Standard blocking solutions, e.g. casein or BSA-based buffers, may compromise the accessibility of analyte to peptide probes. A solution to this point is to chemically quench reactive surface groups that remain after printing of peptides. Alternatively, MacBeath and Schreiber, first attached a BSA layer to a surface aldehyde function and then activated lysine, aspartate and glutamate residues on BSA with *N*,*N*′-disuccinimidyl carbonate to yield BSA-*N*-hydroxysuccinimide (BSA-NHS) (11). Subsequently, the NHS function is made to react with amines on printed peptides forming covalent urea or amide linkages. Remaining NHS functions were finally quenched with glycine. This approach allows an accessible

display of peptide probes on a protein monolayer; however, cross-reaction with side-chain nucleophiles remains a source of failed orientation as described above.

The oriented immobilization of unprotected peptide products faces a similar problem as the chemoselective ligation of peptide fragments to assemble proteins (73). Thus, many of these ligation techniques have been adapted for the immobilization of peptides to final display supports. Reactions leading to the formation of thioester (74, 75), thioether (76–79) and oxime have been specially instructive in early reports of chemoselective immobilization approaches (41, 80, 81). Additionally, the use of a Diels–Alders reaction (42, 82) and semicarbamide chemistry have been described. Also here, pioneer work from MacBeath et al. set up a protocol for immobilization of small molecules and peptides (83). The method relies on the Michael Addition. Glass slides were functionalized with maleimide groups that are expected to form a thioether linkage to thiol moieties contained in printed molecules. Unspecific binding of protein analyte to remaining surface maleimide functions was not observed. As for peptide probes, oriented covalent immobilization can only be expected when exclusively one (N-terminal) cysteine residue is present in the sequence. Another advanced approach is the derivatization of aminosilane coatings by coupling acetal-protected glyoxylic acid and deprotecting with dilute HCl to obtain a surface-bound glyoxylyl function with a free aldehyde (41). Immobilization can be achieved through formation of oxime bonds with N-terminal oxy-amine functions and thiazolidine rings with N-terminal cysteines. Quenching can be carried out with hydroxylamine derivates.

By adapting reactions originally described for the native chemical ligation (84, 85), Yao and co-workers have shown that derivatization of aminosilane slides to thioester functions allows the chemoselective reaction with 1,2-aminothiols of N-terminal cysteines (80, 81). The sulphydryl moiety of N-terminal cysteine-containing peptides is first captured by thioester bond formation and, subsequently, bond stability is achieved by spontaneous rearrangement leading to a native amide bond and regenerating the free sulphydryl on cysteine.

A complete different mechanism for oriented chemoselective immobilization applies affinity/avidity binding between biotinylated peptides and avidin-coated surfaces. The non-covalent interaction, with an approximate K_D of 10^{-15} M, is stronger than most maturated antibody/antigen reactions and very stable under normal assay conditions (ca. neutrally buffered solutions with low ionic strength). The avidin coat provides a protein layer preventing nonspecific binding to the glass surface, eliminates the need of blocking with casein or BSA-based buffers and, thus, releases peptide probes from eventual masking effects (80, 81, 86–88).

4. Applications of Peptide Arrays

The first instance of decision for array content often depends on which of two kinds of analysis should be carried out. The dedicated analytical devices for each case have been referred to as functional or analytical arrays (10). Although this distinction may help to plan an experiment or to link entries in databanks, the use of these concepts is very ambiguous. A functional assay analyses the chemical activity between sample and probes, i.e. molecular binding or substrate modifications. The sample is usually a purified analyte which is challenged with relevant peptide probes defined by some systematic collection of peptides, e.g. the targets of a consensus on a sequence data bank or frame and length versions of a protein sequence. Generally, previous knowledge of the assayed activities in terms of affinity or turnover is not required and densitometric measurements are reliable when qualitative.

Analytical assays, on the contrary, rely on quantitative densitometric measurements. In these assays, immobilized probes are usually chosen to have a (known) high affinity and specificity for the assayed samples. The analysed activity refers mostly to the concentration of one or multiple analytes in a biological sample like cell extracts, sera or CSF. Samples from different biological states may be compared on separate devices or by challenging one device with separately labelled samples. Depending on the number of different analytes assayed and the analyte–peptide specificities, the amount of different peptide features required to differentiate between biological states is not as extensive as in explorative functional assays. Preliminary work has shown the potential of peptide arrays for applications like protein expression level, affinity profiling and, eventually, clinical diagnostics. Nevertheless, especially in the later case, further development is required.

Functional peptide assays have found a wide range of applications: mapping antibody epitopes, protein domain-based molecular interactions, as well as investigation of enzyme–substrate–inhibitor interactions especially for kinases, proteases phosphatases, methylase and isomerases (20, 21, 28, 40, 53–58). The explorative approach demands an extensive representation of protein function and diversity. The sequence content of peptide arrays may be flexibly designed to screen the functions of protein analytes and collect the diversity of potential interaction partners. Three array designs may be considered as standards: combinatorial libraries, frame-shifted arrays and proteomic arrays.

Combinatorial libraries define each feature of a peptide array by a consensus sequence similar to $X_1X_2B_3B_4X_5X_6$ having one or more defined amino acid positions (symbolized B_n) and a number of randomized positions (symbolized X_n) (89). Only one amino

acid is introduced at the defined positions, while a mixture of amino acids is introduced at the randomized position. This results in a sub-library of different sequences in each single array feature. The B positions vary from feature to feature of the array. The drawback of this approach is the intensive effort required to deconvolute positive library features in order to unambiguously identify the epitope or substrate sequence.

The use of peptide arrays dedicated to identify linear epitopes of antibody samples is well reported (19, 55, 87, 90–93). This approach consists in the generation of peptide probes by shifting a short sequence frame (usually 8–15 amino acids in length) along the complete antigen sequence. In a similar fashion, the binding motives of a protein interaction domain may be identified by shifting short frames along the complete sequence of a validated interaction partner (94). This type of arrays is generally referred to as epitope or ligand mapping, respectively. Although this approach can only display continuous sequences, some of the functionality of discontinuous motives may be retained by continuous fragments of it. Recently, Tessier and Lindquist (95) synthesized large libraries of overlapping peptides derived from the sequence of the N-terminal and middle regions of the yeast Sup35 prion protein. Labelled Sup35 protein was captured by a very small set of peptides. These recognition elements selectively promoted the assembly of amyloid fibrils in a species-specific fashion.

Analogous to proteomic arrays, where the complete set of proteins coded by ORFs of the yeast or other organisms is immobilized, exploration of the epitope, ligand or substrate sequence space may be defined by targets of a consensus across annotated proteins in a sequence database (96–99). Thousands of peptide probes may be required to cover such space, eventually compromising the recommended use of probe repeats. In some cases, further restrictions may be required to display the concerned sequence diversity on one platform. Technological innovation, especially concerning miniaturization, is required to improve these limitations.

On the belief that advancing applications will rely on analytical approaches, we point out two selected applications which emphasize the quantitative potential of peptide array technologies. First, we price the systems biology approach to place proteins in a larger context of function, specially the construction of binary protein interaction networks. A quantitative protein interaction network of a human proteomic set of SH2 and PTB domains with phosphorylated peptides from ErbB receptor tyrosine kinases has been reported (100). The network is able to describe the selectivity of recruitment of proteins to the receptor at various affinity thresholds or in authors' words, "[the network] reveals [that] receptor tyrosine kinases [...] differ in the extent to which they become more promiscuous when overexpressed". This and an associated publication (101) show the importance of applying different

analyte concentrations to obtain saturation curves from which essential experimental parameters, such as estimates of the amount of active probes on the surface, can be calculated. The mentioned parameter is not biologically relevant, but still a determinant of signal intensity which must be taken into account in high-throughput quantitative approaches (101, 102).

Although this point is essential to allow reliable diagnostic and predictive applications, as well as to reveal the organization of cells and organisms at a systems level, to our best knowledge, only a few reports analyse the performance of peptide array-based assays to deliver accurate and precise measurements in high-throughput fashion (101–103). Nevertheless, several works have demonstrated that synthetic peptides are informative molecular probes to draw profiles of immunoreactivities (86, 90, 91, 93, 104–106). The basic thought behind these applications is that the immunological experience of an individual determines the propensity of its antibody repertoire for resistance or susceptibility to autoimmune and pathogen diseases (107, 108). Under this paradigm, Quintana et al. challenged the IgG antibody repertoire of cyclophosphamide-accelerated diabetes (CAD) mice, before and after CAD induction, with 266 different antigen probes including proteins, synthetic peptides, nucleotides and phospholipids (19 mice with bleeding samples 1 day before and 1 month after injection) (109). Serum samples from all mice were analysed by superparamagnetic clustering (SPC) as an inherent mechanism for robust and stable two-way clustering of reactivities. Of special interest is the result that the sets of most informative probes for diagnosis and prediction were very different, suggesting that informative probes may challenge IgG antibodies that are not direct causative agents of the analysed disease. Synthetic peptides derived from heat-shock proteins behaved specially well as probes – being the only probe type which was informative for both diagnosis and prediction.

Acknowledgments

Financial support from Detsche Froschungsgemeinschaft (SFB449) is gratefully acknowledged.

References

1. Bier, F. F., von Nickisch-Rosenegk, M., Ehrentreich-Forster, E., Reiss, E., Henkel, J., Strehlow, R., and Andresen, D. (2008) DNA microarrays, *Adv Biochem Eng Biotechnol* **109**, 433–53.
2. Southern, E. M. (2001) DNA microarrays. History and overview, *Methods Mol Biol* **170**, 1–15.
3. DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y.,

Su, Y. A., and Trent, J. M. (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer, *Nat Genet* **14**, 457–60.

4. Szathmary, E., and Smith, J. M. (1995) The major evolutionary transitions, *Nature* **374**, 227–32.
5. Anderson, L., and Seilhamer, J. (1997) A comparison of selected mRNA and protein abundances in human liver, *Electrophoresis* **18**, 533–7.
6. Chen, G., Gharib, T. G., Huang, C. C., Thomas, D. G., Shedden, K. A., Taylor, J. M., Kardia, S. L., Misek, D. E., Giordano, T. J., Iannettoni, M. D., Orringer, M. B., Hanash, S. M., and Beer, D. G. (2002) Proteomic analysis of lung adenocarcinoma: identification of a highly expressed set of proteins in tumors, *Clin Cancer Res* **8**, 2298–305.
7. Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast, *Mol Cell Biol* **19**, 1720–30.
8. Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J. C., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L., and Hochstrasser, D. F. (1996) From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis, *Biotechnology (NY)* **14**, 61–5.
9. Aebersold, R. (2003) Constellations in a cellular universe, *Nature* **422**, 115–16.
10. Phizicky, E., Bastiaens, P. I. H., Zhu, H., Snyder, M., and Fields, S. (2003) Protein analysis on a proteomic scale, *Nature* **422**, 208–15.
11. MacBeath, G., and Schreiber, S. L. (2000) Printing proteins as microarrays for high-throughput function determination, *Science* **289**, 1760–3.
12. Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R. A., Gerstein, M., and Snyder, M. (2001) Global analysis of protein activities using proteome chips, *Science* **293**, 2101–5.
13. Zhu, H., Klemic, J. F., Chang, S., Bertone, P., Casamayor, A., Klemic, K. G., Smith, D., Gerstein, M., Reed, M. A., and Snyder, M. (2000) Analysis of yeast protein kinases using protein chips, *Nat Genet* **26**, 283–9.
14. Frank, R., Güler, S., Krause, S., and Lindenmaier, W. (1991) Facile and rapid spot-synthesis of large numbers of peptides on membrane sheets., *in* "Peptides 1990, Proc 21st Eur Peptide Symp." (Giralt, E., and Andreu, D., Eds.), pp. 151–52, ESCOM Science Publishers B. V., Leiden.
15. Fodor, S. P., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T., and Solas, D. (1991) Light-directed, spatially addressable parallel chemical synthesis, *Science* **251**, 767–73.
16. Gutte, B., and Merrifield, R. B. (1969) The total synthesis of an enzyme with ribonuclease A activity, *J Am Chem Soc* **91**, 501–2.
17. Merrifield, R. B. (1965) Automated synthesis of peptides, *Science* **150**, 178–85.
18. Fields, G. B., and Noble, R. L. (1990) Solid phase peptide synthesis utilizing 9-fluorenylmethoxycarbonyl amino acids, *Int J Pept Protein Res* **35**, 161–214.
19. Geysen, H. M., Meloen, R. H., and Barteling, S. J. (1984) Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid, *Proc Natl Acad Sci USA* **81**, 3998–4002.
20. Frank, R. (1992) Spot-synthesis: an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support, *Tetrahedron*, 9217–32.
21. Frank, R. (2002) The SPOT-synthesis technique: synthetic peptide arrays on membrane supports–principles and applications, *J. Immunol. Methods* **267**, 13–26.
22. Goede, A., Jaeger, I. S., and Preissner, R. (2005) SUPERFICIAL–surface mapping of proteins via structure-based peptide library design, *BMC Bioinformatics* **6**, 223.
23. Ekins, R. P. (1998) Ligand assays: from electrophoresis to miniaturized microarrays, *Clin Chem* **44**, 2015–30.
24. Ekins, R., Chu, F., and Biggart, E. (1990) Multispot, multianalyte, immunoassay, *Ann Biol Clin (Paris)* **48**, 655–66.
25. Ekins, R. P. (1989) Multi-analyte immunoassay, *J Pharm Biomed Anal* **7**, 155–68.
26. Joos, T. O., Stoll, D., and Templin, M. F. (2002) Miniaturised multiplexed immunoassays, *Curr Opin Chem Biol* **6**, 76–80.
27. Templin, M. F., Stoll, D., Schrenk, M., Traub, P. C., Vohringer, C. F., and Joos, T. O. (2002) Protein microarray technology, *Drug Discov Today* **7**, 815–22.
28. Reimer, U., Reineke, U., and Schneider-Mergener, J. (2002) Peptide arrays: from macro to micro, *Curr Opin Biotechnol* **13**, 315–20.
29. Schutkowski, M., Reimer, U., Panse, S., Dong, L., Lizcano, J. M., Alessi, D. R., and Schneider-Mergener, J. (2004) High-

content peptide microarrays for deciphering kinase specificity and biology, *Angew Chem Int Ed Engl* **43**, 2671–4.

30. El Khoury, G., Laurenceau, E., Dugas, V., Chevolot, Y., Merieux, Y., Duclos, M. C., Souteyrand, E., Rigal, D., Wallach, J., and Cloarec, J. P. (2007) Acid deprotection of covalently immobilized peptide probes on glass slides for peptide microarrays, *Conf Proc IEEE Eng Med Biol Soc* **2007**, 2242–6.
31. Gao, X., Pellois, J. P., Na, Y., Kim, Y., Gulari, E., and Zhou, X. (2004) High density peptide microarrays. *In situ* synthesis and applications, *Mol Divers* **8**, 177–87.
32. Pellois, J. P., Wang, W., and Gao, X. (2000) Peptide synthesis based on t-Boc chemistry and solution photogenerated acids, *J Comb Chem* **2**, 355–60.
33. Pellois, J. P., Zhou, X., Srivannavit, O., Zhou, T., Gulari, E., and Gao, X. (2002) Individually addressable parallel peptide synthesis on microchips, *Nat Biotechnol* **20**, 922–6.
34. Eichler, J., Beyermann, M., and Bienert, M. (1989) Application of cellulose paper as support in simultaneous solid phase peptide synthesis, *Colect Czech Chem Commun* **54**, 1746–52.
35. Frank, R., and Döring, R. (1988) Simultaneous multiple peptide synthesis under continuous flow conditions on cellulose paper disks as segmental solid supports, *Tetrahedron* **44**, 6031–40.
36. Eichler, J., Bienert, M., Stierandova, A., and Lebl, M. (1991) Evaluation of cotton as a carrier for solid-phase peptide synthesis, *Peptide Res* **4**, 296–307.
37. Schmidt, M., and Eichler, J. (1993) Multiple peptide synthesis using cellulose-based carriers: synthesis of substance P – diastereoisomers and their histamine-releasing activity, *Bioorg Med Chem Lett* **3**, 441–46.
38. Daniels, S. B., Bernatowicz, M. S., Coull, J. M., and Köster, H. (1989) Membranes as solid supports for peptide synthesis, *Tetrahedron Lett* **30**, 4345–48.
39. Wang, Z., and Laursen, R. A. (1992) Multiple peptide synthesis on polypropylene membranes for rapid screening of bioactive peptides, *Pep Res* **5**, 275–80.
40. Wenschuh, H., Volkmer-Engert, R., Schmidt, M., Schulz, M., Schneider-Mergener, J., and Reineke, U. (2000) Coherent membrane supports for parallel microsynthesis and screening of bioactive peptides, *Biopolymers* **55**, 188–206.
41. Falsey, J. R., Renil, M., Park, S., Li, S., and Lam, K. S. (2001) Peptide and small molecule microarray for high throughput cell adhesion and functional assays, *Bioconjug Chem* **12**, 346–53.
42. Houseman, B. T., Huh, J. H., Kron, S. J., and Mrksich, M. (2002) Peptide chips for the quantitative evaluation of protein kinase activity, *Nat Biotechnol* **20**, 270–4.
43. Jonsson, U., Fagerstam, L., Ivarsson, B., Johnsson, B., Karlsson, R., Lundh, K., Lofas, S., Persson, B., Roos, H., Ronnberg, I. et al. (1991) Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology, *Biotechniques* **11**, 620–7.
44. Malmqvist, M. (1993) Biospecific interaction analysis using biosensor technology, *Nature* **361**, 186–7.
45. Angenendt, P., and Glokler, J. (2004) Evaluation of antibodies and microarray coatings as a prerequisite for the generation of optimized antibody microarrays, *Methods Mol Biol* **264**, 123–34.
46. Angenendt, P., Glokler, J., Murphy, D., Lehrach, H., and Cahill, D. J. (2002) Toward optimized antibody microarrays: a comparison of current microarray support materials, *Anal Biochem* **309**, 253–60.
47. Angenendt, P., Glokler, J., Sobek, J., Lehrach, H., and Cahill, D. J. (2003) Next generation of protein microarray support materials: evaluation for protein and antibody microarray applications, *J Chromatogr A* **1009**, 97–104.
48. Seurynck-Servoss, S. L., Baird, C. L., Miller, K. D., Pefaur, N. B., Gonzalez, R. M., Apiyo, D. O., Engelmann, H. E., Srivastava, S., Kagan, J., Rodland, K. D., and Zangar, R. C. (2008) Immobilization strategies for single-chain antibody microarrays, *Proteomics* **8**, 2199–210.
49. Seurynck-Servoss, S. L., Baird, C. L., Rodland, K. D., and Zangar, R. C. (2007) Surface chemistries for antibody microarrays, *Front Biosci* **12**, 3956–64.
50. Seurynck-Servoss, S. L., White, A. M., Baird, C. L., Rodland, K. D., and Zangar, R. C. (2007) Evaluation of surface chemistries for antibody microarrays, *Anal Biochem* **371**, 105–15.
51. Sobek, J., Aquino, C., and Schlapbach, R. (2007) Quality considerations and selection of surface chemistry for glass-based DNA, peptide, antibody, carbohydrate, and small molecule microarrays, *Methods Mol Biol* **382**, 17–31.
52. Sobek, J., Bartscherer, K., Jacob, A., Hoheisel, J. D., and Angenendt, P. (2006) Microarray

technology as a universal tool for high-throughput analysis of biological systems, *Comb Chem High Throughput Screen* **9**, 365–80.

53. Reineke, U., Volkmer-Engert, R., and Schneider-Mergener, J. (2001) Applications of peptide arrays prepared by the SPOT-technology, *Curr Opin Biotechnol* **12**, 59–64.
54. Frank, R. (2002) High-density synthetic peptide microarrays: emerging tools for functional genomics and proteomics, *Comb Chem High Throughput Screen* **5**, 429–40.
55. Frank, R., and Overwin, H. (1996) SPOT synthesis. Epitope analysis with arrays of synthetic peptides prepared on cellulose membranes, *Methods Mol Biol* **66**, 149–69.
56. Hilpert, K., Winkler, D. F., and Hancock, R. E. (2007) Peptide arrays on cellulose support: SPOT synthesis, a time and cost efficient method for synthesis of large numbers of peptides in a parallel and addressable fashion, *Nat Protoc* **2**, 1333–49.
57. Kramer, A., and Schneider-Mergener, J. (1998) Synthesis and screening of peptide libraries on continuous cellulose membrane supports, *Methods Mol Biol* **87**, 25–39.
58. Winkler, D. F., and Campbell, W. D. (2008) The spot technique the spot technique: synthesis and screening of peptide macroarrays on cellulose membranes, *Methods Mol Biol* **494**, 47–70.
59. Ay, B., Landgraf, K., Streitz, M., Fuhrmann, S., Volkmer, R., and Boisguerin, P. (2008) Using hydroxymethylphenoxy derivates with the SPOT technology to generate peptides with authentic C-termini, *Bioorg Med Chem Lett* **18**, 4038–43.
60. Ay, B., Streitz, M., Boisguerin, P., Schlosser, A., Mahrenholz, C. C., Schuck, S. D., Kern, F., and Volkmer, R. (2007) Sorting and pooling strategy: a novel tool to map a virus proteome for CD8 T-cell epitopes, *Biopolymers* **88**, 64–75.
61. Ay, B., Volkmer, R., and Boisguerin, P. (2007) Synthesis of cleavable peptides with authentic C-termini: an application for fully automated SPOT synthesis, *Tetrahedron Letters* **48**, 361–64.
62. Volkmer-Engert, R., Hoffmann, B., and Schneider-Mergener, J. (1997) Stable attachment of the HMB-linker to continuous cellulose membranes for parallel solid phase spot synthesis, *Tetrahedron Lett* **38**, 1029–32.
63. Kramer, A., Reineke, U., Dong, L., Hoffmann, B., Hoffmuller, U., Winkler, D., Volkmer-Engert, R., and Schneider-Mergener, J. (1999) Spot synthesis: observations and optimizations, *J Pept Res* **54**, 319–27.
64. Weiser, A. A., Or-Guil, M., Tapia, V., Leichsenring, A., Schuchhardt, J., Frommel, C., and Volkmer-Engert, R. (2005) SPOT synthesis: reliability of array-based measurement of peptide binding affinity, *Anal Biochem* **342**, 300–11.
65. Boisguerin, P., Ay, B., Radziwill, G., Fritz, R. D., Moelling, K., and Volkmer, R. (2007) Characterization of a putative phosphorylation switch: adaptation of SPOT synthesis to analyze PDZ domain regulation mechanisms, *Chembiochem* **8**, 2302–7.
66. Wiedemann, U., Boisguerin, P., Leben, R., Leitner, D., Krause, G., Moelling, K., Volkmer-Engert, R., and Oschkinat, H. (2004) Quantification of PDZ domain specificity, prediction of ligand affinity and rational design of super-binding peptides, *J Mol Biol* **343**, 703–18.
67. Fodor, S. P., Rava, R. P., Huang, X. C., Pease, A. C., Holmes, C. P., and Adams, C. L. (1993) Multiplexed biochemical assays with biological chips, *Nature* **364**, 555–6.
68. Stears, R. L., Martinsky, T., and Schena, M. (2003) Trends in microarray analysis, *Nat Med* **9**, 140–5.
69. MacBeath, G. (2007) Protein Arrays: Preparation of Microscope Slides, Cold Spring Harb. Protoc., doi:10.1101/pdb.prot4629.
70. Angenendt, P. (2005) Progress in protein and antibody microarray technology, *Drug Discov Today* **10**, 503–11.
71. Chatterjee, D. K., Sitaraman, K., Baptista, C., Hartley, J., Hill, T. M., and Munroe, D. J. (2008) Protein microarray on-demand: a novel protein microarray system, *PLoS ONE* **3**, e3265.
72. Hartmann, M., Roeraade, J., Stoll, D., Templin, M. F., and Joos, T. O. (2009) Protein microarrays for diagnostic assays, *Anal Bioanal Chem* **393**, 1407–16.
73. Borgia, J. A., and Fields, G. B. (2000) Chemical synthesis of proteins, *Trends Biotechnol* **18**, 243–51.
74. Muir, T. W., Dawson, P. E., and Kent, S. B. (1997) Protein synthesis by chemical ligation of unprotected peptides in aqueous solution, *Methods Enzymol* **289**, 266–98.
75. Schnolzer, M., and Kent, S. B. (1992) Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease, *Science* **256**, 221–5.
76. Hashida, S., Imagawa, M., Inoue, S., Ruan, K. H., and Ishikawa, E. (1984) More useful maleimide compounds for the conjugation of Fab'

to horseradish peroxidase through thiol groups in the hinge, *J Appl Biochem* **6**, 56–63.

77. Inouye, S., and Sato, J. (2008) Recombinant aequorin with a reactive cysteine residue for conjugation with maleimide-activated antibody, *Anal Biochem* **378**, 105–7.
78. Uto, I., Ishimatsu, T., Hirayama, H., Ueda, S., Tsuruta, J., and Kambara, T. (1991) Determination of urinary Tamm-Horsfall protein by ELISA using a maleimide method for enzyme-antibody conjugation, *J Immunol Methods* **138**, 87–94.
79. Yoshitake, S., Yamada, Y., Ishikawa, E., and Masseyeff, R. (1979) Conjugation of glucose oxidase from *Aspergillus niger* and rabbit antibodies using N-hydroxysuccinimide ester of N-(4-carboxycyclohexylmethyl)-maleimide, *Eur J Biochem* **101**, 395–9.
80. Lesaicherre, M. L., Uttamchandani, M., Chen, G. Y., and Yao, S. Q. (2002) Developing site-specific immobilization strategies of peptides in a microarray, *Bioorg Med Chem Lett* **12**, 2079–83.
81. Uttamchandani, M., Chen, G. Y., Lesaicherre, M. L., and Yao, S. Q. (2004) Site-specific peptide immobilization strategies for the rapid detection of kinase activity on microarrays, *Methods Mol Biol* **264**, 191–204.
82. Houseman, B. T., and Mrksich, M. (2002) Towards quantitative assays with peptide chips: a surface engineering approach, *Trends Biotechnol* **20**, 279–81.
83. MacBeath, G., Koehler, A. N., and Schreiber, S. L. (1999) Printing small molecules as microarrays and detecting protein–ligand interactions en masse, *J Am Chem Soc* **121**, 7967–68.
84. Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent, S. B. (1994) Synthesis of proteins by native chemical ligation, *Science* **266**, 776–9.
85. Green, N. M., and Toms, E. J. (1973) The properties of subunits of avidin coupled to sepharose, *Biochem J* **133**, 687–700.
86. Andresen, H., Grotzinger, C., Zarse, K., Kreuzer, O. J., Ehrentreich-Forster, E., and Bier, F. F. (2006) Functional peptide microarrays for specific and sensitive antibody diagnostics, *Proteomics* **6**, 1376–84.
87. Andresen, H., Zarse, K., Grotzinger, C., Hollidt, J. M., Ehrentreich-Forster, E., Bier, F. F., and Kreuzer, O. J. (2006) Development of peptide microarrays for epitope mapping of antibodies against the human TSH receptor, *J Immunol Methods* **315**, 11–8.
88. Lesaicherre, M. L., Lue, R. Y., Chen, G. Y., Zhu, Q., and Yao, S. Q. (2002) Intein-mediated biotinylation of proteins and its application in a protein microarray, *J Am Chem Soc* **124**, 8768–9.
89. Shin, D. S., Kim, D. H., Chung, W. J., and Lee, Y. S. (2005) Combinatorial solid phase peptide synthesis and bioassays, *J Biochem Mol Biol* **38**, 517–25.
90. Chiari, M., Cretich, M., Corti, A., Damin, F., Pirri, G., and Longhi, R. (2005) Peptide microarrays for the characterization of antigenic regions of human chromogranin A, *Proteomics* **5**, 3600–3.
91. Cerecedo, I., Zamora, J., Shreffler, W. G., Lin, J., Bardina, L., Dieguez, M. C., Wang, J., Muriel, A., de la Hoz, B., and Sampson, H. A. (2008) Mapping of the IgE and IgG4 sequential epitopes of milk allergens with a peptide microarray-based immunoassay, *J Allergy Clin Immunol* **122**, 589–94.
92. Jarvinen, K. M., Beyer, K., Vila, L., Chatchatee, P., Busse, P. J., and Sampson, H. A. (2002) B-cell epitopes as a screening instrument for persistent cow's milk allergy, *J Allergy Clin Immunol* **110**, 293–7.
93. Shreffler, W. G., Lencer, D. A., Bardina, L., and Sampson, H. A. (2005) IgE and IgG4 epitope mapping by microarray immunoassay reveals the diversity of immune response to the peanut allergen, Ara h 2, *J Allergy Clin Immunol* **116**, 893–9.
94. Komuro, A., Saeki, M., and Kato, S. (1999) Association of two nuclear proteins, Npw38 and NpwBP, via the interaction between the WW domain and a novel proline-rich motif containing glycine and arginine, *J Biol Chem* **274**, 36513–9.
95. Tessier, P. M., and Lindquist, S. (2007) Prion recognition elements govern nucleation, strain specificity and species barriers, *Nature* **447**, 556–61.
96. Pawson, T., and Nash, P. (2003) Assembly of cell regulatory systems through protein interaction domains, *Science* **300**, 445–52.
97. Tong, A. H., Drees, B., Nardelli, G., Bader, G. D., Brannetti, B., Castagnoli, L., Evangelista, M., Ferracuti, S., Nelson, B., Paoluzi, S., Quondam, M., Zucconi, A., Hogue, C. W., Fields, S., Boone, C., and Cesareni, G. (2002) A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules, *Science* **295**, 321–4.
98. Hu, H., Columbus, J., Zhang, Y., Wu, D., Lian, L., Yang, S., Goodwin, J., Luczak, C., Carter, M., Chen, L., James, M., Davis, R., Sudol, M., Rodwell, J., and Herrero, J. J. (2004) A map of WW domain family interactions, *Proteomics* **4**, 643–55.

99. Landgraf, C., Panni, S., Montecchi-Palazzi, L., Castagnoli, L., Schneider-Mergener, J., Volkmer-Engert, R., and Cesareni, G. (2004) Protein interaction networks by proteome peptide scanning, *PLoS Biol* **2**, E14.

100. Jones, R. B., Gordus, A., Krall, J. A., and MacBeath, G. (2006) A quantitative protein interaction network for the ErbB receptors using protein microarrays, *Nature* **439**, 168–74.

101. Gordus, A., and MacBeath, G. (2006) Circumventing the problems caused by protein diversity in microarrays: implications for protein interaction networks, *J Am Chem Soc* **128**, 13668–9.

102. Tapia, V., Bongartz, J., Schutkowski, M., Bruni, N., Weiser, A., Ay, B., Volkmer, R., and Or-Guil, M. (2007) Affinity profiling using the peptide microarray technology: a case study, *Anal Biochem* **363**, 108–18.

103. Nahtman, T., Jernberg, A., Mahdavifar, S., Zerweck, J., Schutkowski, M., Maeurer, M., and Reilly, M. (2007) Validation of peptide epitope microarray experiments and extraction of quality data, *J Immunol Methods* **328**, 1–13.

104. Fournel, S., and Muller, S. (2003) Synthetic peptides in the diagnosis of systemic autoimmune diseases, *Curr Protein Pept Sci* **4**, 261–74.

105. Duburcq, X., Olivier, C., Malingue, F., Desmet, R., Bouzidi, A., Zhou, F., Auriault, C., Gras-Masse, H., and Melnyk, O. (2004) Peptide–protein microarrays for the simultaneous detection of pathogen infections, *Bioconjug Chem* **15**, 307–16.

106. Durauer, A., Berger, E., Schuster, M., Wasserbauer, E., Himmler, G., Loibner, H., Mudde, G. C., and Jungbauer, A. (2006) Peptide arrays for the determination of humoral responses induced by active immunization with a monoclonal antibody against EpCAM, *J Immunol Methods* **317**, 114–25.

107. Cohen, I. R. (2007) Biomarkers, self-antigens and the immunological homunculus, *J Autoimmun* **29**, 246–9.

108. Quintana, F. J., Merbl, Y., Sahar, E., Domany, E., and Cohen, I. R. (2006) Antigen-chip technology for accessing global information about the state of the body, *Lupus* **15**, 428–30.

109. Quintana, F. J., Hagedorn, P. H., Elizur, G., Merbl, Y., Domany, E., and Cohen, I. R. (2004) Functional immunomics: microarray analysis of IgG autoantibody repertoires predicts the future response of mice to induced diabetes, *Proc Nattl Acad Sci USA* **101**(Suppl 2), 14615–21.

Chapter 2

Peptide Arrays for Enzyme Profiling

Alexandra Thiele, Johannes Zerweck, and Mike Schutkowski

Abstract

Enzymes are key molecules in signal transduction pathways. However, only a small fraction of more than 500 predicted human kinases, 250 proteases and 250 phosphatases is characterized so far. Peptide microarray-based technologies for extremely efficient profiling of enzyme substrate specificity emerged in the last years. Additionally, patterns of enzymatic activities could be used to fingerprint the status of cells or organisms. This technology reduces set-up time for HTS assays and allows the identification of downstream targets. Moreover, peptide microarrays enable optimization of enzyme substrates. A comprehensive overview regarding enzyme profiling using peptide microarrays is presented with special focus on assay principles.

Key words: Enzyme, substrate specificity, kinase, protease, phosphatase, peptide microarray, spot synthesis.

1. Introduction

Phosphorylation of proteins by protein kinases plays an essential role in the regulation of cellular processes such as signal transduction, cell proliferation and viability, differentiation, apoptosis and metabolism. Information about substrate proteins and peptides is necessary to integrate kinases into their biological networks. This can provide the basis for understanding molecular origins of diseases and for potentially developing tools for therapeutic intervention. The discovery of more than 500 members of these enzymes in the human genome stimulated a growing interest in protein kinases. Consequently, high-throughput technologies for determining kinase substrates have become a prerequisite for elucidating the huge number of potential phosphorylation events triggered by these kinases. This demand can be perfectly matched by peptide (micro)arrays, which have proved to be powerful tools

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_2,

for the rapid delineation of molecular recognition events. In this review we describe the application of peptide arrays for enzyme profiling with focus on kinase, phosphatase and protease research.

The scope of this review is peptide (micro)arrays on planar supports. Thus, we exclude the following topics from the considerations: peptide synthesis on polymeric pins, surfaces modified with only one peptide rather than a peptide collection as often applied in material sciences, read-out methods like surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF-MS) with peptides immobilized on the surface and surface plasmon resonance with immobilized peptides in standard configurations as well as small molecule and protein (micro)arrays.

2. Chemistry of Peptide Array Preparation

There are two main principles for the preparation of peptide arrays: in situ synthesis directly on the array surface and immobilization of pre-synthesized peptide derivatives. In general, in situ synthesis has a number of advantages compared to immobilization of pre-synthesized peptides. Normally, yields of peptides synthesized on surfaces are high and consistent over the entire support surface from one array region to another. It also permits combinatorial strategies for constructing large arrays of peptides in a few coupling steps. In 1991 two different technologies for the in situ preparation of peptide arrays were published. Light-directed, spatially addressable parallel chemical synthesis (1) is a synthesis technology permitting extreme miniaturization of array formats; however, it involves sophisticated and rather tedious synthesis cycles. A major problem is the novel set of chemistries. If these chemistries are not optimized, final quality of the surface-bound peptides will give false-positive (if an impurity is active) and/or false-negative results (if the target peptide sequence was not synthesized). An interesting alternative to circumvent this limitation is the use of photo-generated acids in combination with standard Boc-chemistry (2–4). Nevertheless, the use of photolithographic masks combined with solid-phase peptide synthesis is relatively labour intensive.

Alternatively, the SPOT synthesis concept developed by Frank is the stepwise synthesis of peptides on planar supports, such as functionalized cellulose membranes, applying standard Fmoc-based peptide chemistry (5, 6). SPOT synthesis is technically very simple and flexible and does not require any expensive laboratory automation or synthesis hardware. Nevertheless, the degree of miniaturization is significantly lower as compared to light-directed, spatially addressable parallel chemical synthesis. SPOT synthesis is very flexible and economic relative to other techniques and was

transformed from a semi-automatic procedure (7) into a fully automated system (www.jpt.com). The basic principle involves the spatially addressed deposition of defined volumes of activated amino acid derivatives directly onto a planar surface such as functionalized cellulose, aminated polypropylene (8) or aminopropylsilylated glass slides (9). The areas contacted by the droplets represent individual micro-reactors allowing the formation of a covalent bond between the amino acid derivative and the surface function. The resulting spot size is defined by the dispensed volume as well as the physical properties of the surface used. This SPOT synthesis has been reviewed extensively (8, 10–20).

A very elegant form of spatially addressed compound deposition makes use of modified colour laser printers. The cartridges are filled with a solvent/amino acid derivative mixture (the high melting point of the solvent yields a toner-like powder) resulting in an activated amino acid solution during the laser-induced melting process (21, 22). Analogous ink-jet delivery of activated amino acids to appropriate functionalized surfaces, such as membranes, microscope slides (23) or spinning surfaces in a CD-format (24) for automated synthesis of peptides, have been developed by a number of companies, but is not yet commercially available.

When large numbers of peptide arrays with the same sequences are required, immobilization of pre-synthesized peptides is more economical than in situ synthesis. Immobilization is also the method of choice for long peptide sequences, which normally have to be purified to obtain high-quality products. Chemoselective immobilization reactions are of particular interest in the preparation of peptide arrays because they allow control over both the orientation and the density of the attached peptides.

One intrinsic advantage of using chemoselective reactions via N-terminal reactivity tags is the introduced reactivity purification step. The resulting target peptide derivative is contaminated by acetylated truncated sequences only if the chemical moiety mediating the chemoselective reaction with the appropriately modified surface is attached to the N-terminus of target peptide and the peptide synthesis protocol is modified by introduction of capping steps subsequent to each coupling reaction. Deposition of this mixture results in formation of a covalent bond exclusively between the full-length target peptide derivative and the surface. The chemically "inert" truncated (and acetylated) sequences can be simply removed during subsequent washing steps. Thus, chemoselective reactions allow the generation of peptide arrays displaying purified peptides that are free of truncated sequences (*see* **Fig. 2.1**).

Different chemoselective reactions were used for peptide microarray preparation in connection with enzyme profiling experiments. An aldehyde function at the surface of glass slides in combination with amino-oxy-acetyl moieties in the peptides (15, 25–29) or cysteinyl residues (25, 30, 31) was used for the preparation of

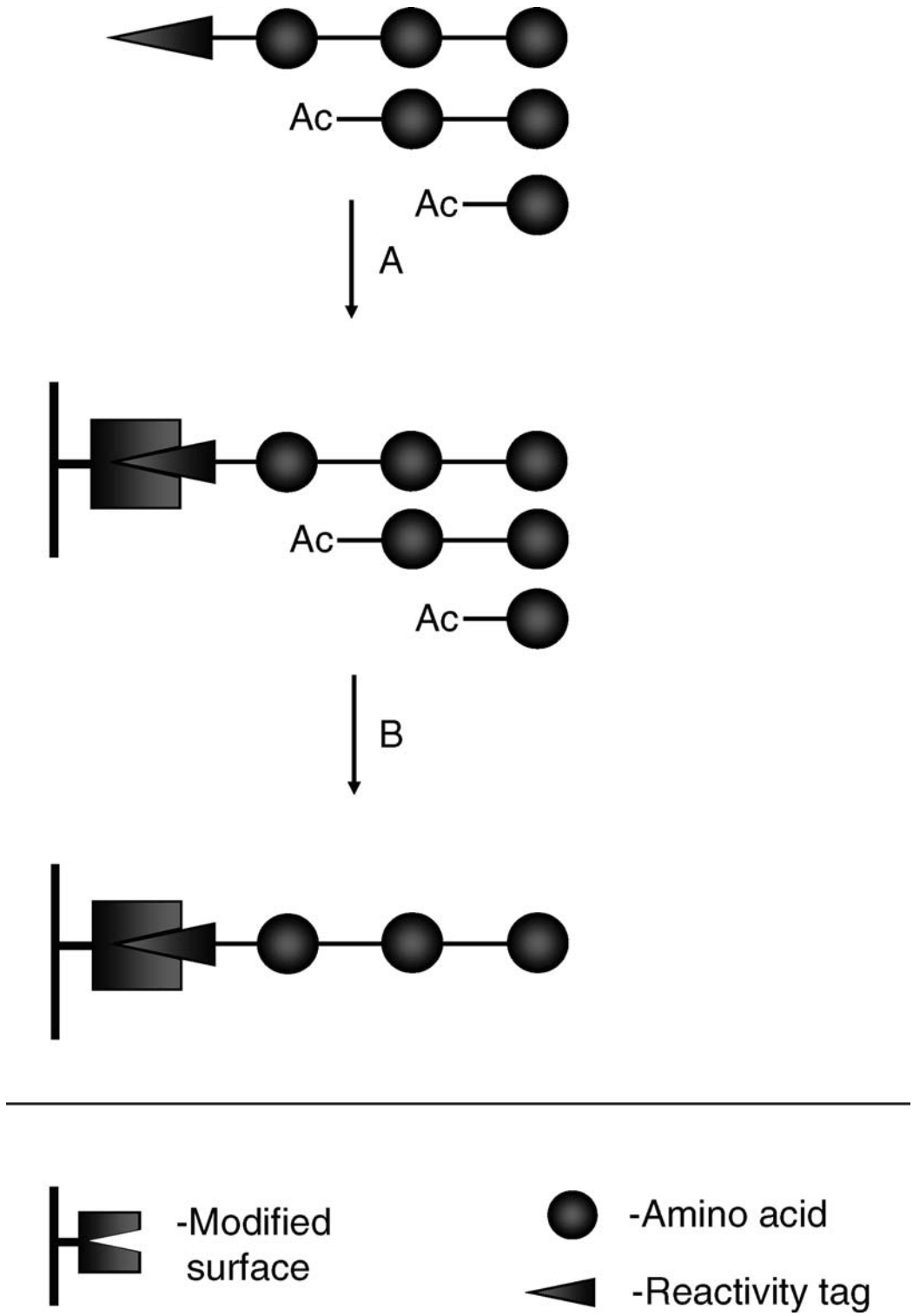


Fig. 2.1. Purification of Peptides During Chemoselective Immobilization. (**A**) Deposition of peptides mixtures and chemoselective immobilization of crude peptides which contain truncated sequences resulting from introduced acetylation

peptide microarrays on glass slides. The reaction between cysteine residues and surface-bound maleimide groups was used for preparation of peptide microarrays for kinase (32–37) and protease profiling (38). It could be demonstrated that native chemical ligation, introduced by Dawson et al. (39), is well suited for effective attachment of kinase substrate peptides containing an N-terminal cysteine residue to thioester-modified glass slides (40–42). A more sophisticated reaction for oriented immobilization of peptide derivatives was introduced by Houseman et al. (43). A Diels–Alder reaction between benzoquinone groups on self-assembled monolayers and cyclopentadiene–peptide conjugates led to efficient covalent attachment of kinase substrate peptides that were efficiently phosphorylated by c-Src kinase (43). Formation of amide bonds by Staudinger ligation between azide-modified phosphopeptides and appropriately phosphin-displaying glass surfaces was used for preparation of phosphopeptide microarrays enabling profiling of protein tyrosine phosphatase activities (44). Regioselective immobilization of poly(desoxythymidin)-modified kinase substrates onto differently coated glass slides was reported for PKA and c-Src (45). Photocleavable acrylamide labelled cysteine-containing kinase substrates were incorporated into peptide–acylamide copolymer hydrogel surfaces, and v-Abl- or c-Abl-mediated phosphorylation was detected by MALDI-TOF/TOF subsequent to laser-induced cleavage at the ß-(2-nitrophenyl)-ß-alanine residue (46). There are several additional chemistries used for the chemoselective immobilization of peptides onto different surfaces (comprehensively reviewed in (19)) like formation of covalent bonds by reaction of salicylhydroxamic acids with 1,3-phenydiboronic acid derivatives (47–49) or semicarbazides with aldehydes (50–54), but no applications for enzyme profiling have been described so far.

The surface-bound peptide's accessibility to the proteins or enzymes used in screening has also been identified as a critical factor. Insertion of a spacer between the peptide and the surface is an effective way to circumvent this potential problem. Generally, all linker molecules introduced to transform a given surface function into a functional group suitable for amino acid or peptide attachment can be considered as spacers. Such spacers can improve the efficiency of enzyme/substrate or antibody/peptide interactions on surfaces, as demonstrated with FLAG epitope peptides recognized by the monoclonal anti-FLAG M2 antibody (55). Additionally, for protein tyrosine kinase $p60^{c\text{-}src}$ only incorporation of the long and hydrophilic 1-amino-4,7,10-trioxa-13-tridecanamine succinimic acid building block spacer allowed effective phosphorylation of the glass surface-bound peptides (25). Similar linker structure was used

Fig. 2.1. (continued) steps subsequent to every coupling reaction. Truncated and acetylated peptides are non-reactive in contrast to the full-length target peptide equipped with the reactivity tag. (**B**) Washing steps subsequent to chemoselective immobilization yield purified covalently immobilized target peptide.

to space apart peptide nucleic acid (PNA) tags from potential protease substrates (56) or protein tyrosine kinase substrates (57). Inamori et al. described that insertion of a PEG spacer between chemoselective attachment point and the kinase substrate sequence improved phosphorylation efficiency by c-Src but not by PKA (32). Moreover, insertion of hydrophilic dextran structures between the surface and the presented peptides was described as necessary for efficient kinase substrate interaction (58).

An interesting alternative to spacers is the use of proteins decorated with peptides. MacBeath and Schreiber used covalently attached bovine serum albumin as a spacer molecule to present kinases with p42MAPK, PKA and CKII peptide substrates covalently attached to amino acid side chains of the albumin protein (59). Alternatively, substrate peptides fused genetically to the C-terminus of human leptin were immobilized onto aldehyde-modified glass slides (60). Using this method sensitivity was increased by three orders of magnitude compared to other microarray approaches with PKA and leptin–kemptide fusions. Another method is the decoration of streptavidin with biotinylated peptides or phosphopeptides (61). Sun et al. were able to demonstrate that biotinylated phosphopeptides immobilized onto streptavidin-coated glass surfaces could be efficiently dephosphorylated by phosphatases or isomerized by peptidyl-prolyl-*cis/trans*-isomerases (61). Nevertheless, Kimura et al. could not find detectable phosphorylation of streptavidin-bound biotinylated PKA substrates by PKA using fluorescently labelled anti-phosphopeptide antibodies but were able to demonstrate phosphorylation of the same substrates if regioselectively immobilized via polythymine–deoxyribonucleoside tag onto aminosilane- or polycarbodiimide-coated glass surfaces (45).

3. Library Types

A considerable number of different library types have been used for enzyme profiling and substrate identification with peptide arrays. One can define two general types of libraries: knowledge-based libraries comprising peptides with sequences that are derived from naturally occurring proteins, and libraries that are designed de novo, i.e. either consisting of randomly generated single peptides or peptide mixtures based on combinatorial principles.

The first type of knowledge-based library is of particular interest when the protein target for the enzyme is known. Identification of the actual phospho-acceptor residue or cleavage site is achieved by scans of overlapping peptides ("peptide scans") derived from

the protein's sequence (*see* **Fig. 2.2A**) (29). Alternatively, libraries of peptides only covering the sequence around each potential phospho-acceptor residue or protease cleavage site have been used (62, 63). The availability of high-density peptide microarrays

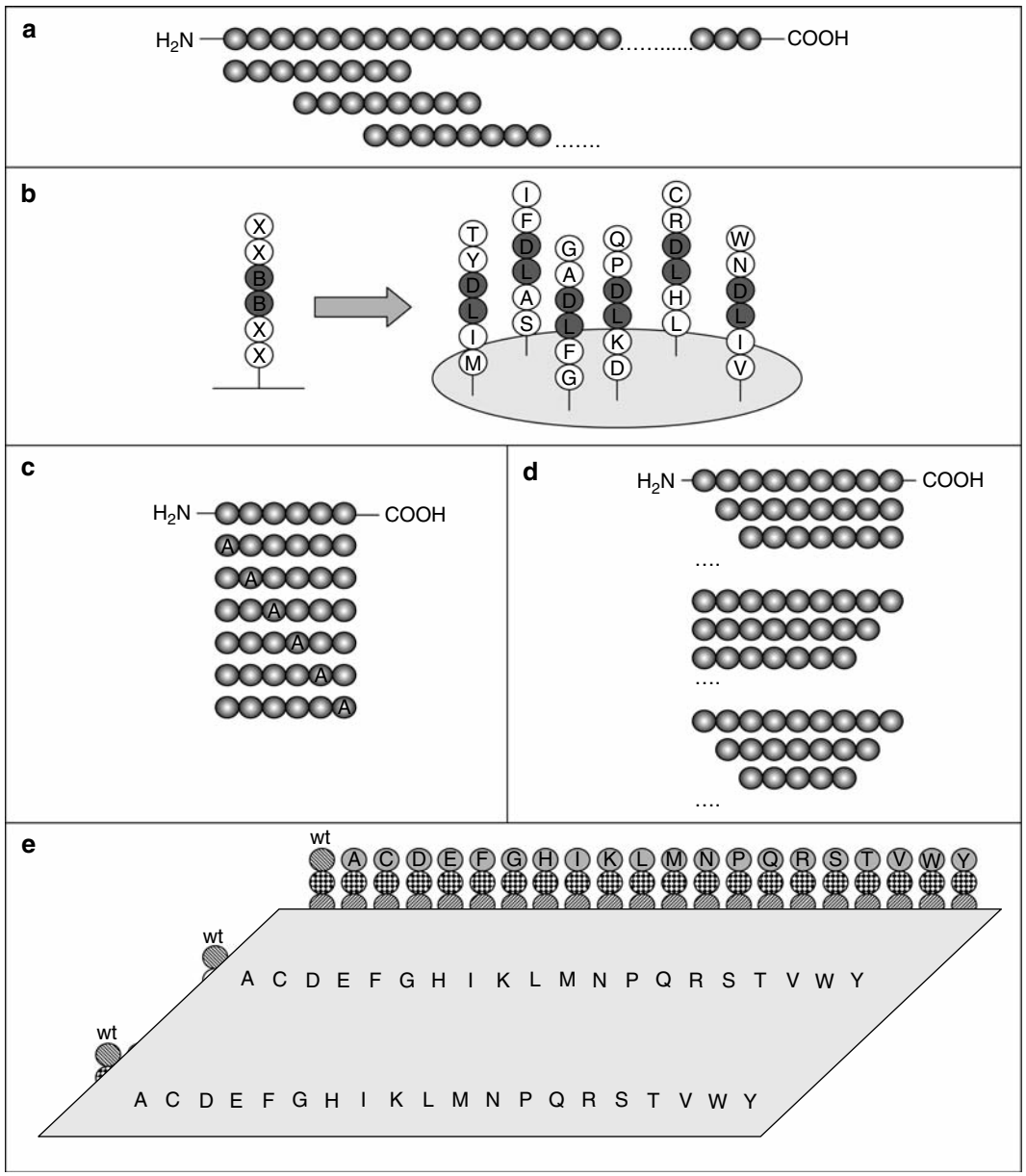


Fig. 2.2. Different Library Types. (**A**) The amino acid sequence of the protein under investigation is used to generate short linear overlapping 8-mer peptides shifted by three amino acids (peptide scan). (**B**) Peptide mixtures with defined positions *B* and randomized positions X. (**C**) Amino acid substitution scan (alanine scan) of a hexameric peptide. (**D**) Truncation library with N-terminal, C-terminal and bi-directional stepwise truncations. (**E**) Complete substitutional analysis of a 3-meric peptide.

enabled the systematic extension of this approach in a "proteomics-like manner", where addressing groups of proteins comprising all human peptidyl-prolyl-*cis/trans*-isomerases in the form of overlapping peptides resulting in more than 3250 peptides or covering the complete proteome of human cytomegalovirus (17,181 overlapping peptides) immobilized onto one standard industry glass slide in duplicates. Incubation of these high-content peptide microarrays with kinases and fluorescence scanning subsequent to treatment with phosphospecific dye yielded proteome-wide detection of phosphorylation sites (64). In a similar approach, the sequences of experimentally identified phosphorylation sites taken from databases (i.e. Swissprot (65); Phosphobase (66)) and the literature were comprehensively evaluated on peptide arrays (15, 26, 27, 29, 35, 36, 67–76). The use of such peptides increases the probability of finding substrates for a given kinase since each peptide's sequence is known to be phosphorylated by a kinase at least in the context of the natively folded protein. One problem associated with knowledge-based libraries is the uncertainty of determination which residue is phosphorylated when multiple phospho-acceptor sites occur in one individual peptide. When using overlapping peptide scans, this information can be extracted from increasing and decreasing signal intensities appearing along consecutive peptides. This is illustrated by scanning myelin basic protein phosphorylated by cAMP-dependent kinase, as shown in **Fig. 2.3**. Statistical analysis and alignment of the sequences proved to be useful in the case of collections of annotated phosphorylation sites (27). Identification and alignment of key residues in the different substrate peptides allow a reliable assignment of the actual phospho-acceptor residue(s).

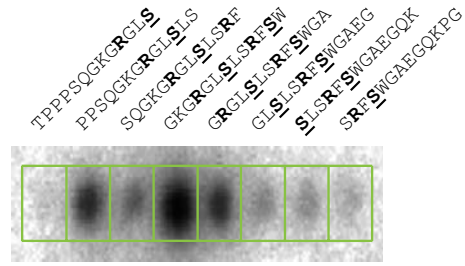


Fig. 2.3. Determination of Phosphorylation Sites Using Overlapping Peptide Scans. Section of the phosphorimage of a peptide microarray displaying a scan of overlapping peptides derived from myelin basic protein MBP (13-mers overlapping by 11 amino acids) after phosphorylation with protein kinase A in the presence of [γ-^{32}P]-ATP. Serine residues shown to be phosphorylated in the native MBP (143) are written in bold and underlined. The key residue arginine is shown in *bold*. The strongest signal is observed when both phosphorylation sites are in the central region of the substrate. When the respective phosphorylation site is positioned at one of the peptide termini, phosphorylation is not effective.

An extension of the knowledge-based libraries concept is the introduction of post-translational modifications, e.g. phosphorylation, methylation or acetylation, within the substrate sequences. This more adequately mimics the natural environment in which phosphorylation occurs, allowing the detection of peptides that become substrates only after an initial priming modification event. Such modifications can be introduced on-chip enzymatically after chemical synthesis of the unmodified peptides (29) or chemically by either "off-chip" using modified building blocks during the course of peptide synthesis synthesis (29) or by on-chip-chemistry using chemoselective chemical reactions as demonstrated by selective acetylation of lysine side chains in microarray-bound peptides (77).

Another type of knowledge-based libraries allows the mapping of protein interactions involving two discontinuous components that are far apart in the primary polypeptide structure but form a composite phosphorylation/dephosphorylation site in the natively folded protein. Two separate peptides are synthesized independently by a double peptide synthesis method on a single spot allowing the detection of synergistic pairs of peptides for protein tyrosine phosphatase 1B and for serine/threonine kinase Erk2 (78).

For de novo detection of kinase substrates, both combinatorial approaches and randomly generated libraries of single peptides proved to be useful. Combinatorial libraries have one or more defined amino acid positions and a number of randomized or degenerated positions (41, 58, 79–83). Only one particular amino acid is introduced at the defined positions while a mixture of amino acids is introduced at the randomized positions, resulting in a sub-library of different sequences in each single spot (*see* **Fig. 2.2B**). The number of individual sequences per spot depends on the number of randomized positions and the number of different building blocks used for these positions. A very high diversity can be achieved due to the huge number of different peptides. Once the amino acids that are productive for phosphorylation by a given kinase have been identified at the defined positions, the remaining randomized positions must be deconvoluted using follow-up libraries. Combinatorial libraries were successfully used with cellulose membranes as the solid support. However, representation of each single sequence of a peptide mixture is not guaranteed when using peptide microarrays with a low concentration of peptide per spot on planar surfaces (41).

The tremendous miniaturization of peptide libraries possible on planar surfaces such as glass slides enables the application of randomly generated libraries of single peptides that cover a significant, although not complete part of the potential sequence space. Such randomly generated libraries for kinase substrate identification and kinase profiling have a defined phospho-acceptor residue and random sequences in the flanking areas (28). In

contrast to combinatorial libraries, each spot represents one single sequence. If information on the consensus sequence for the substrates of a kinase is available, the random libraries can be biased by introducing defined positions derived from the consensus sequences. Randomly generated libraries show higher sequence diversity compared to knowledge-based libraries that are biased towards known kinase substrates. This can be an advantage in searching for selective substrates for closely related enzymes.

Substrate characteristics, i.e. key interaction residues, can be deduced from all these library types using statistical analysis, provided that the number of identified substrates is high enough (27, 28, 80–84). Alternatively, different library types based on single substrate sequences, such as alanine scans (*see* **Fig. 2.2C**), deletion (*see* **Fig. 2.2D**) (41) and substitutional libraries (*see* **Fig. 2.2E**) (26, 67, 85, 86), permit comprehensive substrate characterization.

4. Assays and Detection

Measuring the activity of enzymes that modify peptides on microarrays involves either addition of chemical moieties to displayed peptides (kinases, acetyl transferases, glycosyltransferases, ADP-ribosyl transferases, etc.) or release of a part of the immobilized peptide derivative (proteases, phosphatases, demethylases, etc.). In principle, all enzyme-modifying peptides or proteins can be applied to screen peptide arrays. However, so far only studies with kinases, phosphatases, proteases, (*see* **Table 2.1**), lysine methyltransferases (87, 88), isomerases (89, 90), glycosyltransferases (91, 92), ADP-ribosyltransferases (93), hydrolases (94), esterases (95) and SUMO ligases (96) have been described using peptide arrays or peptide microarrays. Following chapter will review application of peptide arrays for kinase, phosphatase and protease research in more detail.

There are three general assay principles applied to detect phosphorylated peptides on peptide arrays. One way is to incorporate a radioactive label during the phosphorylation reaction using [γ-$^{32/33}$P]-ATP (*see* **Fig. 2.4A**). Subsequently, quantification of incorporated radioactivity is achieved using either a phosphorimager (27–29) or X-ray films, or alternatively photographic emulsions that deposit silver grains directly onto the glass surface (59). This procedure has a low limit of detection and is only influenced by the selectivity of the kinase. Incubation protocols have been described for peptide arrays prepared by SPOT synthesis (97, 98) and peptide microarrays (27–29, 71, 99). However, for reasons of operational safety, ease of handling and waste disposal, radioactive detection methods are increasingly avoided.

Table 2.1
Bibliography of applications using peptide (micro)arrays for enzyme profiling and substrate identification. The table lists screening molecules, library types, array preparation technologies and each publication citation

Screening molecule	Library	Technology	Reference
Kinases and Phosphatases			
Protein kinase A (PKA); cytoplasmic (kinase) domains of transforming growth factor ß (TGFß) type I and II receptors; γ^{32}P-ATP	Several combinatorial libraries with fixed amino acids and randomized positions (17 amino acids excluding Cys, Ser, Thr) at different positions	SODA	(58)
Protein tyrosine kinase Lyn, anti-phospho-tyrosine antibody and POD-labelled second antibody for detection	Panel of 23 PKCδ-derived 15-mer peptides	SPOT	(62)
Catalytic subunit of protein kinase A or cGMP activated protein kinase G and γ^{32}P-ATP	Combinatorial library of the type $XXXO_1O_2XXX$ and iterative deconvolution libraries	SPOT	(83)
– Anti-porcine tubulin-tyrosine ligase (TTL) mab 1D3 – Streptavidin-alkaline phosphatase – cAMP-dependent protein kinase	– Substitutional and truncation analysis of a TTL-derived peptide; combinatorial library (XXB_1B_2XX) and deconvolution libraries – Combinatorial library (XXB_1B_2XX) and deconvolution libraries – Combinatorial library (XXB_1B_2XX) and deconvolution libraries	SPOT	(145)
– Casein kinase I (CKI) or II (CKII) or protein kinase C (PKC) or catalytic subunit of protein kinase A (PKA) and γ^{32}P-ATP – Potato acid phosphatase	– Peptides derived from the phosphorylation sites RRASVA, QKRPSQRAKYL, DDDDEESITRR, DDDSDDDAAAA – RRASS*VA, QKRPS*QRAKYL, DDDDEES*ITRR, DDDS*DDDAAAA (S* = phospho-serine)	SPOT	(146)
Protein kinase A (PKA), protein kinase C (PKC), casein kinase I (CKI), casein kinase II (CKII) and γ^{32}P-ATP	Panel of human phosphorylation sites	SPOT	(67)
–^{125}I-Calmodulin	– phosphorylated and unphosphorylated peptides derived from cell–cell adhesion	SPOT	(63)

(continued)

Table 2.1(continued)

Screening molecule	Library	Technology	Reference
– Porcine protein kinase C β and γ ^{32}P-ATP	molecule (C-CAM) isoforms S and L and single site substitution analogs – Partial peptide scans of the C-CAM isoforms S and L		
Androgen receptor protein kinase and γ^{32}P-ATP	Panel of kinase substrates known from literature	SPOT	(135)
Enzyme I of bacterial sugar phosphotransferase system and ^{32}P-phospho*enol*pyruvate	Combinatorial library of the type $XXXXXO_1HO_2XXXXX$ and iterative deconvolution libraries	SPOT	(79)
Catalytic subunit of protein kinase A or cGMP activated protein kinase G and γ^{32}P-ATP	Combinatorial libraries of the type $XXXO_1O_2XXX$ and $XXXRRO_1O_2X$	SPOT	(97)
Catalytic subunit of protein kinase A or cGMP activated protein kinase G and γ^{32}P-ATP	Combinatorial libraries of the type $O_1KARKKSNO_2$, $O_1O_2TQAKRKKSLA$, $O_1O_2KATQAKRKKSLA$, $TQAKRKKSLAO_1O_2$ and $TQAKRKKSLAMAO_1O_2$	SPOT	(80)
Catalytic subunit of protein kinase A or cGMP activated protein kinase G and γ^{32}P-ATP	Combinatorial library of the type $XXXO_1O_2XXX$ and deconvolution libraries: $XXXRKO_1O_2X$, $XRKKKO_1O_2X$, $O_1RKKKKKO_2$, $LRKKKKKHO_1O_2$ and $O_1O_2LRKKKKKH$	SPOT	(84)
GST-Dyrk1A-Δ(500-763) and γ^{32}P-ATP	Partial substitutional analysis of RRRFRPASPLRGPPK	SPOT	(85)
Maize Ca^{2+}-dependent protein kinase (CDPK-1) and γ^{32}P-ATP	Partial substitutional analysis of LARLHSVRER	SPOT	(86)
Protein kinase A (PKA), casein kinase II (CKII), p42-MAP kinase (Erk2) and γ^{32}P-ATP	Kinase substrates Kemptide and Elk1; protein kinase inhibitor 2	DIPP	(59)
– $p60^{c\text{-}src}$ Protein tyrosine kinase and γ^{32}P-ATP – Strepavidin-Cy3 conjugate; avidin-Cy5 conjugate – Anti-human insulin mab HB125 and anti-mouse IgG-Cy5 conjugate – WEHI-231 cells and negative control cells	– Ttds-EEIYGEFF – biotin, HPYPP and WSHPQFEK – Biotin, wGeyidvk, pqrGstG, WSHPQFEK and YGGFL – wGeyidvk	DIPP	(25)
Cyclic GMP-dependent protein kinase (phosphorylation γ^{32}P-ATP)	– $XXXRKB_1B_2X$ (deconvolution libraries described but data not shown)	SPOT	(81)

(continued)

Table 2.1(continued)

Screening molecule	Library	Technology	Reference
^{32}P-labelled cyclic GMP-dependent protein kinase (binding)	– B_1RKKKKKB$_2$ (preceeding libraries described but data not shown)		
– GST-protein tyrosine phosphatase 1B; mixture of anti-phospho-tyrosine antibodies, POD-conjugated anti-mouse antibody	– 15-mer phospho-tyrosine-containing peptides derived from human insulin receptor; degenerate SPOT library of general structure AABX$_1$ZX$_2$BAA (Z = phospho-tyrosine, B = mixture of 20 natural amino acids, X defined amino acids) and peptides derived from human STAT 5A containing both phospho-tyrosine and phospho-threonine residues	SPOT	(116)
– GST-protein tyrosine phosphatase ß; mixture of anti-phospho-tyrosine antibodies, POD-conjugated anti-mouse antibody	– 15-mer phospho-tyrosine-containing peptides derived from human Tie2 receptor containing additional phospho-serine or phospho-threonine residues		
– GST-fusion protein of substrate trapping mutant of protein tyrosine phosphatase 1B (D181A), radioactively labelled by incubation with protein kinase A in the presence of γ^{32}P-ATP	– 15-mer phospho-tyrosine-containing peptides derived from human insulin receptor; degenerate SPOT library of general structure AABX$_1$ZX$_2$BAAA (Z = phospho-tyrosine, B = mixture of 20 natural amino acids, X defined amino acids), progressive alanine substitution of MTRDIYETD YZRKGG (Z = phospho-tyrosine), alanine scan of TRDIYETDYZ RKGGKGL, substitutional analysis for X in MTRDIYETDXZRKGG (Z = phospho-tyrosine)		
c-Src and γ^{32}P-ATP or anti-phospho-tyrosine antibody	Panel of known kinase substrates	DIPP	(43)
c-Src and γ^{32}P-ATP	c-Src substrate	DIPP	(147)
p60 and FITC-labelled anti-phospho-tyrosine antibody	YIYGSFK, ALRRASLG, KGTGYIKTG and monophosphorylated derivatives	DIPP	(30)
PKA or p60 and FITC-labelled anti-phospho-tyrosine or anti-phospho-serine antibody	YIYGSFK, ALRRASLG, YIYGSFK and monophosphorylated derivatives	DIPP	(40)

(continued)

Table 2.1(continued)

Screening molecule	Library	Technology	Reference
NEK6 kinase and γ^{32}P-ATP	Human annotated phosphorylation sites and mutational analysis of GLAKSFGSPNRAY	DIPP	(26)
Protein tyrosine kinase Abl and γ^{32}P-ATP or FITC-labelled anti-phospho-tyrosine antibody	Collection of 720 peptides (13-mer) derived from human annotated phosphorylation sites (databases Swiss-Prot and Phosphobase)	DIPP	(15)
– GST-ERK2 or GST-MEK-EE, anti-phospho-Elk-1 antibody, POD labelled anti-rabbit antibody	– Double peptide synthesis of ELK1-substrate peptide FWSTLSPIAPR, D-loop peptide KGRKPRDLELP and control peptide MNGGAANGRIL	SPOT	(78)
– GST-fusion protein-protein tyrosine phosphatase 1B;anti-phospho-tyrosine antibody, POD-conjugated anti-mouse antibody	– Double peptide synthesized SPOTs containing the PTP-1B substrate IYETDYZRKGG (Z = phospho-tyrosine) and on the second site a scan of 11-mer overlapping peptides derived from the cytoplasmic domain of insulin receptor		
– GST-fusion protein of substrate trapping mutant of protein tyrosine phosphatase 1B (D181A), radioactively labelled by incubation with protein kinase A in the presence of γ^{32}P-ATP	– Double peptide synthesis of insulin receptor derived substrate peptide IYETDYZRKGG (Z = phospho-tyrosine) and binding motif for YAP WW1 domain and p53 binding protein-2, respectively (YPPYPPPPYPS)		
PKA and Abl; Pro-Q Diamond phosphospecific stain or FITC-labelled anti-phospho-tyrosine antibody for detection	Kemptide, p60 c-Src (521–533), delta sleep inducing peptide (DSIP), phosphoDSIP, CamKII peptide (GS1-10), different proteins	DIPP	(100)
GST-proteins of substrate trapping mutants of tyrosine phosphatases (PTP-H1, SAP-1, TC-PTP, PTP-1B) radioactively labelled by incubation with protein kinase A in the presence of γ^{32}P-ATP	7 Peptides (14-mers) derived from human GHR together with the appropriate phospho-tyrosine-containing derivatives	SPOT	(117)
Tyrosine kinase p60 c-Scr and FITC-labelled anti-phospho-tyrosine antibody	Deletion-, alanine-scanning, positional scanning- and full combinatorial mixture libraries of CGG-YIYGSFK (p60 c-src substrate)	DIPP	(41)

(continued)

Table 2.1(continued)

Screening molecule	Library	Technology	Reference
PKA or tyrosine kinase p60 c-Scr and FITC-labelled anti-phospho-serine or antiphospho-tyrosine antibody,respectively	Peptides YIYGSFK and ALRRASLG N-terminally either biotinylated or cysteine-modified	DIPP	(42)
Protein kinase A and tyrosine kinase Abl and Cy5-labelled antibodies	Kemptide–leptin fusion protein and Abl-peptide–leptin fusion protein	DIPP	(60)
– Phosphospecific antibodies MPM2, 3F3/2, PY20 – GST fusions of GRB-SH2, SHP2-CSH2, GRB7-SH2, GRB 10-SH2 – Protein kinase A (PKA), Cdc 15	– Complete L-amino acid substitutional analysis of AXXXX[pS/pT]XXXXA and AXXXX[pYT]XXXXA – Complete L-amino acid substitutional analysis of AXXXX[pYT]XXXXA – Complete L-amino acid substitutional analysis of $AX_1X_1X_1X_1[S/T]X_1X_1X_1X_1A$; X = L-amino acids exept for Cys, X1 = L-amino acids except for Cys, Ser, Thr	SPOT	(82)
Tyrosine kinase Abl and CKII together with FITC-labelled anti-phospho-tyrosine antibody and γ^{32}P-ATP, respectively FITC-labelled anti-phospho-tyrosine antibody	Collection of more than 13,000 peptides (13-mer) derived from human proteins 2923 Phospho-peptides (13-mer) derived from human proteins	DIPP	(27)
Tyrosine kinase Abl and γ^{32}P-ATP	1433 Randomly generated 15-meric peptides	DIPP	(28)
Kinases PDK1, Tie2, CKII, PKA, GSK3 and γ^{32}P-ATP	Peptide scans trough MBP and Tie2, 720 human annotated phosphorylation sites (13-mer), 2923 phospho-peptides (13-mer) derived from human proteins, 1394 peptides derived from activation loops of human kinases	DIPP	(29)
Kinases c-Src, PKA, PKG, CaMKII, CKI, Abl, Erk	7 Peptides each of it is a known specific substrate for one of the kinases	DIPP	(112)
Kinases and phosphatases contained in lysates of IL1-stimulated HeLa cells	Peptide scan through p65 NF-κB, selected peptides from the scan and controls with systematic mutation of Ser, Thr, Tyr to non-phospho-acceptor residues	SPOT	(148)

(continued)

Table 2.1(continued)

Screening molecule	Library	Technology	Reference
Purified PKA or cell lysates from human peripheral blood mononuclear cells before and after stimulation with lipopolysaccharide in the presence of radioisotopically labelled γ^{33}P-ATP followed by phosphorimaging	192 Nonapeptides providing consensus sequences across the mammalian kinome including eight rhodamine-labelled irrelevant control peptides	DIPP	(73)
Protein kinase A and Biotin-ATP (adenosine 5́-triphosphate γ-biotinyl-3,6,9-trioxaundecanediamine) followed by avidin-stabilized gold nanoparticles and signal amplification by silver deposition, detection by resonance light scattering of nanoparticles	Kemptide derivative LRRASLG and negative control peptide LRRAGLG chemoselectively attached via N-terminal amino function	DIPP	(109)
Protein kinase A and tyrosine kinase c-Src, readout either radioisotopically labelled γ^{33}P-ATP or by biotinylated zinc(II) chelate phosphate sensor N-(5-(2-(+)-biotin aminoethylcarbamoyl)pyridine-2-ylmethyl)-N,Ń,Ǹ-tris(pyridine-2-ylmethyl)-1,3-diaminopropan-2-ol (Phos-tag biotin) and readout using surface plasmon resonace subsequent to incubation with streptavidin and anti-streptavidin antibody	Model substrates for PKA (Kemptide derivative) and c-Src (IYGEFKKK) together with replacements of phosphorylation site by either phosphoamino acid or by alanine and phenylalanine, respectively; chemoselective immobilization via N-terminal cysteine residue	DIPP	(32)
Protein kinase A and tyrosine kinase c-Src; readout either radioisotopically labelled γ^{33}P-ATP or generic anti-phosphoamino acid antibodies or by biotinylated zinc(II) chelate phosphate sensor N-(5-(2-(+)-biotin aminoethylcarbamoyl)pyridine-2-ylmethyl)-N,Ń,Ǹ-tris(pyridine-2-ylmethyl)-1,3-diaminopropan-2-ol (Phos-tag biotin) and readout using surface plasmon resonance subsequent to incubation with streptavidin	Model substrates for PKA (Kemptide derivative) and c-Src (IYGEFKKK) together with replacements of phosphorylation site by either phosphoamino acid or by alanine and phenylalanine, respectively; chemoselective immobilization via N-terminal cysteine residues	DIPP	(33)

(continued)

Table 2.1(continued)

Screening molecule	Library	Technology	Reference
Protein kinase A, protein kinase C, tyrosine kinase c-Src; or forskolin/H-89 stimulated MCF-7 cell lysates; readout fluorescence imaging subsequent to incubation with biotinylated zinc(II) chelate phosphate sensor N-(5-(2-(+)-biotin aminoethylcarbamoyl)pyridine-2-ylmethyl)-N,Ń,Ǹ-tris(pyridine-2-ylmethyl)-1,3-diaminopropan-2-ol (Phos-tag biotin) and Cy5-labelled streptavidin	Derivatives of substrates for PKA (Kemptide derivatives), PKC (LRVQNSLRRRR) and c-Src (IYGEFKKK) together with replacements of phosphorylation site by either phosphoamino acid or by alanine and phenylalanine, respectively; chemoselective immobilization via N-terminal cysteine residues	DIPP	(34)
Casein kinase 2 and radioisotopically labelled γ^{32}P-ATP followed by autoradiography	10-mer peptides derived from human high-mobility group protein SSRP1 together with derivatives where potential phosphorylation site was replaced by alanine residue	SPOT	(149)
Protein kinase A and fluorescent phosphospecific stain Pro-Q Diamond	Sequences derived from R(R/K)XSLG (X = 20 natural amino acids) together with synthetically phosphorylated and phosphorylation negative analogs	LDPS	(101)
GST-fusions of kinases SRPK4 and MPK3 or subcellular fractions (cytosolic or nuclear extracts) and lysates of H_2O_2 treated *Arabidopsis* cells in the presence of radioisotopically labelled γ^{33}P-ATP followed by phosphorimaging	47 Peptides (11-mers) derived from in vivo *Arabidopsis* phosphorylation sites detected by mass spectrometry	DIPP	(150)
Purified protein tyrosine kinases Yes and Fyn or de-yolked zebrafish embryos lysed at 24 hour post fertilization or zebrafish lysates of morpholino-mediated knock down of Fyn, Yes, Wnt1 and Nacre. Kinetics of kinase mediated phosphate transfer were recorded by semi-continuous CCD imaging of fluorescence caused by bound fluoresceine labelled, co-incubated PY20 anti-phospho-tyrosine antibody	144 Peptide derivatives containing 13-meric sequences derived from annotated human tyrosine phosphorylation sites	DIPP	(35)

(continued)

Table 2.1(continued)

Screening molecule	Library	Technology	Reference
Zebrafish embryos lysed at 3 or 5 days post fertilization; kinetics of kinase mediated phosphate transfer to microarray bound substrate peptides were recorded by semi-continuous CCD imaging of fluorescence caused by bound fluoresceine labelled, co-incubated PY20 anti-phospho-tyrosine antibody	144 Peptide derivatives containing 13-meric sequences derived from annotated human tyrosine phosphorylation sites	DIPP	(36)
Lysates of tissue samples (either frozen or fresh) of 30 Barretťs oesophagus patients in the presence of radioisotopically labelled γ^{33}P-ATP followed by phosphorimaging, control incubations in the presence of α-^{33}P-ATP showed no incorporation of radioisotopically labelled phosphate demonstrating that no unspecific binding of ATP or ATP-protein complexes to displayed peptides takes place.	1164 Peptides derived from human phosphorylation/consensus sites including 12 control sequences in duplicates	DIPP	(74)
PKA, PKCα, Erk2, PKCδ, JAK1, CamKII or tyrosine kinases Abl, c-Src and insuline receptor kinase; readout using surface plasmon resonance either subsequent to treatment with biotinylated zinc(II) chelate phosphate sensor N-(5-(2-(+)-biotin aminoethylcarbamoyl)pyridine-2-ylmethyl)-N,Ń,Ǹ-tris(pyridine-2-ylmethyl)-1,3-diaminopropan-2-ol (Phos-tag biotin) followed by streptavidin and anti-streptavidin antibody or subsequent to treatment with anti-phospho-tyrosine antibody, respectively.	26 Different potential substrates with N-terminal cysteine spaced by two glycine residues from the peptide sequence, peptide derivatives were chemoselectively immobilized onto maleimide-modified self-assembled monolayers on gold surfaces	DIPP	(37)
Kinases CK1, Plk3, Akt1, SGK1, PKCζ, PKCα, PKA, AuroraA, AuroraB, Chk2, CaMK2, MK2, Chk1, Cdk5, Cdk2, p38, CK2, Abl, Lck, c-Src, EphB2, VEGFR2, Kit, Flt3, EGFR and	900 Peptide substrates that targeted over 50 different human kinases and 102 controls for monitoring assay performance	DCSD	(103)

(continued)

Table 2.1(continued)

Screening molecule	Library	Technology	Reference
Jak2, readout using fluorescence signal of silica beads decorated with specific oligonucleotide sequences and placed in etched microwells of fiber-optic bundles. Subsequent to incubation with DNA-peptide conjugates all acidic side chains are chemically blocked followed by chemical transformation of phosphate moieties into dye-labelled derivatives. Dye-labelled solution is hybridized to oligonucleotides attached to individual beads. Dye is detected with anti-dye antibody followed by biotinylated anti-mouse antibody and Alexa555 labelled streptavidin using fluorescence imaging. Additionally, mixtures of kinases in the presence or absence of specific inhibitors were used.			
Either purified recombinant KPI-2 (Lmr2, CprK) kinase domain or full-length KPI-2 kinase (immunoprecipitated from HEK293T cells by c-Myc tag) in the presence of radioisotopically labelled γ^{33}P-ATP followed by phosphorimaging	1164 Peptides derived from human phosphorylation/consensus sites including 12 control sequences in duplicates	DIPP	(151)
Kinases Akt1, PKC, PKA, Cdk5, p38, CK2, ERK1 in the presence of radioisotopically labelled γ^{32}P-ATP, readout using phosphorimaging	95 Peptides (15-mers) derived from human phosphorylation sites found with mass spectrometry including negative controls of this phosphorylation sites (Ser/Ala, Thr/Val replacements)	DIPP	(99)
Lysates from *P.pastoris, T.aestivum, C.albicans, A.thaliana, F.solani*, emb. Liver cells, human B- and T-cells and murine macrophages in the presence of radioisotopically labelled γ^{33}P-ATP followed by phosphorimaging; control incubations in the presence of α^{33}P-ATP showed no incorporation of	1164 Peptides (11-mers) derived from human phosphorylation/consensus sites including 12 control sequences in duplicates	DIPP	(72)

(continued)

Table 2.1(continued)

Screening molecule	Library	Technology	Reference
radioisotopically labelled phosphate demonstrating that no unspecific binding of ATP or ATP-protein complexes to displayed peptides takes place.			
Extracts from dexamethason treated or non-treated human CD4+ T-cells and insulin-stimulated 3T3 adipocytes in the presence of radioisotopically labelled γ^{33}P-ATP followed by phosphorimaging	1164 Peptides derived from human phosphorylation/consensus sites including 12 control sequences in duplicates	DIPP	(75)
Extracts from dexamethason treated or non-treated human CD4+ T-cells stimulated for 15 min with anti-CD3 and anti CD28 antibodies in the presence of radioisotopically labelled γ^{33}P-ATP followed by phosphorimaging	1164 Peptides derived from human phosphorylation/consensus sites including 12 control sequences in duplicates	DIPP	(76)
Recombinant tomato MAPKs LeMPK1, LeMPK2 and LeMPK3 and *A.thaliana* kinase MPK4 in the presence of radioisotopically labelled γ^{33}P-ATP followed by phosphorimaging	976 Peptides (mostly 11-mers) derived from human phosphorylation/consensus sites including 12 control sequences in tripicates	DIPP	(68)
Cell lysates from fresh or two-week-old in vitro grown *Arabidopsis* plants and snap-frozen fresh material or frozen cell lysate in the presence of radioisotopically labelled γ^{33}P-ATP followed by phosphorimaging	192 Potential kinase substrates (8- or 9-mers)	DIPP	(69)
Lysates from either mouse lung capillary endothelial cells or from mouse adult fibroblasts infected with either empty or recombinant adenovirus encoding Cre recombinase to remove the gene for big mitogen-activated kinase 1 (BMK1)	1164 Peptides derived from human phosphorylation/consensus sites including 12 control sequences in duplicates	DIPP	(70)

(continued)

Table 2.1(continued)

Screening molecule	Library	Technology	Reference
Kidney lysates from control rats, Ren2 rats, a model of angiotensin II-mediated hypertensive renal damage and Ren2 rat treated with an ACE inhibitor to reduce renal fibrosis; readout using radioisotopically labelled γ^{33}P-ATP followed by phosphorimaging	1164 Peptides derived from human phosphorylation/consensus sites including 12 control sequences in duplicates	DIPP	(152)
Purified kinases PKA, PKC, PKG, Erk1,CK1, c-Src, Abl and phosphatases 2A and alkaline phosphatase or lysates from K562 cells overexpressing oncogenic Bcr-Abl fusion protein, readout via MALDI MS imaging	9 Different peptide substrates including three phosphopeptides equipped with N-terminal cysteine residue	DIPP	(113)
S. cerevisiae kinase Dbf1 in complex with activator Mob1 in the presence of radioisotopically labelled γ^{32}P-ATP followed by phosphorimaging	2296 Peptides (15-mer) derived from annotated human phosphorylation sites	DIPP	(71)
GST-tagged murine Plk4 or catalytically inactive mutant in the presence of radioisotopically labelled γ^{32}P-ATP followed by phosphorimaging	Substitutional analyses of substrate peptides MSYYHHHHHH, RKKKSFYFKKHHH, KKKCTIYGKICHF, GHRQTFLDEFDYA and GTKKSFFYKEVFE plus follow-up substrates combining found optimal residues and subsequent substitutional analyses of improved substrates resulting in total of 1240 cellulose-bound peptides	SPOT	(138)
Protein kinase A and biotin-ATP *(adenosine 5́-triphosphate γ-biotinyl-3,6,9-trioxa-undecanediamine)* followed by avidin-stabilized gold nanoparticles and signal amplification by silver deposition, detection by resonance light scattering of nanoparticles	Kemptide derivative LRRASLG and negative control peptide LRRAGLG chemoselectively attached via N-terminal amino function	DIPP	(111)
PKA and substrates entrapped by sol-gel encapsulation onto epoxy-modified glass surfaces, readout by fluorescence imaging	Kemptide derivative LRRASLG and controls	Diff	(102)

(continued)

Table 2.1(continued)

Screening molecule	Library	Technology	Reference
of bound fluorescent phosphospecific stain Pro-Q Diamond			
Recombinant kinases c-Src kinase, Raf, KDR, cMet, Flt3, Lyn, EGFR, PDGFR and Tie2 in the presence of members of a small compound library, readout by fluorescence imaging subsequent to incubation with phosphospecific antibodies followed by AlexaFluor555-labelled secondary anti-rabbit or anti-mouse antibodies.	Biotinylated peptide substrates like poly-$(Glu_4\text{-}Tyr)_{10}$	DIPP & SPNS	(153)
Recombinant proviral integration site 1 (Pim1) kinase together with substrate peptide in the presence or absence of specific inhibitor hb1217 was dispensed acoustically into hanging droplets of radioisotopically labelled γ^{32}P-ATP in a 10% glycerol/buffer solution, readout by phosphorimaging subsequent to transfer of droplets to phosphocellulose paper	S6 kinase/RSK2 substrate peptide 2 (KKRNRTLTK)	SPNS	(154)
Recombinant c-Src kinase or PKA, readout via fluorescence imaging subsequent to incubation with Cy5-labelled anti-phospho-tyrosine (c-Src) or anti-phosphoGFAP antibody (PKA)	Either biotinylated or poly(desoxythymidin)-modified substrate peptides IYGEFKKK (c-Src substrate) and RRRVTSAARRS (PKA substrate) together with phosphorylated forms (positive control) and with negative controls (Tyr/Phe (for c-Scr) or Ser/Ala (for PKA) substitutions)	DIPP	(45)
More than 30 different recombinant human kinases including PDK1, HMK1, PKC, PKA, Grk2, Grk5, LRRK2, Abl and CKII in the presence of radioisotopically labelled γ^{33}P-ATP followed by phosphorimaging (for the Isomerase-Chip) or human	Either overlapping peptides representing all human peptidyl-prolyl-*cis/trans*-isomerases (isomerase-chip; 3280 peptides) or overlapping peptides representing complete proteome of human cytomegalovirus (CMV-Chip; 17,181 peptides)	DIPP	(64)

(continued)

Table 2.1(continued)

Screening molecule	Library	Technology	Reference
kinases CKII; Abl or Grk5 and fluorescence imaging subsequent to treatment with phosphospecific stain Pro-Q Diamond (CMV-Chip)			
Recombinant protein tyrosine kinases Abl, Her2 and VEGFR2, readout using fluorescence imaging subsequent to hybridization to DNA microarray, treatment with monoclonal anti-phospho-tyrosine antibody PY-20 followed by Cy3-labelled anti-mouse IgG antibody	10,000 PNA-encoded split and mix peptide library of following structure: fluoresceine-Ttds-Phe-Gln-Xaa4-Xaa3-Tyr-Xaa2-Xaa1-Ile-Lys-Ttds-Lys(PNA)-amide with fluoresceine = normalization fluorophore; Ttds = PEG-based-linker moiety, Xaa1, Xaa2 and Xaa4 = Ile,Val, Phe, Pro, Arg, Glu, Lys, Ser, D-Pro, D-Val and Xaa3 = Ile, Val, Ala, Pro, Arg, Glu, Lys, D-Ala, Ser)	DCSD	(57)
Phospho-tyrosine phosphatases PTP1B and PTPμ; readout by fluorescence imaging subsequent to Cy5-labelled anti-phospho-tyrosine antibody	One tetra-, one hexa- and 46 penta-phosphopeptide derivative(s) covering different amino acid residues in different positions relative to the phospho-tyrosine residue were immobilized chemoselectively via C-terminal azide residue	DIPP	(114)
GST-tagged PTP-1B followed by mixture of anti-phospho-tyrosine antibodies (4G10 [1:2000], PY20 [1:1000], PY69 [1:1000] and P-Tyr-100 [1:2000]) and ECL detection	Insulin receptor derived phosphopeptides (Z = phospho-tyrosine residue) MTRDIZETDYYRKGG, MTRDIYETDZYRKGG, MTRDIYETDYZRKGG and non-phosphorylated control peptide MTRDIYETDYYRKGG	SPOT	(115)
Alkaline phosphatase, lambda phosphatase or Ser/Thr phosphatases PP1, PP2B, PP2A and PP2A in combination with of human peptidyl-prolyl-*cis/trans*-isomerase Pin1 followed by staining with fluorescent phosphospecific stain Pro-Q diamond (signal decrease after dephosphorylation)	87 Biotinylated, putative peptide substrates of Ser/Thr phosphatases each containing 11 residues with a centrally located phosphoSer/Thr-Pro moiety which is flanked by four or five residues N- or C-terminally, respectively. Sequences were derived from known substrates of human peptidyl-prolyl-*cis/trans*-isomerase Pin1	DIPP	(61)
Proteases			

(continued)

Table 2.1(continued)

Screening molecule	Library	Technology	Reference
Chymotrysin, papain	Combinatorial dipeptide libraries (400 spots) for substrate identification	SPOT	(118)
Trypsin	Substitution libraries GGR*B*G and GGK*B*G (***B*** = 20 amino acids)	SPOT	(126)
Trypsin	Internally quenched libraries of the type XXBXX and XXB_1B_2XX for substrate identification	SPOT	(124)
Porcine pancreatic elastase	Substitutional analyses and truncation analyses of a peptide derived from the third domain of turkey ovomucoid inhibitor (inhibitor identification)	SPOT	(155)
Trypsin; *Escherichia coli* outer membrane protease T *(OmpT)*	Internally quenched substitution analog libraries of known substrates (substrate specificity mapping)	SPOT	(125)
Porcine pancreatic elastase	Combinatorial library of all 400 dipeptide combinations (N-terminally labelled with FITC) from the genetically encoded amino acids	SPOT	(156)
Caspase-3	Complete L-amino acid substitutional analysis of the caspase-3 substrate Abz-VDQMDGW (Abz = amino benzoic acid)	SPOT	(121)
Caspase-3	Substitutional analysis of a known substrate; protein sequence derived, combinatorial, positional scanning and randomly generated peptide libraries (data not shown) for substrate identification	SPOT	(122)
Cathepsin C and cathepsin L	Peptidic cathepsin inhibitors tethered to fluorescence-labelled peptide nucleic acids	DCSD	(131)
Elastase, cathepsin G	Partial peptide scan and substitution analogs of the transferrin receptor stalk	SPOT	(123)
Caspase 3, granzyme B, Jurkat cell lysates or granzyme B activated apoptotic Jurkat cell lysates	Peptide derivatives tethered to fluoresceine-labelled peptide nucleic acids	DCSD	(130)

(continued)

Table 2.1(continued)

Screening molecule	Library	Technology	Reference
Lysyl-endopeptidase (LEP), chymotrypsin, V8 protease; BSA and ConA as negative controls; TLCK and Boc-Glu as inhibitors	3 Substrate peptides with LEP, chymotrypsin, or V8 protease specificity	Diff	(129)
Serine proteases (protein C, plasma kallikrein, factor VIIa, factor Ixaß, factor XIa, factor alphaXIIa, complement-C1s, -C1r, -D, tryptase, trypsin, subtilisin Calsberg, cathepsin G) and papain-like cysteine proteases (cathepsin B, H, K, L, S, V, rhodesain, papain, chymopapain, ficin, stem bromelain) followed by fluorescence scanning using CCD-based imager	361 Individual peptide derivatives of general structure Ac-Ala-P3-P2-Arg/Lys-fluorophore with fluorophore = 7-amino-4-carbamoylmethylcoumarin and P3/P2 = all proteinogenic amino acids except Cys	SPNS	(157)
Purified human proteases factor Xa, thrombin, plasmin, urokinase plasminogen activator and bovine or salmon thrombin followed by fluorescence scanning	361 Individual peptide derivatives of general structure Ac-Ala-P3-P2-Arg/Lys-fluorophore with fluorophore = 7-amino-4-carbamoylmethylcoumarin and P3/P2 = all proteinogenic amino acids except Cys	SPNS	(158)
Recombinant human caspase 3 followed by either two-dimensional surface plasmon resonance or fluorescent readout	Gluthationylated gold chip treated with caspase 3 substrate sequence DEVD flanked by GST and EGFP	DIPP	(159)
10 different plasma proteases (kallikrein, factors XIIa, XIa, IXa, VIIa, Xa, thrombin, activated protein C, plasmin and µPA)	361 Individual peptide derivatives of general structure Ac-Ala-P3-P2-Arg/Lys-fluorophore with fluorophore = 7-amino-4-carbamoylmethylcoumarin and P3/P2 = all proteinogenic amino acids except Cys	SPNS	(160)
Recombinant caspases 2, 4 and 6 and human thrombin, human plasmin, alpha-chymotrypsin from bovine pancreas; readout using fluorescence microscope equipped with either 4́6-diamino-2-phenylindole-, FITC- and TRITC-epifluorescence cubes and CCD camera	Caspase substrates Ac-YVAD-Mca, Ac-VDVAD-Mca, Ac-DEVD-Mca, Ac-VEID-Mca, Ac-IETD-Mca, Ac-LEHD-Mca in the presence of respective peptide aldehyde inhibitors or of 352 different compounds as potential inhibitors	SPNS	(161)

(continued)

Table 2.1(continued)

Screening molecule	Library	Technology	Reference
	Boc-VPR-Mca, (Z-FR)2-R110 and different quenched BODIPY-labelled casein substrates for thrombin, human plasmin and chymotrypsin, respectively		
Chymopapain and subtilisin followed by dual colour fluorescence imaging subsequent to hybridization to DNA microarray	Peptide nucleic acid (PNA) encoded split and mix FRET based peptide library of following structure: TAMRA-ßAla-Ser-Xaa4-Xaa3-Xaa2-Xaa1-Ala-Lys(FAM)-Ttds-Lys(PNA)-amide with TAMRA = quencher/normalization fluorophor; FAM = fluorophor, Ttds = PEG-based-linker moiety, Xaa = random positions with amino acid residues Ala, Asp, Phe, Lys, Leu, Asn, Pro, Ser, Val, Tyr	DCSD	(56)
Purified DegP protease (HtrA, protease Do) from *E. coli*, readout of fluorescence in aliquots of supernatant caused by aminobenzoic carboxamide (Abz) moiety of proteolytic fragments released from cellulose-bound Abz-labelled peptides	Overlapping peptides representing complete P pillin subunit in form of either 7-meric peptides overlapping by four amino acid residues or 12-meric peptides overlapping by nine residues	SPOT	(144)
Caspases 3, 6 and 8 or lysates from staurosporin stimulated CHO cells; readout using surface plasmon resonance imaging subsequent to treatment with protease followed by streptavidin	Caspase substrates DEVD, VEID, IETD and negative controls EVEE, VIEE, TIEE, respectively, flanked by cysteinyl-PEG-linker N-terminally and by GGSK(biotinyl)-amide C-terminally	DIPP	(31)
Purified caspase 3 or lysates of CHO cells treated with staurosporin followed by fluorescence imaging	N-terminally cysteine-tagged caspase 3 substrate ADEVDA and negative control AEVEEA flanked by Cy3 fluorophore and QSY-quencher moiety. Positive control without quencher included	DIPP	(38)
Trypsin or human glandular kallikrein 2; readout of fluorescence in aliquots of supernatant caused by aminobenzoic carboxamide (Abz) moiety of proteolytic	Cleavage sites derived from human seminal proteins including alanine scans, substitutions of key residues and incorporation of D-amino acids	SPOT	(120)

(continued)

Table 2.1(continued)

Screening molecule	Library	Technology	Reference
fragments released from cellulose-bound Abz-labelled peptides			
Recombinant and purified soluble ADAM8 protease containing pro- and metalloprotease domain in the absence and presence of EDTA, readout of fluorescence in aliquots of supernatant caused by aminobenzoic carboxamide (Abz) moiety of proteolytic fragments released from cellulose-bound Abz-labelled peptides	Eight peptides (10-mer) derived from myelin basic protein cleavage site YGSLP-QKAQG	SPOT	(119)
α-Chymotrypsin followed by fluorescence imaging subsequent to treatment with Cy5-labelled streptavidin	Biotinylated peptides of general structure biotin-GA-P_1-G-linker with P1= A, F, G, I, L, P, V, W	SPOT	(9)
Caspase-3, thrombin, plasmin or lysates from Jurkat cell treated with or without fas-activating antibody CH-11 or frozen plasma samples from pooled normal individuals and individuals on warfarin therapy; readout by fluorescence imaging subsequent to hybridization to DNA microarray	192 Member PNA-encoded rhodamine peptide conjugate library of general structure (Ac-Xaa1-Xaa2-Xaa3-Xaa4)2-rhodamine derivative-Lys(PEG-spacer-PNA-tag)-amide with Xaa1= Asp,Arg,Leu; Xaa2= Phe,Val,Thr,Pro, Xaa3= Asp,Arg,Thr,Pro and Xaa4= Asp,Arg,Nle,Pro	DCSD	(133)
Trypsin, thrombin and granzyme B followed by fluorescence imaging	Thrombin substrates Ac-Leu-Gly-Pro-Lys-ACC-linker, Ac-Nle-Thr-Pro-Lys-ACC-linker and granzyme B substrate Ac-Ile-Glu-Pro-Asp-ACC-linker (with ACC= 7-amino-4-carbamoylmethyl coumarin and Nle=norleucine) or 361-member library of general structure Ac-Ala-Xaa1-Xaa2-Lys-ACC-linker with Xaa1=Xaa2= all proteinogenic anino acid residues without cysteine)	DIPP	(128)
Recombinant proteases trypsin, chymotrypsin, endoproteinase AspN, endoproteinase Lys-C,	More than 1000 DNA encoded biotinylated peptides derivatives with penta histidine tag,	DCSD	(127)

(continued)

Table 2.1(continued)

Screening molecule	Library	Technology	Reference
endoproteinase Glu-C, endoproteinase Arg-C, thrombin, factor Xa, HRV1, different caspases and matrix metallo proteinases, enterokinase and tobacco etch virus protease, or biological samples like cell extracts from apoptotic Jurkat cells or Jurkat cell lines overexpressing oncogen Bcl2	sequences represent either cleavage sites taken from data base MEROPS or synthetical libraries generated by randomization of residues critical within consensus sequences for known proteases		
Diverse Enzymes			
Protein lysine methyltransferase G9a and radioisotopically labelled [methyl-^{3}H]-*S*-adenosyl-L-methionine	Substitutional analysis of N-terminal 21-meric peptide derived from human histone H3 including derivatives mimicking post-translationally modified variants like Arg8 methylation and Thr3, Thr6, Ser10, or Thr11 phosphorylation	SPOT	(87)
Histone lysine methyltransferase Dim-5 and D209A, D209S, E227A, E227S mutants and radioisotopically labelled [methyl-^{3}H]-*S*-adenosyl-L-methionine	Substitutional analysis of N-terminal 21-meric peptide derived from human histone H3	SPOT	(88)
Isoform T2 of UDP-*N*-acetyl-α-D-galactosamine:polypeptide:*N*-acetylgalactosaminyltransferase family	10 Peptides derived from known substrates of O-GlcNAc transferase including truncated versions and threonine/serine replacements chemoselectively immobilized via amino group of lysine side chain	DIPP	(162)
Pertussis toxin (*Bordetella pertussis*) and ATP + ^{32}P-NAD-ADP ribosylation	5- to 25-mer C-terminal peptides of the G-Protein α-subunits of G_{i3}, G_i, G_s, G_{o1}, G_{o2}, G_{oX1}, T_{rod}, G_z, $G_{q/11}$ and G_h, partial substitutional analysis of the 16-mer C-terminal peptide of the G_{i3} α-subunit	SPOT	(93)
Trigger factor (TF) (*Escherichia coli*) and TF-fragments	Peptide scans of *E. coli* proteins: EF-Tu (elongation factor Tu), MetE (methionine biosynthesis	SPOT	(163)

(continued)

Table 2.1(continued)

Screening molecule	Library	Technology	Reference
	enzyme), ICDH (isocitrate dehydrogenase) GlnRS (glutamine-tRNA-synthetase), alkaline phosphatase, β-galactosidase, FtsZ (involved in cell division), GBP (galactose binding protein), L2 (a ribosomal protein), lambda cI, pro OmpA (pro-outer membrane protein A), sigma 32, SecA (involved in secretion and murine DHFR (dehydrofolate reductase), yeast cytochrome B2, F1β- and Su9-ATPase subunits, *Photinus pyralis* luciferase, RNaseT1 from *Aspergillus oryzae* and human PrP (prion protein)		
Human peptidyl-prolyl-*cis/trans*-isomerase Pin1, sequential electro-blotting to PVDF-membranes, incubation of PVDV-membranes with anti-Pin1 antibody followed by secondary antibody	Peptide scan of human Cdc25C including phosphorylated S/T-P moieties	SPOT	(89)
Catalytic domain of human peptidyl-prolyl-*cis/trans*-isomerase Pin1, sequential electro-blotting to PVDF-membranes, incubation of PVDV-membranes with anti-Pin1 antibody followed by secondary antibody	1000 Phospho-threonine-containing (pThr) pentapeptides of general structure Ac-Xaa-pThr-Yaa-Zaa-Gln(linker)-COOH with Xaa and Zaa = 10 non-proteinogenic aromatic amino acids and Zaa = 10 proline derivatives resulting in 10 × 10 × 10 = 1000 individual phosphopeptide derivatives	SPOT	(90)
Readout using anti-SUMO-1 antibody subsequent to incubation with enzymes E1 and E2 necessary for efficient SUMO transfer in the presence of SUMO-1. Optical density was measured using CCD camera system after treatment with biotinylated anti-mouse IgG followed by HRP labelled	10 RanGAP1 derived peptides (9–15mers containing SUMO consensus sequence LKSE and derivatives with Lys to Ala exchange) and 15 peptides derived from known SUMOylation sites	DIPP & Diff	(96)

(continued)

Table 2.1(continued)

Screening molecule	Library	Technology	Reference
streptavidin and 2,2́-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid-dependent formation of green colour			
Trigger factor and DnaK from *E. coli*	Peptides derived from trigger factor and DnaK substrate proteins	SPOT	(164)
Peptide detection by MALDI-TOF MS – Concanavalin A (lectin binding) – β-1,4-Galactosyltransferase – galactosidase	Ac-IYAAPKKKC-NH2 – α-Mannose – *N*-acetylglucosamine – *N*-acetyllactosamine	DIPP	(165)
Rhodococcus rhodochrous epoxide hydrolase, *Electrophorus electricus* acetylcholine esterase, bovine pancreas trypsin, bovine intestinal mucosa alkaline phosphatase	5 substrates for the 4 hydrolases among them being 2 peptidic substrates for trypsin and caspases	DIPP	(94)

SPOT = spot synthesis of peptides
LDPS = light directed peptide synthesis
DIPP = directed immobilization of pre-synthesized peptides
DCSD = DNA/PNA chips as sorting device
SPNS = solution phase and nano spray
SODA = synthesis on defined areas
CLPP = cross-linking of pre-synthesized peptides
Diff. = different technologies

Alternatively, phospho-peptides can be detected using either phosphospecific antibodies (*see* **Fig. 2.4B**,C) or phosphospecific chelators (**Fig. 2.4D**,E) that are labelled with a detection moiety. The detection moiety can be a fluorescent label such as fluorescein (27, 30, 40) or an enzyme, for example horseradish peroxidase generating a chemiluminescent signal in combination with an appropriate substrate (78). The detection moiety can be coupled either directly to the anti-phosphoamino acid antibody (*see* **Fig. 2.4B**) (15, 27, 41) or to a secondary antibody (*see* **Fig. 2.4B**) (43, 78). Quality of anti-phosphoamino acid antibodies as detection tools was compared to the radioactive detection method generally recognized as the golden standard with respect to reliability. Here, the different detection procedures were applied to peptide microarrays on glass slides with 694 peptides derived from annotated phosphorylation sites from human

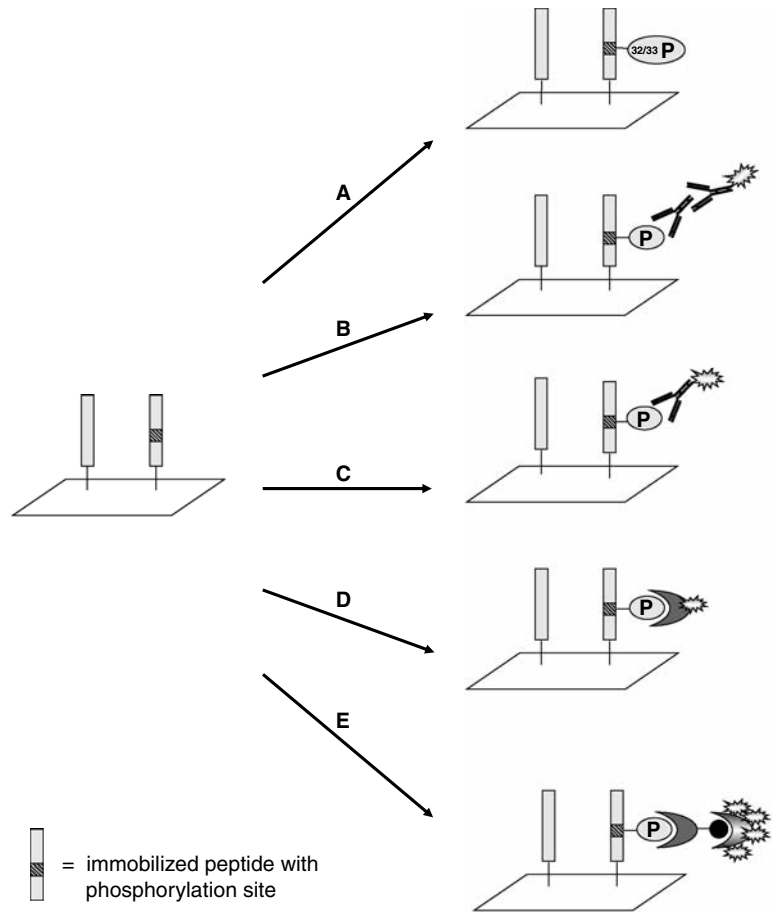


Fig. 2.4. Assay Principles for the Detection of Peptide Phosphorylation on Peptide Arrays. (**A**) The array is incubated with the kinase of interest in the presence of [γ-^{32}P or γ-^{33}P]-ATP and detection is performed by autoradiography. Alternatively, phosphorylation is measured with a generic anti-phosphoamino acid antibody in combination with a labelled secondary antibody (**B**) or with a directly fluorescently or chemoluminescently labelled anti-phosphoamino acid antibody (**C**). (**D**) Phospho-peptide detection using fluorescently labelled phosphoamino acid chelator. (**E**) Phospho-peptide detection using biotinylated phosphoamino acid chelator followed by fluorescently labelled streptavidin.

proteins together with all their possible 2234 monophosphorylated derivatives (27). Monoclonal anti-phospho-tyrosine antibodies only showed reliable results with no detectable binding to non-phosphorylated amino acids. Additionally, such antibodies have a limited false-negative rate and low cross-reactivity to phospho-serine or phospho-threonine residues. Anti-phospho-serine antibodies, however, had an extremely high false-negative rate, while anti-phospho-threonine antibodies showed significant cross-reactivities for peptides with phospho-tyrosine. Similar results were described using peptide microarrays displaying more than 6000 phospho-tyrosine peptides derived from human phosphorylation sites in triplicates

(77). In conclusion, the detection of phospho-peptides using antibodies seems to be limited to the detection of phosphotyrosine residues, at least until suitable anti-phospho-serine and anti-phospho-threonine antibodies are available.

Additionally, phosphoamino acid chelators coupled to a detection moiety can be used (*see* **Fig. 2.4D**). Martin et al. described the Pro-Q Diamond phospho-sensor dye which recognizes phospho-peptides with remarkably little cross-reactivity and a low false-negative rate (100). It was demonstrated that this dye could be applied to kinase profiling (77, 101, 102) and phosphatase profiling (61) using high-density peptide microarrays. A very similar approach was reported using biotinylated zinc(II) chelate phosphate sensor *N*-(5-(2-(+)-biotin aminoethylcarbamoyl)pyridine-2-ylmethyl)-*N*,*N*́,*N*̀-tris(pyridine-2-ylmethyl)-1,3-diaminopropan-2-ol (Phos-tag biotin, *see* **Fig. 2.4E**) (32–34, 37).

Finally, phosphopeptides could be transformed either enzymatically or chemically with high selectivity into labelled (phospho)peptides. Shults et al. reported carbodiimide-mediated, selective formation of a covalent bond between a fluorescent dye derivative and the phosphate moiety of phosphopeptides generated by incubation with kinases (103). Alternatively, Akita et al. used ß-elimination to transformed cellulose-surface bound phosphopeptides generated by protein kinase A selectively into dehydroalanine-containing peptides which could be labelled with fluorescence dye derivatives (104). Enzymatic transformation of the phosphate moiety could be performed using derivatives of ATP substituted at the γ-phosphate residue (105, 106). Kerman and Kraatz used γ-thio-derivative of ATP to detect transfer of thio-phosphate to surface-bound peptides electrochemically using gold nanoparticles (107). It could be demonstrated that electrochemical detection of kinase-mediated transfer of modified phosphate residues is more efficient if electro-active adenosine-5́-[γ-ferrocene] triphosphate is used as co-substrate for protein kinase C (108). ATP derivatives biotinylated at the γ-phosphate residue were used to detect PKA-mediated generation of surface-bound biotinylated phosphopeptides using avidin-stabilized gold nanoparticles amplified by silver deposition (109). Use of biotinylated ATP derivatives in kinase assays seems to be an interesting alternative to existing technologies (110).

In principle, many other detection principles such as surface plasmon resonance and mass spectrometry, which are comprehensively described in a recent review (19), are possible. Examples have been described for the detection of phosphorylation at peptide microarrays using resonance light scattering (109, 111), surface plasmon resonance (32, 33, 37, 43) and MALDI-TOF mass spectrometry (112, 113).

Generally, each assay suitable for kinase profiling on peptide (micro)arrays could be used for analysis of phosphatase-mediated release of phosphate moieties from phosphopeptides, too. Starting surface-bound phosphopeptides could be generated either enzymatically (78) or chemically (61, 78, 114–117). Detection of phosphatase action on (micro)array bound phosphopeptides by signal decrease subsequent to treatment with anti-phospho-tyrosine antibodies (78, 115, 116) or using phosphospecific dyes (61) was demonstrated.

Several different assay principles have been developed to measure protease activity on peptide arrays. Optimal assays lead to increased signal intensity upon substrate cleavage either by reading generated signal in released proteolytic fragment (*see* **Fig. 2.5**) or in still surface-bound peptide fragment (*see* **Fig. 2.6**). Several assays of this type have been described: (1) the first assay was developed by Duan and Laursen and is based upon peptide arrays prepared on polyaminoethylmethacrylamide membranes by the SPOT method (*see* **Fig. 2.5A**) (118). The array comprised all

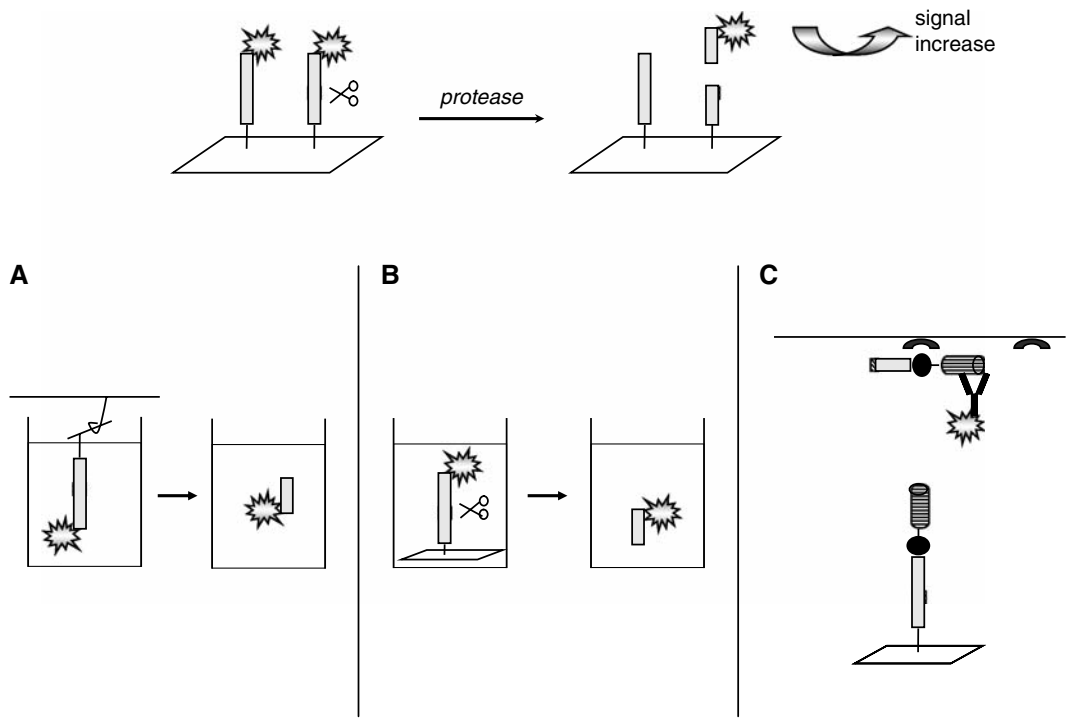


Fig. 2.5. Detection of Proteolytic Cleavage on Peptide Arrays Reading Released Fragment. (**A**) N-terminally labelled immobilized peptides in the 96-well plate format, cleavage yields fluorescently labelled fragment released into the well of the microtitre plate; membrane disc with bound peptide fragment is released before fluorescence imaging (118). (**B**) Alternative assay with N-terminally labelled immobilized peptides in the 96-well plate format, membrane disc with bound peptide or peptide fragment will stay in the well of the microtitre plate but released fragment is separated from membrane by transfer of aliquots of reaction solution from supernatant to other microtitre plate which yields fluorescence signals for cleaved peptides after fluorescence imaging (119, 120, 123, 144), and (**C**) indirect detection of peptide cleavage subsequent to electrotransfer of the released N-terminal peptide fragment containing antibody epitope (126).

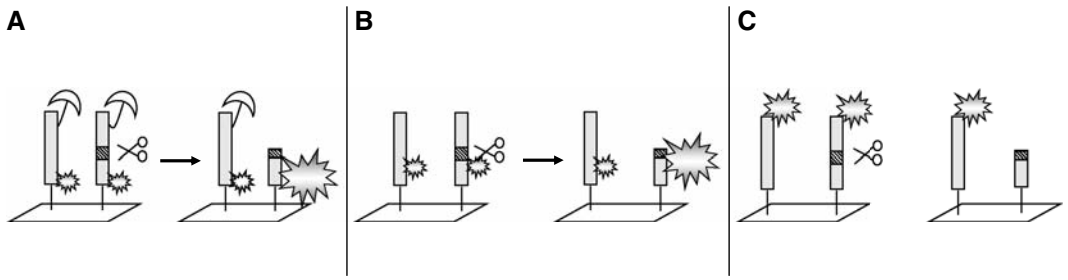


Fig. 2.6. Detection of Proteolytic Cleavage on Peptide Arrays Reading Surface-Bound Fragment. (**A**) Internally fluorescence quenched peptides on arrays, subsequent to cleavage peptide fragment containing quenching moiety is washed away yielding fluorescence increase for cleaved peptides (56, 123, 125). (**B**) Peptides having a fluorogenic group C-terminal to the scissile bond which increases fluorescence subsequent to cleavage (128, 133), and (**C**) peptides with a fluorescence dye at the free terminus for array-based assays with decreasing fluorescence signal upon cleavage.

400 possible dipeptides derived from genetically encoded amino acids with an N-terminally coupled fluoresceinyl thiocarbamyl moiety. These peptides were punched out and attached to pins in a microtitre plate lid. Subsequently, they were suspended in wells of a 96-well microtitre plate filled with protease solution. After specified reaction times, the spots were removed in order to quantify the fluorescence dye coupled to the cleaved-off N-terminal peptide fragment. This method was evaluated using chymotrypsin and papain. (2) To avoid laborious pin attachment, a modified assay involves immersing substrate spots (amino benzoic acid as fluorescence dye) in wells filled with the protease solution (**Fig. 2.5B**) (119, 120). At various times small aliquots are pipetted into new wells and cleavage is quantified using a fluorescence microtitre plate reader. This assay was employed to identify and characterize caspase-3 substrates using substitutional analyses of a known peptide substrate, a peptide scan, combinatorial libraries and randomly generated sets of peptides (121–123). The major disadvantage of these two assay principles is that the peptide array has to be dissected, essentially abandoning the benefits of array technologies. (3) This led to the introduction of peptide arrays with internally quenched peptides (*see* **Fig. 2.6A**). Compartmentalization of the cleavage reaction is not necessary and increasing signal intensity is observed. This technique was evaluated using combinatorial peptide libraries and substitutional analyses of substrate peptides incubated with trypsin (124, 125) and subsequently employed to determine the substrate specificity of the integral membrane protease OmpT of *Escherichia coli* (125). (4) A sophisticated but rather tedious procedure involves peptides coupled to cellulose membranes by their C-terminus and having an antibody epitope tag with a biotinylated lysine residue at the N-terminus (*see* **Fig. 2.5C**). Cleavage releases the N-terminal part of a substrate peptide including the epitope tag and the biotin moiety. This fragment is affinity-blotted onto a streptavidin-coated PVDF membrane and detected via an enzyme-conjugated

antibody (126). A similar procedure was used by Kozlov et al. in combination with DNA-encoded substrates and DNA microarrays as sorting device. Potential cleavage site peptides flanked by biotin on one side and penta histidine tag/DNA tag on the other side were incubated with different proteases or cell lysates. Streptavidin-coated magnetic beads were used to remove non-cleaved peptides and biotinylated cleavage fragments. Remaining members of the library represent His-tagged DNA-encoded fragments of cleaved substrates only which could be detected and deconvoluted using anti-penta histidine antibody and fluorescence imaging subsequent to hybridization onto appropriate DNA microarrays (127). (5) Peptide derivatives containing a substituted fluorogenic group C-terminal to the scissile bond are immobilized on glass slides resulting in peptide microarrays (*see* **Fig. 2.6B**). In a proof-of-concept study it was demonstrated that the protease trypsin cleaved the amino acyl–fluorophore bond (94). A very similar assay principle using longer peptides successfully determined the substrate specificities of trypsin, granzyme B and thrombin employing peptide microarrays on glass slides generated by chemoselective peptide immobilization (128). (6) Peptides with a fluorescence dye at the free terminus are applied for array-based protease assays with a decreasing signal upon cleavage (*see* **Fig. 2.6C**).

Recently, a novel principle was described for profiling proteolytic activities using semi-wet peptide microarrays and differences in the partition coefficients of peptide substrates and released fluorophores (129). Lysyl endopeptidase treatment released an environmentally sensitive fluorophore resulting in a blue shift of the emission maximum from 540 to 508 nm, along with two-fold higher fluorescence intensity.

In a very elegant experiment, peptidic inhibitors tethered to fluorescence-tagged peptide nucleic acids (PNA) were used to profile inhibitor specificity against different cysteine proteases (130, 131). The peptide nucleic acid tag encodes the structure of the attached peptide derivative and, therefore, allows spatially addressed deconvolution after hybridization to Affymetrix GenFlex oligonucleotide microarray. This approach was extended to the profiling of enzymatic activities of proteases (56, 127, 132, 133) and kinases (57, 103) using PNA/DNA-encoding. Use of oligonucleotides and in particular peptide nucleic acids (PNAs) to encode libraries was reviewed comprehensively (134).

5. Substrate Identification

Different scenarios for the identification of kinase substrates are possible. Combinatorial and randomly generated libraries can be

applied if no information about potential protein substrates is available. Pioneering work in this field was carried out using low-density peptide arrays on cellulose membranes. In this format combinatorial libraries were used to identify substrates for PKA (*see* **Fig. 2.2B**) (58, 82–84, 97), PKG (81, 83, 97) and the budding yeast kinase CDC15 (82). This approach was also successful using peptide microarrays for p60c-src (41). A randomly generated library of 1433 tyrosine-containing single peptides on a peptide microarray was used to identify new substrates for c-Abl (28).

While these approaches are suitable for identifying kinase substrate peptides de novo, a demanding question in biology is the identification of in vivo protein substrates for kinases in order to integrate novel kinases into their biological context and signal transduction networks. The data generated from combinatorial or random libraries are of limited use for answering such questions. The resulting substrate sequences are usually not found in nature and natural substrates can only be deduced by looking for similar naturally occurring sequences. Knowledge-based libraries are used to overcome this problem.

In cases where a protein substrate of a certain kinase is known, and the aim is to identify the actual site of phosphorylation within the target protein, two straightforward approaches are to use a selection of peptides containing the potential phospho-acceptor residues of the target protein, or a peptide scan of the target protein. For example, three peptides could be identified as substrates for Lyn kinase by using libraries of 15-mer peptides generated from the sequence around each tyrosine residue in PKC (62). Decapeptide sequences derived from the cytoplasmic domains of C-CAM revealed a single specific phosphorylation site for PKC (63). Overlapping peptide scans were used to determine PKA phosphorylation sites in myelin basic protein (**Fig. 2.2**), and the autophosphorylation sites as well as sites for CK2-mediated phosphorylation in the tyrosine kinase Tie2 (29).

Even though initial proof-of-concept experiments with 18 cellulose membrane bound peptides derived from protein sequences phosphorylated by PKC in vivo (67), plus later studies, were successful (135), the full power of the knowledge-based approach emerged when applied to high-density, high-content peptide microarrays.

Peptide microarrays displaying the sequences of 710 human annotated phosphorylation sites revealed peptide substrates for NEK6 (26), Abl (27) and PKA/CK2/GSK3 (29). More advanced libraries were used for PDK1, HMK and Plk1 (*see* **Fig. 2.7**) or CK2 (11,096 peptides from cytoplasmic domains of human membrane proteins and 2304 human annotated phosphorylation sites) (27, 71), PDK1 (1394 peptides derived from the activation loops of human kinases) (29) and CK2 or

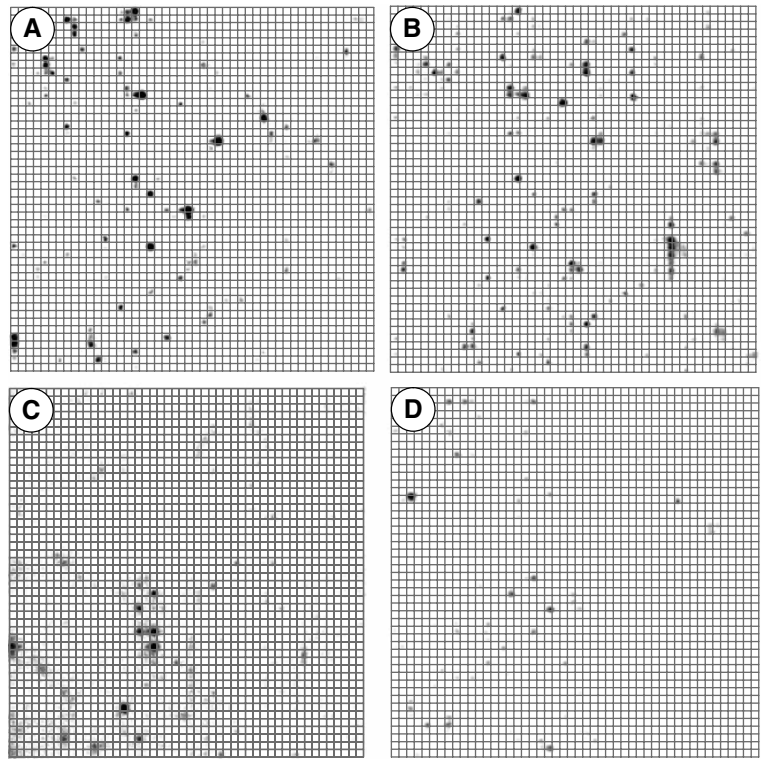


Fig. 2.7. Profiling of Four Different Kinases. Peptide microarrays display 2304 annotated human phosphorylation site derived peptides immobilized chemoselectively via N-terminal end in triplicates. Phosphorimages of one subarray obtained after incubation with the respective kinase in the presence of [γ-^{32}P]-ATP are shown. (**A**) PDK1 in the presence of 50 μM activator PIFtide. (**B**) Hydrophobic motif kinase (HMK). (**C**) Proline-directed kinase KIS. (**D**) Polo-like kinase 1.

Gsk3 (17,181 peptides representing overlapping peptides covering complete proteome of cytomegalovirus) (64). Peptide substrates identified in the microarray experiments were superior to known peptide substrates for PDK1 and NEK6, as demonstrated by determination of catalytic constants in solution phase experiments (26, 29).

An additional application of these microarrays is the detection of priming phosphorylation events. In such processes substrates for certain kinases are generated upon previous phosphorylation with another kinase on different phosphoacceptor amino acids of the substrate. This was shown for the system CK2 as priming kinase and GSK3 as second kinase with a library of 694 annotated human phosphorylation-site peptides where all corresponding CK2 monophosphorylated derivatives were produced by incorporating phosphoamino acid building blocks during synthesis (29).

6. Substrate Optimization

From the beginning, the use of peptide arrays in kinase research focused on kinase substrate optimization in terms of substrate efficiency and selectivity. Using cAMP- and cGMP-dependent protein kinases (PKA and PKG) as model enzymes, Tegge et al. (83) applied peptide arrays on cellulose membranes to identify substrates from combinatorial libraries with the format Ac-XXXO_1O_2XXX. X represents mixtures of all 20 proteinogenic amino acids, while O_1 and O_2 represent individual amino acids defined for one spot but varying between different spots in a library. Applying all 20 naturally encoded amino acids at these defined positions will result in $20 \times 20 = 400$ spots or peptide mixtures. After a first screening round of this initial library, the best two amino acids at positions 1 and 2 are retained throughout the optimization cycles and two new positions are defined. This procedure is iteratively repeated until each position is refined and one single peptide sequence per spot is obtained. This strategy led to the identification of a new, very efficient peptide substrate for PKG, and selected a PKA substrate with properties very similar to the known kemptide peptide (136). Extending this approach to 12- and 14-meric peptides yielded substrates highly specific for PKG (80). Substituting the phospho-acceptor residue by alanine resulted in specific inhibitory peptides. Analysis of these results revealed a central role for PKG in the modulation of vascular contractility (84, 81).

Different combinatorial libraries led to the deconvolution of substrate sequences for PKA and type I and II TGF-α receptor kinases using porous polyethylene discs as the solid support (58).

Toomik and Ek (67) used the SPOT technology to synthesize an optimization library for PKC substrates, with the flanking residues of a known substrate substituted by different amino acids, leading to specific and efficient PKC substrates. Similar experiments led to optimized substrate sequences for the calcium-dependent kinase from maize seedlings (86).

A different approach that proved to be very powerful for mapping antibody epitopes is a complete substitutional analysis (137) (for review *see* (19)). For example, substitutional analysis of a histone H3-derived peptide on cellulose membranes led to the discovery of DYRKtide, which is a very efficient peptide substrate for DYRK1a (85). These strategies applied to peptide microarrays were also used to determine the substrate requirements of NEK6 (26) and p60c-src (41). Moreover, iterative substitutional analyses lead to optimized substrates for murine Plk4 (138).

However, when a good database of substrates and non-substrates is available from peptide array experiments the generation of weight matrices is a valuable alternative for kinase specificity analysis.

Weight matrices are a representation of the probability of each amino acid occurring at a certain position relative to the phospho-acceptor residue based on statistical evaluation of peptide microarray data. Phosphorylation of a library comprising 1433 randomly generated peptides by Abl resulted in a weight matrix that was successfully used to predict bona fide kinase substrates (28).

7. Miscellaneous

A new high-throughput tool connecting solution-phase kinase activity assays with immobilized format analysis via biotin–streptavidin interaction and phosphorimaging was introduced by Panse et al. (27). Following casein kinase II reactions in 384-well microtitre plates in the presence of [γ-^{32}P]-ATP, aliquots of reaction solution were transferred to a streptavidin-coated membrane to create a peptide array composed of 720 different 13-mer peptides derived from human phosphorylation sites. Biotinylated substrate peptides were radioactively labelled due to the incorporated phosphate moiety and could be easily detected by phosphorimaging of the membrane after washing steps (27). A similar approach was used in combination with positional scanning libraries for deciphering substrate specificity of plant protein kinases (139), *S. cerevisiae* kinases Ime2 and Cdk1 (140) and eight different human protein serine/threonine kinases (141).

Zhu et al. described comprehensive analysis of yeast kinases using elastomer sheets with imprinted microwells mounted onto microscope slides (142). The microwells were loaded with 17 different known substrate proteins using the cross-linker 3-glycidoxypropyltrimethylsilane and incubated with 119 different yeast kinase-GST fusion proteins in the presence of radioisotope labelled ATP. Subsequent to phosphorimaging, it could be demonstrated that this technology permits the identification of novel kinase activities. Generally, each protein microarray could be used for the analysis of kinase activities. Nevertheless, the correct alignment of identified phosphorylation events to a given peptide sequence is impaired by the number of potential phospho-acceptor residues within a protein. Additionally, signal quantification is difficult due to possible multiple phosphorylations within one protein.

Another approach is the immobilization of kinase substrates on the bottom of a microtitre plate well, perfectly combining both formats. Electrochemiluminescence detection can be used for signal readout, subsequent to incubation of the patterned peptide array inside the well with kinases and appropriately labelled phosphospecific antibodies.

An extension of the microarray approach is the determination of K_i values directly on microarrays. Houseman et al. demonstrated efficient concentration-dependent inhibition of c-Src activity for c-Src inhibitors quercetin, tyrphostin and PP1 by applying different kinase/inhibitor mixtures to a substrate-coated slide under a layer of mineral oil. Droplets formed due to the oil layer reaction, allowing their spatial resolution (43). Additionally, the authors were able to demonstrate that their peptide microarrays on monolayers of alkanethiolates self-assembled on gold are fully compatible with surface plasmon resonance spectroscopy. This MALDI MS imaging was extended to six different single kinases or two phosphatases, kinase mixtures and K562 cell lysates in the absence and presence of inhibitors in combination with microfluidic devices (113).

References

1. Fodor, S. P., J. L. Read, M. C. Pirrung, L. Stryer, A. T. Lu & D. Solas (1991) Light-directed, spatially addressable parallel chemical synthesis. Science, 251, 767–73.
2. Pellois, J. P., X. Zhou, O. Srivannavit, T. Zhou, E. Gulari & X. Gao (2002) Individually addressable parallel peptide synthesis on microchips. Nat Biotechnol, 20, 922–6.
3. LeProust, E., J. P. Pellois, P. Yu, H. Zhang, X. Gao, O. Srivannavit, E. Gulari & X. Zhou (2000) Digital light-directed synthesis. A microarray platform that permits rapid reaction optimization on a combinatorial basis. J Comb Chem, 2, 349–54.
4. Komolpis, K., O. Srivannavit & E. Gulari (2002) Light-directed simultaneous synthesis of oligopeptides on microarray substrate using a photogenerated acid. Biotechnol Prog, 18, 641–6.
5. Frank, R., S. Güler, S. Krause & W. Lindenmaier (1991). Peptides 1990 proceedings of the 21st European peptide symposium (Eds. Giralt, E., Andreu, D.) Escom Leiden, 151–152.
6. Frank, R. (1992) Spot-synthesis: an easy technique for the positionally adressable, parallel chemical synthesis on a membrane support. Tetrahedron, 48, 9217–9232.
7. Frank, R. (1995) Simultaneous and combinatorial chemical synthesis techniques for the generation and screening of molecular diversity. J Biotechnol, 41, 259–72.
8. Wenschuh, H., R. Volkmer-Engert, M. Schmidt, M. Schulz, J. Schneider-Mergener & U. Reineke (2000) Coherent membrane supports for parallel microsynthesis and screening of bioactive peptides. Biopolymers, 55, 188–206.
9. Kim, D. H., D. S. Shin & Y. S. Lee (2007) Spot arrays on modified glass surfaces for efficient SPOT synthesis and on-chip bioassay of peptides. J Pept Sci, 13, 625–33.
10. Frank, R. & H. Overwin (1996) SPOT synthesis. Epitope analysis with arrays of synthetic peptides prepared on cellulose membranes. Methods Mol Biol, 66, 149–69.
11. Kramer, A. & J. Schneider-Mergener (1998) Synthesis and screening of peptide libraries on continuous cellulose membrane supports. Methods Mol Biol, 87, 25–39.
12. Kramer, A., U. Reineke, L. Dong, B. Hoffmann, U. Hoffmuller, D. Winkler, R. Volkmer-Engert & J. Schneider-Mergener (1999b) Spot synthesis: observations and optimizations. J Pept Res, 54, 319–27.
13. Reineke, U., A. Kramer & J. Schneider-Mergener (2001a) Epitope mapping with synthetic peptides prepared by SPOT synthesis. (Eds. Konterman, R., Dübel, S.) Antibody Engineering (Springer Lab Manual). Berlin: Springer Verlag, 443–459.
14. Reineke, U., R. Volkmer-Engert & J. Schneider-Mergener (2001c) Applications of peptide arrays prepared by the SPOT-technology. Curr Opin Biotechnol, 12, 59–64.
15. Reimer, U., U. Reineke & J. Schneider-Mergener (2002) Peptide arrays: from macro to micro. Curr Opin Biotechnol, 13, 315–20.
16. Frank, R. (2002b) The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports – principles and applications. J Immunol Methods, 267, 13–26.

17. Frank, R. & J. Schneider-Mergener (2002) SPOT-Synthesis – Scope and Applications. Peptide Arrays on Membrane Supports: Synthesis and Application, Koch, J., Mahler, M., 1–22.
18. Frank, R. (2002a) High-density synthetic peptide microarrays: emerging tools for functional genomics and proteomics. Comb Chem High Throughput Screen, 5, 429–40.
19. Reineke, U., J. Schneider-Mergener & M. Schutkowski (2004) Peptide arrays in proteomics and drug discovery. Micro/Nano technology for genomics and proteomics (Eds. Ozkan, M., Heller, M.) Volume 2,161–282.
20. Hilpert, K., D. F. Winkler & R. E. Hancock (2007) Peptide arrays on cellulose support: SPOT synthesis, a time and cost efficient method for synthesis of large numbers of peptides in a parallel and addressable fashion. Nat Protoc, 2, 1333–49.
21. Beyer, M., T. Felgenhauer, F. Ralf Bischoff, F. Breitling & V. Stadler (2006) A novel glass slide-based peptide array support with high functionality resisting non-specific protein adsorption. Biomaterials, 27, 3505–14.
22. Beyer, M., A. Nesterov, I. Block, K. Konig, T. Felgenhauer, S. Fernandez, K. Leibe, G. Torralba, M. Hausmann, U. Trunk, V. Lindenstruth, F. R. Bischoff, V. Stadler & F. Breitling (2007) Combinatorial synthesis of peptide arrays onto a microchip. Science, 318, 1888.
23. Antohe, B. V. & P. W. Cooley (2007) In situ synthesis of peptide microarrays using inkjet microdispensing. Methods Mol Biol, 381, 299–312.
24. Adler, G., R. Türk, N. Frank, W. Zander, W. Wu, J. Volkmer-Engert, J. Schneider-Mergener & H. Gausepohl (1999). Proceedings of the international symposium on innovation and perspectives in solid phase synthesis (Ed. Epton, R.) Mayflower Worldwide 221–222.
25. Falsey, J. R., M. Renil, S. Park, S. Li & K. S. Lam (2001) Peptide and small molecule microarray for high throughput cell adhesion and functional assays. Bioconjug Chem, 12, 346–53.
26. Lizcano, J. M., M. Deak, N. Morrice, A. Kieloch, C. J. Hastie, L. Dong, M. Schutkowski, U. Reimer & D. R. Alessi (2002) Molecular basis for the substrate specificity of NIMA-related kinase-6 (NEK6). Evidence that NEK6 does not phosphorylate the hydrophobic motif of ribosomal S6 protein kinase and serum- and glucocorticoid-induced protein kinase in vivo. J Biol Chem, 277, 27839–49.
27. Panse, S., L. Dong, A. Burian, R. Carus, M. Schutkowski, U. Reimer & J. Schneider-Mergener (2004) Profiling of generic anti-phosphopeptide antibodies and kinases with peptide microarrays using radioactive and fluorescence-based assays. Mol Divers, 8, 291–9.
28. Rychlewski, L., M. Kschischo, L. Dong, M. Schutkowski & U. Reimer (2004) Target specificity analysis of the Abl kinase using peptide microarray data. J Mol Biol, 336, 307–11.
29. Schutkowski, M., U. Reimer, S. Panse, L. Dong, J. M. Lizcano & D. R. Alessi (2004) High-content peptide microarrays for deciphering kinase specificity and biology. Angew Chem, 116, 2725–2728.
30. Lesaicherre, M. L., M. Uttamchandani, G. Y. Chen & S. Q. Yao (2002) Developing site-specific immobilization strategies of peptides in a microarray. Bioorg Med Chem Lett, 12, 2079–83.
31. Inoue, Y., T. Mori, G. Yamanouchi, X. Han, T. Sonoda, T. Niidome & Y. Katayama (2008) Surface plasmon resonance imaging measurements of caspase reactions on peptide microarrays. Anal Biochem, 375, 147–9.
32. Inamori, K., M. Kyo, K. Matsukawa, Y. Inoue, T. Sonoda, K. Tatematsu, K. Tanizawa, T. Mori & Y. Katayama (2008) Optimal surface chemistry for peptide immobilization in on-chip phosphorylation analysis. Anal Chem, 80, 643–650.
33. Inamori, K., M. Kyo, Y. Nishiya, Y. Inoue, T. Sonoda, E. Kinoshita, T. Koike & Y. Katayama (2005) Detection and quantification of on-chip phosphorylated peptides by surface plasmon resonance imaging techniques using a phosphate capture molecule. Anal Chem, 77, 3979–85.
34. Shigaki, S., T. Yamaji, X. Han, G. Yamanouchi, T. Sonoda, O. Okitsu, T. Mori, T. Niidome & Y. Katayama (2007) A peptide microarray for the detection of protein kinase activity in cell lysate. Anal Sci, 23, 271–5.
35. Lemeer, S., C. Jopling, F. Naji, R. Ruijtenbeek, M. Slijper, A. J. Heck & J. den Hertog (2007) Protein–tyrosine kinase activity profiling in knock down zebrafish embryos. PLoS One, 2, e581.
36. Lemeer, S., R. Ruijtenbeek, M. W. Pinkse, C. Jopling, A. J. Heck, J. den Hertog & M. Slijper (2007) Endogenous phosphotyrosine signaling in zebrafish embryos. Mol Cell Proteomics, 6, 2088–99.
37. Mori, T., K. Inamori, Y. Inoue, X. Han, G. Yamanouchi, T. Niidome & Y. Katayama

(2007) Evaluation of protein kinase activities of cell lysates using peptide microarrays based on surface plasmon resonance imaging. Anal Biochem, 375, 223–31.

38. Han, A., T. Sonoda, J. H. Kang, M. Murata, N. I. T & Y. Katayam (2006) Development of a fluorescence peptide chip for the detection of caspase activity. Comb Chem High Throughput Screen, 9, 21–5.
39. Dawson, P. E., T. W. Muir, I. Clark-Lewis & S. B. Kent (1994) Synthesis of proteins by native chemical ligation. Science, 266, 776–9.
40. Lesaicherre, M. L., M. Uttamchandani, G. Y. Chen & S. Q. Yao (2002) Antibody-based fluorescence detection of kinase activity on a peptide array. Bioorg Med Chem Lett, 12, 2085–8.
41. Uttamchandani, M., E. W. Chan, G. Y. Chen & S. Q. Yao (2003) Combinatorial peptide microarrays for the rapid determination of kinase specificity. Bioorg Med Chem Lett, 13, 2997–3000.
42. Uttamchandani, M., G. Y. Chen, M. L. Lesaicherre & S. Q. Yao (2004) Site-specific peptide immobilization strategies for the rapid detection of kinase activity on microarrays. Methods Mol Biol, 264, 191–204.
43. Houseman, B. T., J. H. Huh, S. J. Kron & M. Mrksich (2002) Peptide chips for the quantitative evaluation of protein kinase activity. Nat Biotechnol, 20, 270–4.
44. Kohn, M., R. Wacker, C. Peters, H. Schroder, L. Soulere, R. Breinbauer, C. M. Niemeyer & H. Waldmann (2003) Staudinger ligation: a new immobilization strategy for the preparation of small-molecule arrays. Angew Chem Int Ed Engl, 42, 5830–4.
45. Kimura, N., T. Okegawa, K. Yamazaki & K. Matsuoka (2007) Site-specific, covalent attachment of poly(dT)-modified peptides to solid surfaces for microarrays. Bioconjug Chem, 18, 1778–85.
46. Parker, L. L., S. B. Brueggemeier, W. J. Rhee, D. Wu, S. B. Kent, S. J. Kron & S. P. Palecek (2006) Photocleavable peptide hydrogel arrays for MALDI-TOF analysis of kinase activity. Analyst, 131, 1097–104.
47. Lynch, M., C. Mosher, J. Huff, S. Nettikadan, J. Johnson & E. Henderson (2004) Functional protein nanoarrays for biomarker profiling. Proteomics, 4, 1695–702.
48. Wiley, J. P., K. A. Hughes, R. J. Kaiser, E. A. Kesicki, K. P. Lund & M. L. Stolowitz (2001) Phenylboronic acid–salicylhydroxamic acid bioconjugates. 2. Polyvalent immobilization of protein ligands for affinity chromatography. Bioconjug Chem, 12, 240–50.
49. Stolowitz, M. L., C. Ahlem, K. A. Hughes, R. J. Kaiser, E. A. Kesicki, G. Li, K. P. Lund, S. M. Torkelson & J. P. Wiley (2001) Phenylboronic acid–salicylhydroxamic acid bioconjugates. 1. A novel boronic acid complex for protein immobilization. Bioconjug Chem, 12, 229–39.
50. Melnyk, O., X. Duburcq, C. Olivier, F. Urbes, C. Auriault & H. Gras-Masse (2002) Peptide arrays for highly sensitive and specific antibody-binding fluorescence assays. Bioconjug Chem, 13, 713–20.
51. Olivier, C., A. Perzyna, Y. Coffinier, B. Grandidier, D. Stievenard, O. Melnyk & J. O. Durand (2006) Detecting the chemoselective ligation of peptides to silicon with the use of cobalt-carbonyl labels. Langmuir, 22, 7059–65.
52. Carion, O., V. Souplet, C. Olivier, C. Maillet, N. Medard, O. El-Mahdi, J. O. Durand & O. Melnyk (2007) Chemical micropatterning of polycarbonate for site-specific peptide immobilization and biomolecular interactions. Chembiochem, 8, 315–22.
53. Coffinier, Y., S. Szunerits, C. Jama, R. Desmet, O. Melnyk, B. Marcus, L. Gengembre, E. Payen, D. Delabouglise & R. Boukherroub (2007) Peptide immobilization on amine-terminated boron-doped diamond surfaces. Langmuir, 23, 4494–7.
54. Coffinier, Y., C. Olivier, A. Perzyna, B. Grandidier, X. Wallart, J. O. Durand, O. Melnyk & D. Stievenard (2005) Semicarbazide-functionalized Si(111) surfaces for the site-specific immobilization of peptides. Langmuir, 21, 1489–96.
55. Wegner, G. J., H. J. Lee & R. M. Corn (2002) Characterization and optimization of peptide arrays for the study of epitope–antibody interactions using surface plasmon resonance imaging. Anal Chem, 74, 5161–8.
56. Diaz-Mochon, J. J., L. Bialy & M. Bradley (2006) Dual colour, microarray-based, analysis of 10,000 protease substrates. Chem Commun (Camb), 3984–6.
57. Pouchain, D., J. J. Diaz-Mochon, L. Bialy & M. Bradley (2007) A 10,000 member PNA-encoded peptide library for profiling tyrosine kinases. ACS Chem Biol, 2, 810–8.
58. Luo, K., P. Zhou & H. F. Lodish (1995) The specificity of the transforming growth factor beta receptor kinases determined by a spatially addressable peptide library. Proc Natl Acad Sci USA, 92, 11761–5.

59. MacBeath, G. & S. L. Schreiber (2000) Printing proteins as microarrays for high-throughput function determination. Science, 289, 1760–3.
60. Lee, S. J. & S. Y. Lee (2004) Microarrays of peptides elevated on the protein layer for efficient protein kinase assay. Anal Biochem, 330, 311–6.
61. Sun, H., C. H. Lu, M. Uttamchandani, Y. Xia, Y. C. Liou & S. Q. Yao (2008) Peptide Microarray for High-Throughput Determination of Phosphatase Specificity and Biology. Angew Chem Int Ed Engl, 47, 1698–1702.
62. Szallasi, Z., M. F. Denning, E. Y. Chang, J. Rivera, S. H. Yuspa, C. Lehel, Z. Olah, W. B. Anderson & P. M. Blumberg (1995) Development of a rapid approach to identification of tyrosine phosphorylation sites: application to PKC delta phosphorylated upon activation of the high affinity receptor for IgE in rat basophilic leukemia cells. Biochem Biophys Res Commun, 214, 888–94.
63. Edlund, M., K. Wikstrom, R. Toomik, P. Ek & B. Obrink (1998) Characterization of protein kinase C-mediated phosphorylation of the short cytoplasmic domain isoform of C-CAM. FEBS Lett, 425, 166–70.
64. Thiele, A., J. Zerweck, M. Weiwad, G. Fischer & M. Schutkowski (2009) High density peptide microarrays for reliable identification of phosphorylation sites and upstream kinases. Methods in molecular biology, Peptide Microarrays (Eds. Cretich, M. and Chiari, M).
65. Boeckmann, B., A. Bairoch, R. Apweiler, M. C. Blatter, A. Estreicher, E. Gasteiger, M. J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilbout & M. Schneider (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res, 31, 365–70.
66. Kreegipuu, A., N. Blom & S. Brunak (1999) PhosphoBase, a database of phosphorylation sites: release 2.0. Nucleic Acids Res, 27, 237–9.
67. Toomik, R. & P. Ek (1997) A potent and highly selective peptide substrate for protein kinase C assay. Biochem J, 322 (Pt 2), 455–60.
68. Stulemeijer, I. J., J. W. Stratmann & M. H. Joosten (2007) Tomato mitogen-activated protein kinases LeMPK1, LeMPK2, and LeMPK3 are activated during the Cf-4/Avr4-induced hypersensitive response and have distinct phosphorylation specificities. Plant Physiol, 144, 1481–94.
69. Ritsema, T., J. Joore, W. van Workum & C. M. Pieterse (2007) Kinome profiling of Arabidopsis using arrays of kinase consensus substrates. Plant Methods, 3, 3.
70. Hayashi, M., C. Fearns, B. Eliceiri, Y. Yang & J. D. Lee (2005) Big mitogen-activated protein kinase 1/extracellular signal-regulated kinase 5 signaling pathway is essential for tumor-associated angiogenesis. Cancer Res, 65, 7699–706.
71. Mah, A. S., A. E. Elia, G. Devgan, J. Ptacek, M. Schutkowski, M. Snyder, M. B. Yaffe & R. J. Deshaies (2005) Substrate specificity analysis of protein kinase complex Dbf2-Mob1 by peptide library and proteome array screening. BMC Biochem, 6, 22.
72. Diks, S. H., K. Parikh, M. van der Sijde, J. Joore, T. Ritsema & M. P. Peppelenbosch (2007) Evidence for a minimal eukaryotic phosphoproteome? PLoS One, 2, e777.
73. Diks, S. H., K. Kok, T. O'Toole, D. W. Hommes, P. van Dijken, J. Joore & M. P. Peppelenbosch (2004) Kinome profiling for studying lipopolysaccharide signal transduction in human peripheral blood mononuclear cells. J Biol Chem, 279, 49206–13.
74. van Baal, J. W., S. H. Diks, R. J. Wanders, A. M. Rygiel, F. Milano, J. Joore, J. J. Bergman, M. P. Peppelenbosch & K. K. Krishnadath (2006) Comparison of kinome profiles of Barrett's esophagus with normal squamous esophagus and normal gastric cardia. Cancer Res, 66, 11605–12.
75. Lowenberg, M., J. Tuynman, M. Scheffer, A. Verhaar, L. Vermeulen, S. van Deventer, D. Hommes & M. Peppelenbosch (2006) Kinome analysis reveals nongenomic glucocorticoid receptor-dependent inhibition of insulin signaling. Endocrinology, 147, 3555–62.
76. Lowenberg, M., J. Tuynman, J. Bilderbeek, T. Gaber, F. Buttgereit, S. van Deventer, M. Peppelenbosch & D. Hommes (2005) Rapid immunosuppressive effects of glucocorticoids mediated through Lck and Fyn. Blood, 106, 1703–10.
77. Zerweck, J., A. Masch & M. Schutkowski (2008) Peptide microarrays for profiling of modification state specific antibodies. Methods in molecular biology, epitope mapping protocols, 2nd edition (Eds. Reineke, R. and Schutkowski, M.) Chapter 12 Vol. 524, 169–80.
78. Espanel, X., S. Walchli, T. Ruckle, A. Harrenga, M. Huguenin-Reggiani & R. Hooft van Huijsduijnen (2003) Mapping of synergistic components of weakly interacting protein–protein motifs using arrays of paired peptides. J Biol Chem, 278, 15162–7.

79. Mukhija, S., L. Germeroth, J. Schneider-Mergener & B. Erni (1998) Identification of peptides inhibiting enzyme I of the bacterial phosphotransferase system using combinatorial cellulose-bound peptide libraries. Eur J Biochem, 254, 433–8.
80. Dostmann, W. R., C. Nickl, S. Thiel, I. Tsigelny, R. Frank & W. J. Tegge (1999) Delineation of selective cyclic GMP-dependent protein kinase Ialpha substrate and inhibitor peptides based on combinatorial peptide libraries on paper. Pharmacol Ther, 82, 373–87.
81. Dostmann, W. R., W. Tegge, R. Frank, C. K. Nickl, M. S. Taylor & J. E. Brayden (2002) Exploring the mechanisms of vascular smooth muscle tone with highly specific, membrane-permeable inhibitors of cyclic GMP-dependent protein kinase Ialpha. Pharmacol Ther, 93, 203–15.
82. Rodriguez, M., S. S. Li, J. W. Harper & Z. Songyang (2004) An oriented peptide array library (OPAL) strategy to study protein–protein interactions. J Biol Chem, 279, 8802–7.
83. Tegge, W., R. Frank, F. Hofmann & W. R. Dostmann (1995) Determination of cyclic nucleotide-dependent protein kinase substrate specificity by the use of peptide libraries on cellulose paper. Biochemistry, 34, 10569–77.
84. Dostmann, W. R., M. S. Taylor, C. K. Nickl, J. E. Brayden, R. Frank & W. J. Tegge (2000) Highly specific, membrane-permeant peptide blockers of cGMP-dependent protein kinase Ialpha inhibit NO-induced cerebral dilation. Proc Natl Acad Sci USA, 97, 14772–7.
85. Himpel, S., W. Tegge, R. Frank, S. Leder, H. G. Joost & W. Becker (2000) Specificity determinants of substrate recognition by the protein kinase DYRK1A. J Biol Chem, 275, 2431–8.
86. Loog, M., R. Toomik, K. Sak, G. Muszynska, J. Jarv & P. Ek (2000) Peptide phosphorylation by calcium-dependent protein kinase from maize seedlings. Eur J Biochem, 267, 337–43.
87. Rathert, P., A. Dhayalan, M. Murakami, X. Zhang, R. Tamas, R. Jurkowska, Y. Komatsu, Y. Shinkai, X. Cheng & A. Jeltsch (2008a) Protein lysine methyltransferase G9a acts on non-histone targets. Nat Chem Biol, 4, 344–6.
88. Rathert, P., X. Zhang, C. Freund, X. Cheng & A. Jeltsch (2008b) Analysis of the substrate specificity of the dim-5 histone lysine methyltransferase using Peptide arrays. Chem Biol, 15, 5–11.
89. Lu, P. J., X. Z. Zhou, M. Shen & K. P. Lu (1999) Function of WW domains as phosphoserine- or phosphothreonine-binding modules. Science, 283, 1325–8.
90. Wildemann, D., F. Erdmann, B. H. Alvarez, G. Stoller, X. Z. Zhou, J. Fanghanel, M. Schutkowski, K. P. Lu & G. Fischer (2006) Nanomolar inhibitors of the peptidyl prolyl cis/trans isomerase Pin1 from combinatorial peptide libraries. J Med Chem, 49, 2147–50.
91. Houseman, B. T. & M. Mrksich (1999) The Role of Ligand Density in the Enzymatic Glycosylation of Carbohydrates Presented on Self-Assembled Monolayers of Alkanethiolates on Gold. Angewandte Chemie Intl. Ed., 38, 782–5.
92. Fazio, F., M. C. Bryan, O. Blixt, J. C. Paulson & C. H. Wong (2002) Synthesis of sugar arrays in microtiter plate. J Am Chem Soc, 124, 14397–402.
93. von Olleschik-Elbheim, L., A. el Bayâ & M. A. Schmidt (1997). ADP-ribosylation inanimal tissue: Structure, Function and Biology of Mono(ADPriboryl)transferases and Related Enzymes. Plenum Press, New York 1997 (Ed. Haag, K.-N. and Koch-Nolte, F.), 87.
94. Zhu, Q., M. Uttamchandani, D. Li, M. L. Lesaicherre & S. Q. Yao (2003) Enzymatic profiling system in a small-molecule microarray. Org Lett, 5, 1257–60.
95. Zhu, H. & M. Snyder (2003) Protein chip technology. Curr Opin Chem Biol, 7, 55–63.
96. Schwamborn, K., P. Knipscheer, E. van Dijk, W. J. van Dijk, T. K. Sixma, R. H. Meloen & J. P. Langedijk (2008) SUMO assay with peptide arrays on solid support: insights into SUMO target sites. J Biochem, 144, 39–49.
97. Tegge, W. J. & R. Frank (1998) Analysis of protein kinase substrate specificity by the use of peptide libraries on cellulose paper (SPOT-method). Methods Mol Biol, 87, 99–106.
98. Bodem, J. & M. Blüthner (2002) Modification of Immobilized Peptides. Peptide Arrays on Membrane Supports: Synthesis and Application, Koch, J., Mahler, M., 141–151.
99. Collins, M. O., L. Yu, M. P. Coba, H. Husi, I. Campuzano, W. P. Blackstock, J. S. Choudhary & S. G. Grant (2005) Proteomic analysis of in vivo phosphorylated synaptic proteins. J Biol Chem, 280, 5972–82.
100. Martin, K., T. H. Steinberg, L. A. Cooley, K. R. Gee, J. M. Beechem & W. F. Patton (2003) Quantitative analysis of protein phosphorylation status and protein kinase activity on microarrays using a novel fluorescent phosphorylation sensor dye. Proteomics, 3, 1244–55.

101. Zhu, Q., A. Hong, N. Sheng, X. Zhang, A. Matejko, K. Y. Jun, O. Srivannavit, E. Gulari, X. Gao & X. Zhou (2007) microParaflo biochip for nucleic acid and protein analysis. Methods Mol Biol, 382, 287–312.
102. Rupcich, N., J. R. Green & J. D. Brennan (2005) Nanovolume kinase inhibition assay using a sol-gel-derived multicomponent microarray. Anal Chem, 77, 8013–9.
103. Shults, M. D., I. A. Kozlov, N. Nelson, B. G. Kermani, P. C. Melnyk, V. Shevchenko, A. Srinivasan, J. Musmacker, J. P. Hachmann, D. L. Barker, M. Lebl & C. Zhao (2007) A multiplexed protein kinase assay. Chembiochem, 8, 933–42.
104. Akita, S., N. Umezawa, N. Kato & T. Higuchi (2008) Array-based fluorescence assay for serine/threonine kinases using specific chemical reaction. Bioorg Med Chem, 16, 7788–94.
105. Elphick, L. M., S. E. Lee, V. Gouverneur & D. J. Mann (2007) Using chemical genetics and ATP analogues to dissect protein kinase function. ACS Chem Biol, 2, 299–314.
106. Allen, J. J., M. Li, C. S. Brinkworth, J. L. Paulson, D. Wang, A. Hubner, W. H. Chou, R. J. Davis, A. L. Burlingame, R. O. Messing, C. D. Katayama, S. M. Hedrick & K. M. Shokat (2007) A semisynthetic epitope for kinase substrates. Nat Methods, 4, 511–6.
107. Kerman, K. & H. B. Kraatz (2007) Electrochemical detection of kinase-catalyzed thiophosphorylation using gold nanoparticles. Chem Commun (Camb), 5019–21.
108. Song, H., K. Kerman & H. B. Kraatz (2008) Electrochemical detection of kinase-catalyzed phosphorylation using ferrocene-conjugated ATP. Chem Commun (Camb), 4, 502–4.
109. Sun, L., D. Liu & Z. Wang (2007) Microarray-based kinase inhibition assay by gold nanoparticle probes. Anal Chem, 79, 773–7.
110. Green, K. D. & M. K. Pflum (2007) Kinase-catalyzed biotinylation for phosphoprotein detection. J Am Chem Soc, 129, 10–1.
111. Wang, Z., J. Lee, A. R. Cossins & M. Brust (2005) Microarray-based detection of protein binding and functionality by gold nanoparticle probes. Anal Chem, 77, 5770–4.
112. Min, D. H., J. Su & M. Mrksich (2004) Profiling kinase activities by using a peptide chip and mass spectrometry. Angew Chem Int Ed Engl, 43, 5973–7.
113. Su, J., M. R. Bringer, R. F. Ismagilov & M. Mrksich (2005) Combining microfluidic networks and peptide arrays for multi-enzyme assays. J Am Chem Soc, 127, 7280–1.
114. Kohn, M., M. Gutierrez-Rodriguez, P. Jonkheijm, S. Wetzel, R. Wacker, H. Schroeder, H. Prinz, C. M. Niemeyer, R. Breinbauer, S. E. Szedlacsek & H. Waldmann (2007) A microarray strategy for mapping the substrate specificity of protein tyrosine phosphatase. Angew Chem Int Ed Engl, 46, 7700–3.
115. Espanel, X. & R. H. van Huijsduijnen (2005) Applying the SPOT peptide synthesis procedure to the study of protein tyrosine phosphatase substrate specificity: probing for the heavenly match in vitro. Methods, 35, 64–72.
116. Espanel, X., M. Huguenin-Reggiani & R. Hooft van Huijsduijnen (2002) The SPOT technique as a tool for studying protein tyrosine phosphatase substrate specificities. Protein Sci, 11, 2326–34.
117. Pasquali, C., M. L. Curchod, S. Walchli, X. Espanel, M. Guerrier, F. Arigoni, G. Strous & R. H. Van Huijsduijnen (2003) Identification of protein tyrosine phosphatases with specificity for the ligand-activated growth hormone receptor. Mol Endocrinol, 17, 2228–39.
118. Duan, Y. & R. A. Laursen (1994) Protease substrate specificity mapping using membrane-bound peptides. Anal Biochem, 216, 431–8.
119. Naus, S., S. Reipschlager, D. Wildeboer, S. F. Lichtenthaler, S. Mitterreiter, Z. Guan, M. L. Moss & J. W. Bartsch (2006) Identification of candidate substrates for ectodomain shedding by the metalloprotease-disintegrin ADAM8. Biol Chem, 387, 337–46.
120. Janssen, S., C. M. Jakobsen, D. M. Rosen, R. M. Ricklis, U. Reineke, S. B. Christensen, H. Lilja & S. R. Denmeade (2004) Screening a combinatorial peptide library to develop a human glandular kallikrein 2-activated prodrug as targeted therapy for prostate cancer. Mol Cancer Ther, 3, 1439–50.
121. Reineke, U., D. Kurzhals, A. Köhler, C. Blex, J. E. G. McCarthy, P. Li, L. Germeroth & J. Schneider-Mergener (2001b) High throughput screening assay for the identification of protease substrates. Peptides 2000: Proceedings of the Twenty-Sixth European Peptide Symposium (Ed. Martinez, J., Fehrentz, J. A.) Éditions EDK 721.
122. Reineke, U. & U. Hoffmüller (2001). American Biotechnology Laboratory, 50.
123. Kaup, M., K. Dassler, U. Reineke, C. Weise, R. Tauber & H. Fuchs (2002) Processing of the human transferrin receptor at distinct positions within the stalk region by neutrophil elastase and cathepsin G. Biol Chem, 383, 1011–20.

124. Reineke, U., S. Bhargava, M. Schutkowski, C. Landgraf, L. Germeroth, G. Fischer & J. Schneider-Mergener (1999). Peptides 1998: Proceedings of the Twenty-Fifth European Peptide Symposium (Ed. Bajusz, S., Hudecz, F.), Akadèmiai Kiadò, Budapest, 1999, p. 670–671.
125. Dekker, N., R. C. Cox, R. A. Kramer & M. R. Egmond (2001) Substrate specificity of the integral membrane protease OmpT determined by spatially addressed peptide libraries. Biochemistry, 40, 1694–701.
126. Kramer, A., R. Affelt & J. Schneider-Mergener (1999a). Peptides 1998: Proceedings of the Twenty-Fifth European Peptide Symposium (Ed. Bajusz, S., Hudecz, F.) Akadèmiai Kiadò, Budapest, 546.
127. Kozlov, I. A., P. C. Melnyk, J. P. Hachmann, A. Srinivasan, M. Shults, C. Zhao, J. Musmacker, N. Nelson, D. L. Barker & M. Lebl (2008) A high-complexity, multiplexed solution-phase assay for profiling protease activity on microarrays. Comb Chem High Throughput Screen, 11, 24–35.
128. Salisbury, C. M., D. J. Maly & J. A. Ellman (2002) Peptide microarrays for the determination of protease substrate specificity. J Am Chem Soc, 124, 14868–70.
129. Kiyonaka, S., K. Sada, I. Yoshimura, S. Shinkai, N. Kato & I. Hamachi (2004) Semi-wet peptide/protein array using supramolecular hydrogel. Nat Mater, 3, 58–64.
130. Winssinger, N., S. Ficarro, P. G. Schultz & J. L. Harris (2002) Profiling protein function with small molecule microarrays. Proc Natl Acad Sci USA, 99, 11139–44.
131. Winssinger, N., J. L. Harris, B. J. Backes & P. G. Schultz (2001). From split-pool libraries to spatially addressable microarrays and its application to functional proteomic profiling. Angew Chem, 113, 3254–3258.
132. Kozlov, I. A., P. C. Melnyk, J. P. Hachmann, D. L. Barker, M. Lebl & C. Zhao (2007) Evaluation of different chemical strategies for conjugation of oligonucleotides to peptides. Nucleos Nucleot Nucleic Acids, 26, 1353–7.
133. Winssinger, N., R. Damoiseaux, D. C. Tully, B. H. Geierstanger, K. Burdick & J. L. Harris (2004) PNA-encoded protease substrate microarrays. Chem Biol, 11, 1351–60.
134. Pianowski, Z. L. & N. Winssinger (2008) Nucleic acid encoding to program self-assembly in chemical biology. Chem Soc Rev, 37, 1330–6.
135. Moilanen, A. M., U. Karvonen, H. Poukka, O. A. Janne & J. J. Palvimo (1998) Activation of androgen receptor function by a novel nuclear protein kinase. Mol Biol Cell, 9, 2527–43.
136. Kemp, B. E., D. J. Graves, E. Benjamini & E. G. Krebs (1977) Role of multiple basic residues in determining the substrate specificity of cyclic AMP-dependent protein kinase. J Biol Chem, 252, 4888–94.
137. Pinilla, C., J. R. Appel & R. A. Houghten (1993) Functional importance of amino acid residues making up peptide antigenic determinants. Mol Immunol, 30, 577–85.
138. Leung, G. C., C. S. Ho, I. M. Blasutig, J. M. Murphy & F. Sicheri (2007) Determination of the Plk4/Sak consensus phosphorylation motif using peptide spots arrays. FEBS Lett, 581, 77–83.
139. Vlad, F., B. E. Turk, P. Peynot, J. Leung & S. Merlot (2008) A versatile strategy to define the phosphorylation preferences of plant protein kinases and screen for putative substrates. Plant J, 55, 104–17.
140. Holt, L. J., J. E. Hutti, L. C. Cantley & D. O. Morgan (2007) Evolution of Ime2 phosphorylation sites on Cdk1 substrates provides a mechanism to limit the effects of the phosphatase Cdc14 in meiosis. Mol Cell, 25, 689–702.
141. Hutti, J. E., E. T. Jarrell, J. D. Chang, D. W. Abbott, P. Storz, A. Toker, L. C. Cantley & B. E. Turk (2004) A rapid method for determining protein kinase phosphorylation specificity. Nat Methods, 1, 27–9.
142. Zhu, H., J. F. Klemic, S. Chang, P. Bertone, A. Casamayor, K. G. Klemic, D. Smith, M. Gerstein, M. A. Reed & M. Snyder (2000) Analysis of yeast protein kinases using protein chips. Nat Genet, 26, 283–9.
143. Kishimoto, A., K. Nishiyama, H. Nakanishi, Y. Uratsuji, H. Nomura, Y. Takeyama & Y. Nishizuka (1985) Studies on the phosphorylation of myelin basic protein by protein kinase C and adenosine 3':5'-monophosphate-dependent protein kinase. J Biol Chem, 260, 12492–9.
144. Jones, C. H., P. Dexter, A. K. Evans, C. Liu, S. J. Hultgren & D. E. Hruby (2002) *Escherichia coli* DegP protease cleaves between paired hydrophobic residues in a natural substrate: the PapA pilin. J Bacteriol, 184, 5762–71.
145. Frank, R., S. Hoffmann, M. Kiess, H. Lahmann, W. Tegge, C. Behn & H. Gausepohl (1996) Combinatoral peptide and nonpeptide libraries (Ed. Jung, G.). VCH Weinheim 363–386.
146. Toomik, R., M. Edlund, P. Ek, B. Obrink & L. Engstrom (1996) Simultaneously

synthesized peptides on continuous cellulose membranes as substrates for protein kinases. Pept Res, 9, 6–11.

147. Houseman, B. T. & M. Mrksich (2002) Towards quantitative assays with peptide chips: a surface engineering approach. Trends Biotechnol, 20, 279–81.
148. Buss, H., A. Dorrie, M. L. Schmitz, R. Frank, M. Livingstone, K. Resch & M. Kracht (2004) Phosphorylation of serine 468 by GSK-3beta negatively regulates basal p65 NF-kappaB activity. J Biol Chem, 279, 49571–4.
149. Li, Y., D. M. Keller, J. D. Scott & H. Lu (2005) CK2 phosphorylates SSRP1 and inhibits its DNA-binding activity. J Biol Chem, 280, 11869–75.
150. de la Fuente van Bentem, S., D. Anrather, I. Dohnal, E. Roitinger, E. Csaszar, J. Joore, J. Buijnink, A. Carreri, C. Forzani, Z. J. Lorkovic, A. Barta, D. Lecourieux, A. Verhounig, C. Jonak & H. Hirt (2008) Site-specific phosphorylation profiling of arabidopsis proteins by mass spectrometry and peptide chip analysis. J Proteome Res, 7, 2458–2470.
151. Wang, H. & D. L. Brautigan (2006) Peptide microarray analysis of substrate specificity of the transmembrane Ser/Thr kinase KPI-2 reveals reactivity with cystic fibrosis transmembrane conductance regulator and phosphorylase. Mol Cell Proteomics, 5, 2124–30.
152. de Borst, M. H., S. H. Diks, J. Bolbrinker, M. W. Schellings, M. B. van Dalen, M. P. Peppelenbosch, R. Kreutz, Y. M. Pinto, G. Navis & H. van Goor (2007) Profiling of the renal kinome: a novel tool to identify protein kinases involved in angiotensin II-dependent hypertensive renal damage. Am J Physiol Renal Physiol, 293, F428–37.
153. Horiuchi, K. Y., Y. Wang, S. L. Diamond & H. Ma (2006) Microarrays for the functional analysis of the chemical-kinase interactome. J Biomol Screen, 11, 48–56.
154. Wong, E. Y. & S. L. Diamond (2008) Enzyme microarrays assembled by acoustic dispensing technology. Anal Biochem, 81, 509–14.
155. Hilpert, K., G. Hansen, H. Wessner, J. Schneider-Mergener & W. Hohne (2000) Characterizing and optimizing protease/peptide inhibitor interactions, a new application for spot synthesis. J Biochem, 128, 1051–7.
156. Laursen, R. A., C. Zhu & Y. Duan (2001). High-Throughput Organic Synthesis (Ed.Sucholeiki, I.), Marcel Dekker, New York, 113–9.
157. Gosalia, D. N., C. M. Salisbury, J. A. Ellman & S. L. Diamond (2005a) High throughput substrate specificity profiling of serine and cysteine proteases using solution-phase fluorogenic peptide microarrays. Mol Cell Proteomics, 4, 626–36.
158. Gosalia, D. N., C. M. Salisbury, D. J. Maly, J. A. Ellman & S. L. Diamond (2005b) Profiling serine protease substrate specificity with solution phase fluorogenic peptide microarrays. Proteomics, 5, 1292–8.
159. Park, K., J. Ahn, S. Y. Yi, M. Kim & B. H. Chung (2008) SPR imaging-based monitoring of caspase-3 activation. Biochem Biophys Res Commun, 368, 684–9.
160. Gosalia, D. N., W. S. Denney, C. M. Salisbury, J. A. Ellman & S. L. Diamond (2006) Functional phenotyping of human plasma using a 361-fluorogenic substrate biosensing microarray. Biotechnol Bioeng, 94, 1099–110.
161. Gosalia, D. N. & S. L. Diamond (2003) Printing chemical libraries on microarrays for fluid phase nanoliter reactions. Proc Natl Acad Sci USA, 100, 8721–6.
162. Laurent, N., J. Voglmeir, A. Wright, J. Blackburn, N. T. Pham, S. C. Wong, S. J. Gaskell & S. L. Flitsch (2008) Enzymatic glycosylation of peptide arrays on gold surfaces. Chembiochem, 9, 883–7.
163. Patzelt, H., S. Rudiger, D. Brehmer, G. Kramer, S. Vorderwulbecke, E. Schaffitzel, A. Waitz, T. Hesterkamp, L. Dong, J. Schneider-Mergener, B. Bukau & E. Deuerling (2001) Binding specificity of *Escherichia coli* trigger factor. Proc Natl Acad Sci USA, 98, 14244–9.
164. Deuerling, E., H. Patzelt, S. Vorderwulbecke, T. Rauch, G. Kramer, E. Schaffitzel, A. Mogk, A. Schulze-Specking, H. Langen & B. Bukau (2003) Trigger Factor and DnaK possess overlapping substrate pools and binding specificities. Mol Microbiol, 47, 1317–28.
165. Su, J. & M. Mrksich (2002). Using Mass Spectrometry to Characterize Self-Assembled Monolayers Presenting Peptides, Proteins, and Carbohydrates. Angew Chem internatialn edition, 41, 4715–8.

Chapter 3

Using Peptide Array to Identify Binding Motifs and Interaction Networks for Modular Domains

Shawn S.-C. Li and Chenggang Wu

Abstract

Specific protein–protein interactions underlie all essential biological processes and form the basis of cellular signal transduction. The recognition of a short, linear peptide sequence in one protein by a modular domain in another represents a common theme of macromolecular recognition in cells, and the importance of this mode of protein–protein interaction is highlighted by the large number of peptide-binding domains encoded by the human genome. This phenomenon also provides a unique opportunity to identify protein–protein binding events using peptide arrays and complementary biochemical assays. Accordingly, high-density peptide array has emerged as a useful tool by which to map domain-mediated protein–protein interaction networks at the proteome level. Using the Src-homology 2 (SH2) and 3 (SH3) domains as examples, we describe the application of oriented peptide array libraries in uncovering specific motifs recognized by an SH2 domain and the use of high-density peptide arrays in identifying interaction networks mediated by the SH3 domain. Methods reviewed here could also be applied to other modular domains, including catalytic domains, that recognize linear peptide sequences.

Key words: Peptide arrays, oriented peptide array library (or OPAL), SPOT, celluspots, domains, src-homology 2 (or SH2), src-homology 3 (SH3), specificity, linear motif, protein–protein interaction (or PPI), interaction network.

1. Introduction

Since it was first introduced by Merrifield (1) in 1963, solid-phase peptide synthesis (SPPS) has become the method of choice for peptide synthesis. In contrast to the conventional approach which involves the synthesis, purification of intermediates and the assembly of fragments into the final product in solution, SPPS allows for a peptide chain to "grow" on a solid support through cycles of decoupling, neutralization and coupling that are usually carried out in a fully automated fashion. This yields a peptide of

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_3,

predetermined sequence with purity comparable or superior to that produced by a solution-based method. More importantly, SPPS makes it feasible to carry out parallel synthesis of multiple peptides in a single run, a versatility conferred only by the solid-phase approach. Although the same chemistry used in a conventional synthesis is employed in a parallel peptide synthesis, the latter features increased throughput ability and is therefore compatible to contemporary proteomic strategies. Parallel peptide synthesis can be carried out either on resins or on functionalized cellulose membranes. Current technology allows for hundreds of peptides to be synthesized on resins from a single run, while thousands of peptides can be prepared on a nitrocellulose membrane. Peptides produced from either the resin- or the membrane-based approach can be respotted or printed onto a glass slide in low density (as in macroarray) or in high density (as in microarray). More recently, oriented peptide libraries and the corresponding sub-libraries were synthesized on nitrocellulose membranes or on soluble membranes for reprinting. The possibility of synthesizing multiple peptides in an efficient and cost-effective manner and the concurrent development of approaches for peptide array production have opened up new avenues for the application of peptides in biomedical and pharmaceutical research. For example, peptides synthesized in a microarray format have become a powerful tool for high throughput screening of antibody epitopes (2, 3) and for determining protein binding sites (4). Arrayed peptide libraries have been employed to uncover peptide motifs selectively bound by modular interaction domains or recognized by protein kinases or phosphatases (5, 6). As often dictated by the particular application, peptide arrays may be either synthesized in situ on a cellulose membrane or prepared by immobilization of pre-synthesized peptides to a solid support.

2. In Situ Synthesis of Peptide Arrays

The in situ parallel synthesis approach has the primary benefits of consuming small amount of reagents and eliminating the steps associated with purification and subsequent immobilization of the peptides. Since peptides are directly synthesized on a functionalized cellulose membrane, the purity of the peptides on an array is undefined. However, for short peptides (<20 residues), the purity is usually greater than 80%, sufficient for most functional assays. Assuming a 99% coupling efficiency for each residue in a peptide, the purity of the final product would fall under 80% once

the peptide length reaches 23 residues. Peptide purity may also become a real concern when it contains one or more Cys and Met residues or multiple hydrophobic residues in a row. A variety of methods have been developed for in situ synthesis of peptide arrays, of which the SPOT approach has found the widest use.

The SPOT method was described first by Frank et al. in 1992 (7). An SPOT peptide array is synthesized in a stepwise manner by sequentially spotting small volumes (sub-microlitres to microlitres) of activated amino acids to a flat solid substrate such as a functionalized cellulose or a polypropylene membrane or a glass slide. The standard 9-fluorenylmethoxycarbonyl (Fmoc)-based peptide chemistry is used in conjunction with the *t*-butyloxycarbonyl (Boc)-based sidechain protection strategy to allow sidechain deprotection under mild acidic conditions. The general steps for SPOT synthesis and their application in binding assays are illustrated in **Fig. 3.1**. These include the following: (a) the selection of a suitable solid support that meets the chemical and biological requirements of the application; (b) the functionalization of the solid support (often a cellulose membrane) for attachment of the first activated amino acid. A linker or spacer sequence is usually inserted between the membrane support and the first residue to allow maximal accessibility and conformational flexibility of the peptide; (c) the growth of the peptide chain by performing cycles of coupling and deprotection; (d) the deprotection of sidechains

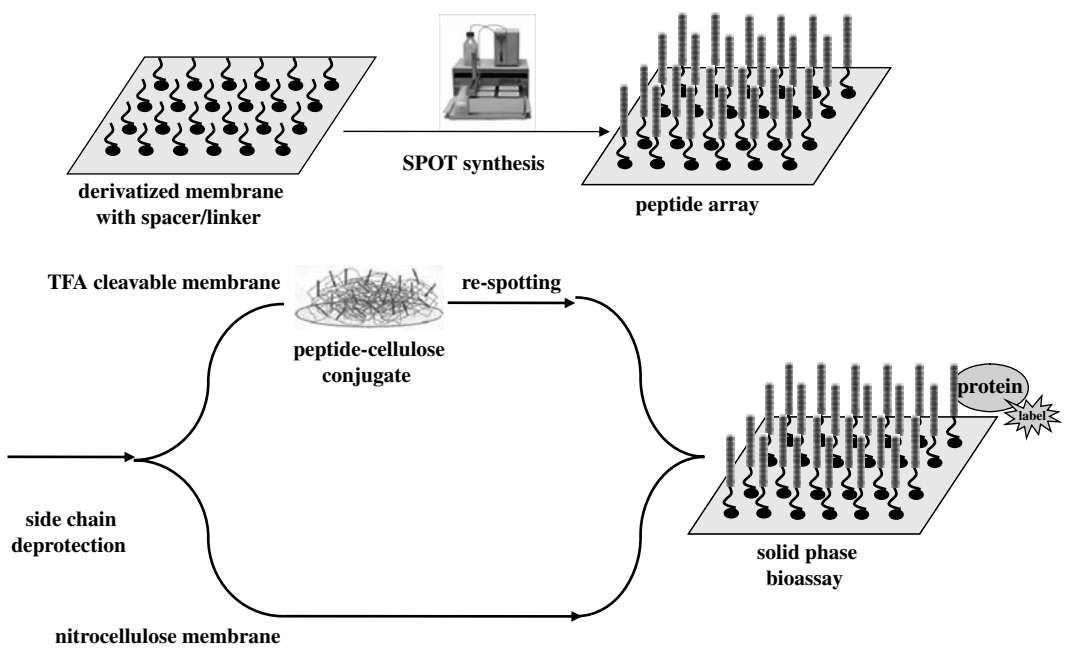


Fig. 3.1. Schematic representation of the SPOT and CelluSpot synthesis and their application in binding assays.

upon the completion of peptide chain assembly. At this step, the peptide array is ready for subsequent biochemical assays. For peptides synthesized on a trifluoroacetic acid (TFA) soluble cellulose membrane, they can be dissolved in organic solvent and respotted at a desired density on another solid support; and (e) the screening of the peptide array for a desired property by performing the corresponding biochemical or binding assays (**Fig. 3.1**).

Besides the SPOT approach, which is suitable for the synthesis of low- to medium-density peptide arrays, a combinatorial synthesis method was recently developed for the production of high-density peptide arrays featuring up to 40,000 peptide spots per cm^2 on a microchip (8). Using digital photolithography and photogenerated acid at the deprotection step, Pellois et al. (9) reported a novel approach by which to perform individually addressable parallel synthesis of peptides on microchips. A unique feature of this strategy is the use of photogenerated acid (PGA) for the removal of acid-labile sidechain protection groups upon light irradiation, which makes it possible, in principle, to synthesize high-density peptide microchips for a variety of pharmaceutical and proteomic applications (10).

3. Immobilization of Pre-made Peptides

An alternative method to produce a peptide array is by immobilization of peptides that have been synthesized and purified or by respotting peptide–cellulose conjugates prepared by the SPOT approach. This strategy is more efficient and cost-effective than in situ synthesis when only a small number of peptides are represented in the array or when multiple copies of the same peptide array are needed.

There are many strategies by which to immobilize a peptide. These include the use of a terminal cysteine residue for attachment to a glyoxylyl glass surface (11) or a surface bearing the maleimide group (12) or disulphide (3), the use of an alkoxyamine-terminated peptide for immobilization to an aldehyde-derivatized surface (13) and the use of biotinylated peptides for immobilization to avidin/streptavidin-coated glass slides (14). Besides these covalent immobilization strategies in which the peptides are covalently attached to a functionalized surface, CelluSpots™, developed by Intavis AG, exploits hydrophobic interaction for non-covalent peptide immobilization (*see* **Chapter 8**).

4. Applications

Over the past decade, peptide arrays have been used both in conventional biochemical and enzymatic analysis and in high throughput binding or epitope-mapping assays. A variety of molecular binding events, including protein–protein interactions (6, 15–21), peptide–enzyme interactions (11, 14), peptide–DNA interactions (22) and peptide–cell interactions (11), can be interrogated by peptide arrays. Peptide arrays have been found to be particularly amenable to identifying or characterizing protein–protein interactions. A number of signalling proteins and interaction domains, such as the tumour necrosis factor receptor-associated factors (TRAFs) (18), the PDZ domains (15), SH3 domains (16, 19, 21), PTB domains (20) and SH2 domains (6), have been studied by peptide arrays. Compared to DNA microarrays, peptide arrays are highly versatile and can be probed for binding to either macromolecules (such as DNA, RNA, protein or lipid) or small organic compounds.

4.1. Peptide Walking

In this application, a series of peptides of overlapping sequences are synthesized in an array to cover a part or the full sequence of a protein. This strategy is akin to "walking" through the protein sequence from the N- to the C-terminus. Depending on the application, peptides in an array can be synthesized at a specific length and with a given size of "overhanging" sequence. This strategy has been used to map the epitope of an antibody or the recognition sequence of an MHC molecule (3, 23). Peptide walking arrays are also useful tools by which to uncover the binding sites of one protein on another (17) and for revealing novel sequence motifs recognized by an interaction domain or by a protein kinase or phosphatase (4, 24–26).

4.2. Oriented Peptide Array Library

Much of our current knowledge on the specificity of protein interaction domains have been gleaned through the use of purified domains to screen peptide libraries that have been synthesized chemically or prepared using phage display (27–30). Songyang and Cantley pioneered the use of the oriented peptide library approach for the determination of binding specificity for a group of SH2 domains (29, 31). This method was subsequently extended to identify consensus motifs recognized by other protein interaction domains and to map the preferred substrates for protein kinases and phosphatases (29, 31). More recently, oriented peptide array libraries (OPAL) synthesized or respotted on a solid support have been used for charting domain or kinase specificity (5, 6, 32).

The SH2 domain is known to recognize short peptide sequences containing a phosphotyrosine (pY) residue and the specificity of an SH2 domain is governed by 3–5 residues C-terminal to the pY (33–39). In a typical OPAL designed for SH2 domain-binding, the pTyr residue, which acts as an anchoring site to allow for basal level binding to an SH2 domain, is held constant in the library (**Fig. 3.2**). This pTyr is flanked at both the N- and C-termini by amino acids to create a peptide suitable for SH2-binding. When a residue in the peptide sequence is replaced by a mixture "X" of all naturally occurring amino acids (except for Cys which makes a peptide prone to dimerization) during the synthesis, this will generate a peptide library containing 19 peptides. If "*n*" residues in a peptide are replaced by X, this will generate a library containing 19^n peptides. To generate a peptide array library, each X in the parent library is scanned with individual amino acids (**Fig. 3.2A**).

The OPAL approach was employed recently to systematically chart the specificity of human SH2 domains (6, 40). This method adopted CelluSpots technology to produce hundreds of identical OPALs with the degenerated sequence of $X^{-2}X^{1}pY^{0}X^{1}X^{2}X^{3}X^{4}$ for the parent library (**Fig. 3.2**). The selectivity of each residue at a position –2 to +4 relative to the pTyr for SH2-binding is determined by replacing the X with a given amino acid during synthesis. As shown in **Fig. 3.2**, the substitution of the X at position –2 by a

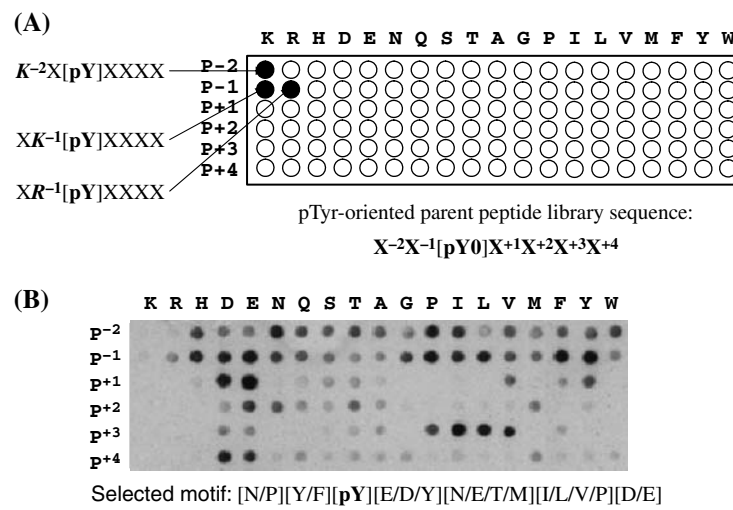


Fig. 3.2. (**A**) Schematic of a pTyr-oriented peptide array library (OPAL). The OPAL is derived from a library with the degenerated sequence XX[pY]XXXX, where X denotes a mixture of all naturally occurring amino acids except Cys, and pY denotes phosphotyrosine. (**B**) Probing of a pY-OPAL by an SH2 domain (from the protein FYN). A *dark* spot suggests a positive selection for the substituted residue at a given position of the peptide. The motif selected by the SH2 domain is indicated at the bottom of the OPAL membrane.

Lys (K) generates a sub-library spot containing the degenerated sequence of $K^{-2}X^{-1}pY^{0}X^{1}X^{2}X^{3}X^{4}$. Thus, a permutation scanning of the X's in the parent library would generate an array of $19 \times 6 = 114$ sub-libraries. In the subsequent binding study, if a given substitution produced a stronger than average binding signal, then the corresponding residue was considered favoured by the SH2 domain under concern (**Fig. 3.2**). As shown in **Fig. 3.2B**, the FYN SH2 domain selected the motif [N/P][Y/F][**pY**][E/D/Y][N/E/T/M][I/L/V/P][D/E] from an OPAL screen.

4.3. Peptide Array Target Screen (PATS)

The PATS approach is similar to the WISE method developed by Landgraf et al. (19) for identifying a protein interaction network in yeast. It is particularly suitable for identifying proteins bound by an interaction domain such as SH2, SH3, PTB, PDZ, etc. (41). When the consensus sequence for an interaction domain is available, this consensus motif can be used to search a protein database in order to retrieve peptides containing the motif. These potential binding targets are then represented on a peptide array and probed for binding to the domain of concern (**Fig. 3.3A**). The signals from the binding assay indicate which peptides (and the corresponding proteins) could be potential targets for the domain (and the

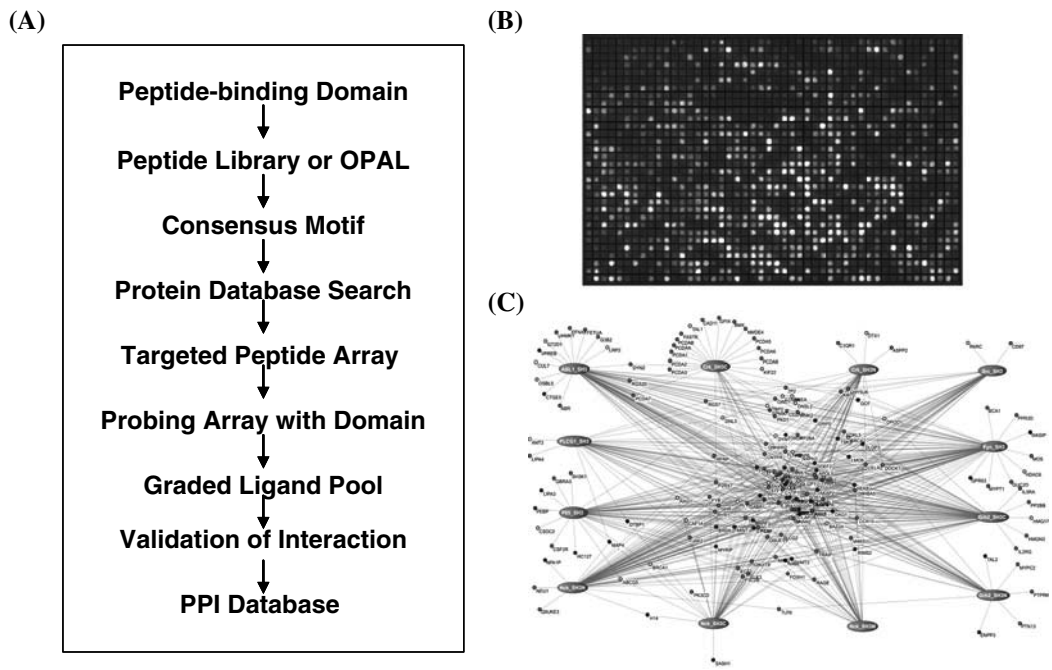


Fig. 3.3. (**A**) Flowchart of the targeted peptide array screening (or PATS) approach for systematic mapping of protein-to-protein interactions based on the principle of domain–linear motif recognition. PPI, protein-to-protein interaction. (**B**) Binding profile of the ABL1 SH3 domain to a 1536 peptide array. Bright (fluorescent) dots indicate positive binding. (**C**) An SH3 domain–protein interaction network identified by the PATS approach. For details, see Jia et al., Proteomics. 2007; 7: 1775–85.

corresponding protein) (**Fig. 3.3B**). We employed PATS in systematically identifying protein interactions mediated by a group of 12 human SH3 domains (**Fig. 3.3C**) (21).

5. Perspectives

Peptide array is gaining widespread usage in the post-genome era and has become a powerful tool in uncovering the molecular basis of protein–protein interactions. With the availability of the human genome sequence and by extension, the unmodified human proteome, it is now possible, in principle, to represent all proteins on a microarray. Because protein array often suffers from difficulty in protein expression and purification and from protein instability, peptide array provides an alternative to represent protein fragments on a solid support in a systematic manner. Another notable advantage of peptide array is that post-translational modifications (PTM) can be readily recapitulated by peptide synthesis. This advantage makes peptide array particularly amenable for studying interactions involving PTMs. We expect peptide array to play an important part in the large-scale identification and characterization of protein–protein interactions and in the elucidation of enzyme–substrate relationships.

References

1. Merrifield, R.B., (1963) Solid-phase peptide synthesis. I. The synthesis of a tetrapeptide. *J Am Chem Soc*, **85**, 2149–54.
2. Frank, R. and Overwin, H., (1996) SPOT-synthesis epitope analysis with arrays of synthetic peptides prepared on cellulose membranes, *Methods Mol Biol*, **66**, 149–169.
3. Wegner, G.J., Lee, H.J., and Corn, R.M., (2002) Characterization and optimization of peptide arrays for the study of epitope–antibody interactions using surface plasmon resonance imaging. *Anal Chem*, **74**, 5161–8.
4. Jia, C.Y., Nie, J., Wu, C., Li, C., and Li, S.S., (2005) Novel Src homology 3 domain-binding motifs identified from proteomic screen of a Pro-rich region. *Mol Cell Proteomics*, **4**, 1155–66.
5. Rodriguez, M., Li, S.S., Harper, J.W., and Songyang, Z., (2004) An oriented peptide array library (OPAL) strategy to study protein–protein interactions. *J Biol Chem*, **279**, 8802–7.
6. Huang, H., Li, L., Wu, C., Schibli, D., Colwill, K., Ma, S., et al., (2008) Defining the specificity space of the human SRC homology 2 domain. *Mol Cell Proteomics*, **7**, 768–84.
7. Frank, R., (1992) Spot-synthesis: an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. *Tetrahedron*, **48**, 9217–32.
8. Beyer, M., Nesterov, A., Block, I., Konig, K., Felgenhauer, T., Fernandez, S., et al., (2007) Combinatorial synthesis of peptide arrays onto a microchip. *Science*, **318**, 1888.
9. Pellois, J.P., Zhou, X., Srivannavit, O., Zhou, T., Gulari, E., and Gao, X., (2002) Individually addressable parallel peptide synthesis on microchips. *Nat Biotechnol*, **20**, 922–6.
10. Shin, D.S., Kim, D.H., Chung, W.J., and Lee, Y.S., (2005) Combinatorial solid

phase peptide synthesis and bioassays. *J Biochem Mol Biol*, **38**, 517–25.

11. Falsey, J.R., Renil, M., Park, S., Li, S., and Lam, K.S., (2001) Peptide and small molecule microarray for high throughput cell adhesion and functional assays. *Bioconjug Chem*, **12**, 346–53.
12. Takahashi, M., Nokihara, K. and Mihara, H., (2003) Construction of a protein-detection system using a loop peptide library with a fluorescence label. *Chem Biol*, **10**, 53–60.
13. Salisbury, C.M., Maly, D.J., and Ellman, J.A., (2002) Peptide microarrays for the determination of protease substrate specificity. *J Am Chem Soc*, **124**, 14868–70.
14. Lesaicherre, M.L., Uttamchandani, M., Chen, G.Y., and Yao, S.Q., (2002) Developing site-specific immobilization strategies of peptides in a microarray. *Bioorg Med Chem Lett*, **12**, 2079–83.
15. Schultz, J., Hoffmuller, U., Krause, G., Ashurst, J., Macias, M.J., Schmieder, P., et al., (1998) Specific interactions between the syntrophin PDZ domain and voltage-gated sodium channels. *Nat Struct Biol*, **5**, 19–24.
16. Cestra, G., Castagnoli, L., Dente, L., Minenkova, O., Petrelli, A., Migone, N., et al., (1999) The SH3 domains of endophilin and amphiphysin bind to the proline-rich region of synaptojanin 1 at distinct sites that display an unconventional binding specificity. *J Biol Chem*, **274**, 32001–7.
17. Knoblauch, N.T., Rudiger, S., Schonfeld, H.J., Driessen, A.J., Schneider-Mergener, J., and Bukau, B., (1999) Substrate specificity of the SecB chaperone. *J Biol Chem*, **274**, 34219–25.
18. Pullen, S.S., Labadia, M.E., Ingraham, R.H., McWhirter, S.M., Everdeen, D.S., Alber, T., et al., (1999) High-affinity interactions of tumor necrosis factor receptor-associated factors (TRAFs) and CD40 require TRAF trimerization and CD40 multimerization. *Biochemistry*, **38**, 10168–77.
19. Landgraf, C., Panni, S., Montecchi-Palazzi, L., Castagnoli, L., Schneider-Mergener, J., Volkmer-Engert, R., et al., (2004) Protein interaction networks by proteome peptide scanning. *PLoS Biol*, **2**, E14.
20. Smith, M.J., Hardy, W.R., Murphy, J.M., Jones, N., and Pawson, T., (2006) Screening for PTB domain binding partners and ligand specificity using proteome-derived NPXY peptide arrays. *Mol Cell Biol*, **26**, 8461–74.
21. Wu, C., Ma, M.H., Brown, K.R., Geisler, M., Li, L., Tzeng, E., et al., (2007) Systematic identification of SH3 domain-mediated human protein–protein interactions by peptide array target screening. *Proteomics*, **7**, 1775–85.
22. Weiler, J., Gausepohl, H., Hauser, N., Jensen, O.N., and Hoheisel, J.D., (1997) Hybridisation based DNA screening on peptide nucleic acid (PNA) oligomer arrays. *Nucleic Acids Res*, **25**, 2792–9.
23. Reineke, U., Kramer, A., and Schneider-Mergener, J., (1999) Antigen sequence- and library-based mapping of linear and discontinuous protein–protein-interaction sites by spot synthesis. *Curr Top Microbiol Immunol*, **243**, 23–36.
24. Buss, H., Dorrie, A., Schmitz, M.L., Frank, R., Livingstone, M., Resch, K., et al., (2004) Phosphorylation of serine 468 by GSK-3beta negatively regulates basal p65 NF-kappaB activity. *J Biol Chem*, **279**, 49571–4.
25. Szallasi, Z., Denning, M.F., Chang, E.Y., Rivera, J., Yuspa, S.H., Lehel, C., et al., (1995) Development of a rapid approach to identification of tyrosine phosphorylation sites: application to PKC delta phosphorylated upon activation of the high affinity receptor for IgE in rat basophilic leukemia cells. *Biochem Biophys Res Commun*, **214**, 888–94.
26. Espanel, X., Huguenin-Reggiani, M., and Van Huijsduijnen, R.H., (2002) The SPOT technique as a tool for studying protein tyrosine phosphatase substrate specificities. *Protein Sci*, **11**, 2326–34.
27. Sparks, A.B., Rider, J.E., and Kay, B.K., (1998) Mapping the specificity of SH3 domains with phage-displayed random-peptide libraries. *Methods Mol Biol*, **84**, 87–103.
28. Songyang, Z., Shoelson, S.E., McGlade, J., Olivier, P., Pawson, T., Bustelo, X.R., et al., (1994) Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav. *Mol Cell Biol*, **14**, 2777–85.
29. Songyang, Z. and Cantley, L.C., (1995) Recognition and specificity in protein tyrosine kinase-mediated signalling. *Trends Biochem Sci*, **20**, 470–5.
30. Songyang, Z. and Liu, D., (2001) Peptide library screening for determination of SH2 or phosphotyrosine-binding domain sequences. *Methods Enzymol*, **332**, 183–95.
31. Turk, B.E. and Cantley, L.C., (2003) Peptide libraries: at the crossroads of proteomics and bioinformatics. *Curr Opin Chem Biol*, **7**, 84–90.
32. Hutti, J.E., Jarrell, E.T., Chang, J.D., Abbott, D.W., Storz, P., Toker, A., et al.,

(2004) A rapid method for determining protein kinase phosphorylation specificity. *Nat Methods*, **1**, 27–9.

33. Sadowski, I., Stone, J.C. and Pawson, T., (1986) A noncatalytic domain conserved among cytoplasmic protein–tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. *Mol Cell Biol*, **6**, 4396–408.
34. Anderson, D., Koch, C.A., Grey, L., Ellis, C., Moran, M.F., and Pawson, T., (1990) Binding of SH2 domains of phospholipase C gamma 1, GAP, and Src to activated growth factor receptors. *Science*, **250**, 979–82.
35. Matsuda, M., Mayer, B.J., Fukui, Y., and Hanafusa, H., (1990) Binding of transforming protein, P47gag-crk, to a broad range of phosphotyrosine-containing proteins. *Science*, **248**, 1537–9.
36. Songyang, Z., Shoelson, S.E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W.G., et al., (1993) SH2 domains recognize specific phosphopeptide sequences. *Cell*, **72**, 767–78.
37. Blaikie, P., Immanuel, D., Wu, J., Li, N., Yajnik, V., and Margolis, B., (1994) A region in Shc distinct from the SH2 domain can bind tyrosine-phosphorylated growth factor receptors. *J Biol Chem*, **269**, 32031–4.
38. Kavanaugh, W.M. and Williams, L.T., (1994) An alternative to SH2 domains for binding tyrosine-phosphorylated proteins. *Science*, **266**, 1862–5.
39. van der Geer, P., Wiley, S., Lai, V.K., Olivier, J.P., Gish, G.D., Stephens, R., et al., (1995) A conserved amino-terminal Shc domain binds to phosphotyrosine motifs in activated receptors and phosphopeptides. *Curr Biol*, **5**, 404–12.
40. Li, L., Wu, C., Huang, H., Zhang, K., Gan, J., and Li, S.S., (2008) Prediction of phosphotyrosine signaling networks using a scoring matrix-assisted ligand identification approach. *Nucleic Acids Res,* **36**, 3263–73.
41. Pawson, T. and Nash, P., (2003) Assembly of cell regulatory systems through protein interaction domains. *Science*, **300**, 445–52.

Chapter 4

Molecular Simulations of Peptides: A Useful Tool for the Development of New Drugs and for the Study of Molecular Recognition

Massimiliano Meli and Giorgio Colombo

Abstract

The study of the molecular recognition and self-organization properties of peptides has emerged in recent years as a very active and diverse field of research, ranging from biomedicine to biotechnology and even to material sciences.

In the case of biomedicine, peptides can be used as ligands of biological receptors to gain insights into the structural, dynamical, and chemical determinants underlying the formation of complexes and identify new effectors of biological processes of interest. In the case of biotechnology and material science, short sequences have been used to understand the sequence determinants of the formation of ordered supra-molecular structures of nanoscale dimensions.

In this work, we will describe our research activities in these two areas of modern chemical biology. In the first part, we will describe the development of a new, specific, potent, and selective anticancer peptide and its use to obtain the information needed to identify a non-peptidic small molecular lead to be used as an inhibitor of cancer growth. In the second part, we will describe the introduction of a new method for the description of the self-organization process at the basis of the growth of ordered supra-molecular structures held together by weak, non-covalent, yet specific interactions.

Key words: Drug design, peptides, self-organization, molecular recognition.

1. Introduction

With the growth in computer power, the role of molecular modeling in bio-molecular studies has been gaining increasing strength over the last few years. Due to the dimensions of the molecules implicated in biological processes, pure quantum chemistry is not the most suitable method for these studies. In contrast, molecular mechanics or mixed methods as QM–MM (quantum mechanics–molecular mechanics) can answer the modeling demands for these systems. Moreover, with the completion of a number of genomes,

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_4,

and the subsequent extraction of genes and hence protein sequences, attention has been focused on the prediction and determination of the 3D structure of proteins. Dynamics is the next logical step in the chain that leads up to bio-molecular function and understanding of biological processes at the supra-molecular level.

Proteins are important targets for drug design and this is greatly facilitated if the 3D structure is known. A molecular modeling study of a bio-system provides a lot of information that often complements experimental data. In many cases a molecular modeling study can provide structural or other data in conditions not accessible for an experiment, i.e. extreme or dangerous experimental conditions, properties on time-scales too short for any experiment. From the point of view of the theory of dynamical systems, bio-molecules are probably among the most complex subjects of study as they not only involve many thousands of degrees of freedom for a given molecule, in addition to the need to model the solvating water environment, but they also span a huge hierarchy of functionally significant time-scales, from nanoseconds to milliseconds and beyond. Although quantum mechanics governs the interactions between the electrons of the atoms and molecules as well as the motions of light particles such as protons, the non-bonded interactions can be very well described by a classical potential-energy function or force field as part of a classical Hamiltonian of the system of interest. As already mentioned, we mainly consider atomic and molecular degrees of freedom with the corresponding classical force fields and classical Newtonian dynamics to sample the degrees of freedom. System sizes, which can be considered, range up to 10^5 or 10^6 atoms or particles. Classical simulations represent the theoretical framework of this chapter and discussion. They can be applied to the general problem of molecular recognition with the final goal to improve our design capabilities and our understanding of biological self-organization phenomena.

2. Overview

The study of the molecular recognition and self-organization properties of peptides has emerged in recent years as a very active and diverse field of research, ranging from biomedicine to biotechnology and even to material sciences (1).

In the case of biomedicine, peptides can be used as ligands of biological receptors to gain insights into the structural, dynamical, and chemical determinants underlying the formation of complexes and identify new effectors of biological processes of interest. In the case of biotechnology and material science, short sequences have been used to understand the sequence determinants of the formation of ordered supra-molecular structures of nanoscale dimensions.

In this work, we will describe our research activities in these two areas of modern chemical biology. In the first part, we will describe the development of a new, specific, potent, and selective anticancer peptide and its use to obtain the information needed to identify a non-peptidic small molecular lead to be used as an inhibitor of cancer growth. In the second part, we will describe new approaches for the description of the self-organization process at the basis of the growth of ordered supra-molecular structures held together by weak, non-covalent, yet specific interactions.

Development of a new anticancer lead: from a peptide to a new non-peptidic molecule. Anticancer agents that selectively kill tumor cells and spare normal tissues are urgently needed. In this work, we describe the engineering of a cell-permeable peptidomimetic, *shepherdin* (2), modeled on the binding interface between the molecular chaperone heat shock protein 90 (Hsp90) and the anti-apoptotic and mitotic regulator survivin. Hsp90 is a molecular chaperone that is involved in multiple signaling pathways for cell proliferation, survival, and cellular adaptation. It is highly and selectively overexpressed in cancer cells and is at the crossroad of a peculiar network of biological pathways. These characteristics indicate that Hsp90 can bet a good candidate for anticancer therapies. The active inhibitor of Hsp90 was identified by peptide scanning of the sequence of survivin, a client protein strongly interacting with the chaperone, followed by affinity studies. This allowed defining the sequence stretch of survivin covering residues 79–87 as the most active peptide. The retroinverso analog of this sequence was synthesized and further characterized as an active and stable inhibitor of Hsp90, and named *shepherdin* (3–5). In our work, we first investigated the dominant conformations of shepherdin in solution through all-atom, explicit solvent molecular dynamics (MD) simulations for a total time span of 400 ns(*6*). Statistical cluster analysis showed that shepherdin displays one main conformation, characterized by the presence of a turn involving residues G83–S84 and an overall hairpin geometry. The remaining clusters were mainly extended conformations, with the peptide backbone groups involved in hydrogen bonding with water.

The most populated conformation was subjected to multiple blind docking experiments on Hsp90 using the Autodock program (*7*). In all cases, the peptide was predicted to bind within the ATP binding site of Hsp90. Statistical analysis of the blind docking results showed that low energy poses are all highly correlated to one another, with a root mean square deviation (RMSD) from the global minimum structure lower than 2.5 Å. Control docking experiments were conducted with the extended structures representative of other clusters, but in those cases it was not possible to univocally identify any particular binding site on Hsp90. The free-energy minimum structure of the Hsp90/*shepherdin* complex was then subjected to two long, 54 and 73 ns, all-atom MD simulations. Analysis of the statistical and time-

dependent distribution of the interactions between functional groups of the ligand and of Hsp90 allowed to identify the most important residues in the stabilization of the bimolecular complex. These structural predictions were first verified by targeted mutations on the Hsp90 binding site. Individual substitutions at the predicted positions decreased binding by 20–60% (2).

These data were also used to define the minimal peptide sequence of *shepherdin* necessary for Hsp90 binding: through this procedure, we could identify *shepherdin* [79–83] as a novel variant able to inhibit Hsp90. This hypochapter was tested experimentally by measuring the levels of expression of Hsp90 client proteins, tumor cell viability, and apoptosis in cells where *shepherdin* [79–83] was administrated (8).

Finally, the structure and dynamics information on the interaction between the chaperone and the inhibiting peptidomimetic was used to develop pharmacophore models, keeping into account the motional and flexibility properties of both the ligand and the receptor. This computational design strategy proved able to identify the non-peptidic small molecule 5-aminoimidazole-4-carboxamide-1-β-D-ribofuranoside as a structurally novel inhibitor of heat shock protein 90 (Hsp90) (9) (**Fig. 4.1**). Experimental tests show that the compound binds the Hsp90 N-domain, destabilizes multiple Hsp90 client proteins in vivo, including survivin, and exhibits antiproliferative and proapoptotic activity in multiple tumor cell lines, while not affecting proliferation of normal human fibroblasts. More in general, we suggest that our all-atom MD-based methodology can be a valuable tool for the rational design of new small molecules, keeping structural and dynamical properties into account.

Study of the dynamics of peptide self-organization. Self-assembly is the process by which nature builds complex 3D multicomponent structures with well-defined functions starting from simple building blocks such as nucleotides, lipids, or aminoacids. In many cases, however, the self-assembly of peptides or proteins into fibrillar structures is the cause of degenerative diseases such as Alzheimer's,

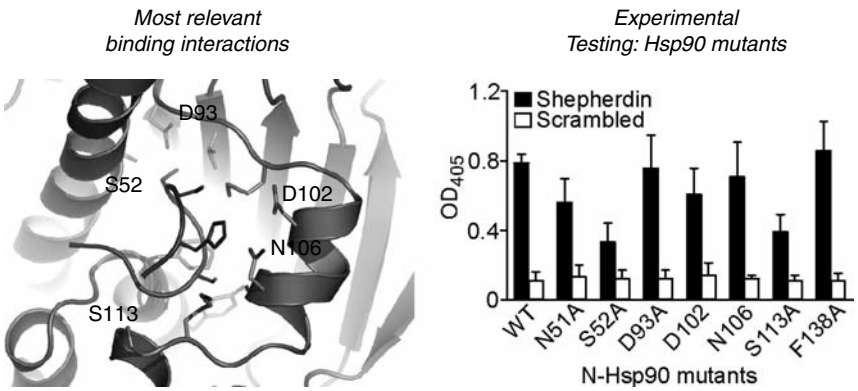


Fig. 4.1. On the right, the most relevant binding interactions between shepherdin-RV and the N-terminal domain of HPS90. On the left, the experimental test of the contacts identified with our docking-molecular dynamics study, from (2).

Parkinson's, type II diabetes, and so on (10). Because protein aggregation is very complex, involving a high number of degrees of freedom and a variety of oligomeric metastable intermediates, the detailed aggregation paths and structural characterization of the intermediates remain to be determined. We focus here on the computational study of the structural and dynamic aspects of the aggregation of a small peptide characterized experimentally as able to form stable β-sheet rich fibrils. We will present and discuss our approach that entails an initial coarse-grained Monte Carlo (MC) scheme followed by all-atom molecular dynamics (MD) simulations in explicit solvent (6). The MC evolution occurs in a simplified free-energy landscape and allows an efficient selection of marginally compact structures which are taken as viable initial conformations for the MD. In the initial coarse-grained part, multiple peptide copies are represented in a simulation box in random starting orientations. The coarse-grained MC structural representation is connected to the one with atomic resolution through a fine-graining reconstruction algorithm (11, 12). This two-stage strategy is used to select and follow the dynamics of several different unrelated conformations of an initial multimolecular aggregate. In several notable cases the initial aggregates evolve toward final structures rich in β-sheet content with interesting details in regards to the register, directionality, and type of interactions among different monomers. The analysis of the various MC–MD trajectories provides valuable insight into the details of the aggregation and self-organization mechanisms and particularly about the delicate influence of local and non-local interactions in steering the process (11, 13) (**Fig. 4.2**).

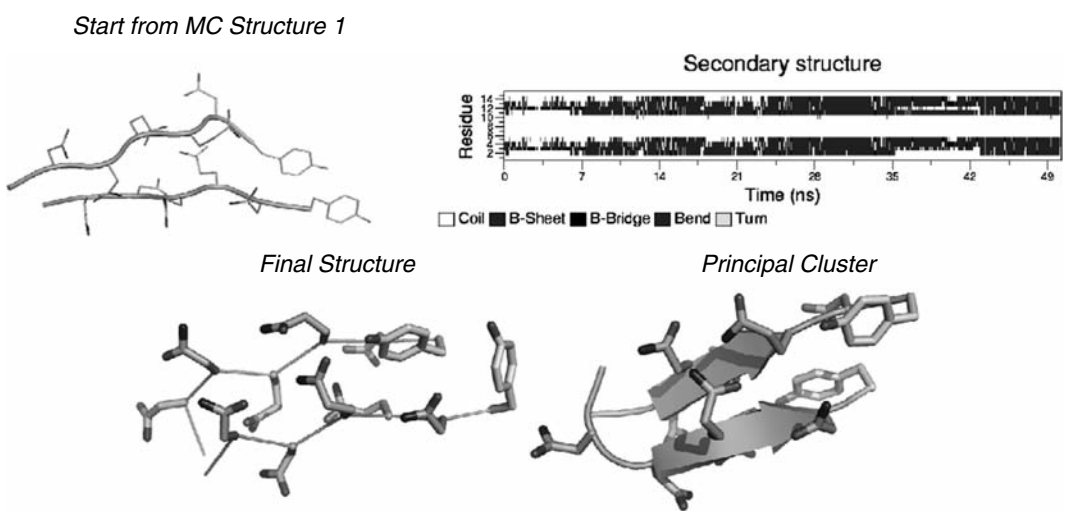


Fig. 4.2. *Top Panel*: the structure from MC simulation and the evolution of the secondary structure during the molecular dynamics simulation. *Bottom line*: the comparison of experimental structure of GNNQQNY peptide crystal and the structure obtained from our MC–MD procedure.

3. Simulation Methods

3.1. Introduction

As it is not possible to observe individual atoms or molecules directly, various models are used to describe and/or predict the properties of a system. In this respect the image of an atom or molecule will strongly depend on the models with which one has been presented. The sophistication of the model we choose depends on the property in which we are interested. Normally it is advisable to choose the simplest representation that will illustrate the property of interest satisfactorily. Over the last decade a combination of computer graphics and molecular modeling techniques has resulted in unprecedented power to create and manipulate 3D models of molecules using computers.

Nevertheless, one should always be extremely cautious. The images created on the screen can be very convincing. When using a program to predict a particular property of the system it is absolutely fundamental to always be aware that the underlying model may or may not bear any relationship to reality.

The general aim of this paragraph is to provide a basic description of all-atom molecular dynamics (MD) simulations of peptides and proteins.

3.2. General Aspects

Although normally represented as static structures, molecules such as proteins, lipids, DNA, carbohydrates, water, etc., are in fact dynamic. Most experimental properties, for example, measure a time average or an ensemble average over the range of possible configurations the molecule can adopt. One way to investigate the range of accessible configurations is to simulate the motions or dynamics of a molecule numerically. This can be done by computing a *trajectory*, i.e., a series of molecular configurations as a function of time by the simultaneous integration of the Newton's equations of motion. MD simulations enable the time-dependent behavior of atomic and molecular systems to be determined, providing a detailed description of the way in which a system changes from one conformation or configuration to another. Importantly, simulations can generate ensembles of representative configurations in such a way that accurate values of structural and thermodynamic properties can be obtained with a reasonable and feasible amount of computation. Statistical analysis is the link between the experimental and simulation world.

In this paragraph we will discuss some of the concepts and equations at the basis of molecular dynamics (MD) simulations.

3.3. Molecular Dynamics

Molecular dynamics simulation (MD) is a technique to compute the equilibrium and transport properties of a classical many-body system. In this context, the word classical indicates that the motions of the particles obey the laws of classical mechanics. This is an excellent approximation for a wide range of materials and applications where electronic motions and reorganizations are not involved, ranging from complex bio-molecular systems to material science to the study of polymers. In molecular dynamics, successive configurations of the system are generated by the application of Newton's laws of motion. The end result is a *trajectory* that specifies the variation of the positions and velocities of the particles in time. On this *trajectory* one calculates the properties of interest as time or ensemble averages.

The trajectory is generated by solving Newton's second law ($F = ma$) for every atom in the system. Newton's second law is actually a differential equation that can be re-written as:

$$\frac{d^2 x_i}{dt^2} = \frac{F_{x_i}}{m_i} \qquad [1]$$

From this equation, it is evident that particle $i$of mass m_i moves along coordinate x_i subject to the force F_{xi} which is due to the presence and interaction with atom i of all other particles in the system. The force can be expressed as the negative derivative of a potential function $V(r_1, r_2, r_3, \ldots\ldots\ldots r_N)$ describing the fundamental types of interactions in the system.

$$F_i = -\frac{\partial V}{\partial r_i} \qquad [2]$$

Roughly speaking, V can be considered as the potential energy of the system as a function of atomic positions. The equations are solved simultaneously in small *time steps*, dt. The system is followed for some time, taking care that the temperature and pressure remain at the required values, and the coordinates are written to an output file at regular intervals generating the trajectory.

A typical potential function for all-atom protein simulations is expressed in the form:

$$\begin{aligned} V(r_1, r_2, \ldots\ldots r_N) = & \sum_{bonds} \frac{1}{2} K_b [b - b_0]^2 \\ & + \sum_{angles} \frac{1}{2} K_\theta [\theta - \theta_0]^2 + \sum_{impropdihedrals} \frac{1}{2} K_\zeta [\zeta - \zeta_0]^2 \\ & + \sum_{dihedrals} K_\varphi [1 + \cos(n\varphi - \delta)] \\ & + \sum_{pairs(i,j)} \Big[C_{12}(i,j)/r_{ij}^{12} - C_6(i,j)/r_{ij}^6 \\ & + q_i q_j / (4\pi\varepsilon_0 \varepsilon_r r_{ij}) \Big] \end{aligned} \qquad [3]$$

The potential V,expressed through equation [3] is also called *Force Field* (see for instance: (16) or (17)). In this chapter we will use predominantly the GROMOS force field as implemented in the Gromacs Package (6). Each term describes a pairwise relation, taking into account the physico-chemical interactions present in the system. In particular:

$$\sum_{bonds} \frac{1}{2} K_b [b - b_0]^2$$

describes the vibrations of covalent bonds around their equilibrium positions b_0. In a classical framework, the vibration is simply expressed in the form of the Hook's law, with a spring constant K_b.

$$\sum_{angles} \frac{1}{2} K_\theta [\theta - \theta_0]^2$$

describes the angular vibration around the equilibrium position, and it is basically expressed with the same form as the previous one.

$$\sum_{impropatopdihedrals} \frac{1}{2} K_\zeta [\zeta - \zeta_0]^2 \text{and} \sum_{dihedrals} K_\varphi [1 + \cos(n\varphi - \delta)]$$

are used to describe dihedral dependencies within the molecule.

The final term in the equation

$$\sum_{pairs(i,j)} \left[C_{12}(i,j)/r_{ij}^{12} - C_6(i,j)/r_{ij}^6 + q_i q_j/(4\pi\varepsilon_0\varepsilon_r r_{ij}) \right]$$

describes all non-bonded interactions. The first two terms take into account van der Waals interactions and decay very fast as a function of distance, while the third one is basically a formulation of the classical electrostatic interaction between two charges (the charge being the formal charge on an atom in a certain environment). Several parameters appear in the *Force Field*, such as the equilibrium distances or the force constants or van der Waals and electrostatic parameters. They have to be determined by either using experimental data, or by fitting to high level ab initio calculations. Clearly, the parametric nature of the *force fields* imposes restrictions to their uses in contexts which are different from the ones they have been developed for.

3.3.1. Setting Up and Running a Molecular Dynamics Simulation

Molecular dynamics simulations are for many aspects very similar to real laboratory experiments. In "real life" an experiment is performed by preparing the sample we wish to study. We then connect our sample to some measuring machine (thermometer, CD spectrometer, NMR, etc.) and we measure the property we are most interested in over a certain amount of time. If, as in

most cases, our measurements are subject to statistical noise, then the higher the number of samples and longer the simulation time, the more accurate our measurements will be. In MD we follow the same approach. In many cases MD simulations are actually defined as "computational experiments." First we have to prepare a sample. This corresponds to choosing the appropriate model system for the quantity we want to calculate. Generally, the model system will consist of N particles which will interact under the action of the potential and forces defined in equations [1]–[3]. The system (and the properties of interest) will evolve as a function of time until they reach an equilibrium stage. After the equilibration has been reached, we can carry out the effective measurement. The choices of the model system, the equilibration time, and the way the measurement is carried out are very sensitive points which have to be faced carefully. Incorrect results or bad artifacts can be generated by using the wrong model for the phenomenon one wants to describe, by using too short equilibration/measurement times, or by not noticing irreversible and chemically meaningless changes that can occur in the system. For details on the specific treatment of interatomic forces and alternative simulation methods see (16–53).

3.3.2. Integrating the Equations of Motion

Now that we have seen how to calculate the forces among the particles, and the kinds of problems that may arise, we can integrate Newton's equations of motion, equation [1] in particular. Algorithms have been developed to do that. As an introduction, let us consider how to calculate the position of a particle at a certain time $t + \Delta t$, $r(t + \Delta t)$. We first can expand in Taylor series the coordinate of a particle around time t.

$$r(t+\Delta t) = r(t) + v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 + \frac{\Delta t^3}{3!}\dddot{r} + O(\Delta t^4) \quad [4]$$

similarly,

$$r(t-\Delta t) = r(t) - v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 - \frac{\Delta t^3}{3!}\dddot{r} + O(\Delta t^4) \quad [5]$$

Summing these two equations, we obtain:

$$r(t+\Delta t) + r(t-\Delta t) = 2r(t) + \frac{f(t)}{m}\Delta t^2 + O(\Delta t^4) \quad [6]$$

or

$$r(t+\Delta t) \approx 2r(t) - r(t-\Delta t) + \frac{f(t)}{m}\Delta t^2 \quad [7]$$

The estimate of the new position contains an error which is of the order of Δt^4, where Δt is the time step in out MD scheme, typically between 1 and 10 fs in the simulation of bio-molecular systems. The algorithm presented here is the so called Verlet algorithm. One should note from [7] that velocities are not used to compute the new position. One can, however, compute the velocity from the knowledge of the trajectory (just like in normal dynamics exercises in physics courses) using:

$$r(t + \Delta t) - r(t - \Delta t) = 2v(t)\Delta t + O(\Delta t^3) \qquad [8]$$

or

$$v(t) = \frac{r(t + \Delta t) - r(t - \Delta t)}{2\Delta t} + O(\Delta t^2) \qquad [9]$$

This expression for the velocity is exact to within an order Δt^2. However, it is not possible to obtain a better estimate of the velocities (and hence of the kinetic energy and temperature of the system) using Verlet-like algorithms.

Once the new positions have been calculated, we can discard the old ones at t–Δt. The current positions become the old ones and the new positions become the current positions. After each time step, we can compute the current temperature, the potential energy, and the total energy. The total energy *must* be conserved, and a good integration scheme has to ensure that in the very first place.

Other integration algorithms can be used. In GROMACS, in particular, the so-called *leap-frog* algorithm is used for the integration of the equations of motion. The *leap-frog* algorithm uses positions $r$at time t and velocities $v$at time t-$\Delta t/2$; it updates the positions and velocities using the force $f(t)$ determined by the positions at time t:

$$v(t + \frac{\Delta t}{2}) = v(t - \frac{\Delta t}{2}) + \frac{f(t)}{m}\Delta t \qquad [10]$$

$$r(t + \Delta t) = r(t) + v(t + \frac{\Delta t}{2})\Delta t \qquad [11]$$

This algorithm can be visualized pictorially (**Fig. 4.3**).

This completes the basic description of the all-atom molecular dynamics simulation technique.

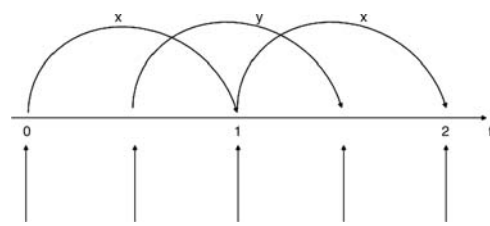


Fig. 4.3. Scheme of the leap-frog algorithm.

3.3.3. Temperature and Pressure Control

MD simulations are generally performed at constant number of particles (N), constant volume (V), and constant energy (E) (NVE ensemble). However, to obtain an easier connection with experiment, it is often desirable to run simulations at constant temperature (T) or pressure (P). The two most common simulation ensembles are in fact the NVT and NPT. In the following we will concentrate on how to control the temperature and the pressure in MD simulations.

3.3.4. Temperature Control

The control of temperature in MD simulations is fundamental for comparison of results with experimental findings. Moreover, it might be of interest to investigate the behavior of a system at different temperatures or to check the temperature-induced unfolding behavior of a protein, DNA stretch, etc. Simulated annealing MD protocols, in which the temperature is lowered in a controlled fashion, are also of interest for NMR or X-ray structure refinements.

Temperature is related to the average kinetic energy:

$$\langle K \rangle_{NVT} = \frac{3}{2} N k_B T \qquad [12]$$

From this relation, it is evident that one easy way to control the temperature would be to scale the kinetic energy of the particles and hence their velocity.

An alternative way to maintain the temperature close to the desired value is to couple the system to an external heath bath that is fixed at the desired temperature. This is the basis of the Berendsen coupling algorithm. The bath acts as a heat reservoir which can supply or remove energy from the system. The velocities are scaled at each time step, such that the rate at which the temperature changes is proportional to the difference in temperature between the bath and the system. If T_0 is defined as the reference temperature and Tas the instantaneous one, we have:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \qquad [13]$$

which means that a temperature deviation decays exponentially with a time constant τ. This method of coupling has the advantage that the strength of coupling can be varied and adapted to the different situations by just changing the scaling factor τ. For equilibration purposes, for instance, the coupling time can be taken quite short (e.g., 0.01 ps, which means that velocities will be rescaled every 0.01 ps), while for equilibrium runs this value can be much higher (0.5 ps). In this case the influence on the dynamics is very low. In general, if τ is large then the coupling is weak. If τ is small the coupling is strong, and when it equals the time step, we simply have a velocity rescaling algorithm.

The main problem with this algorithm is that it does not generate rigorous canonical averages. Velocity rescaling artificially prolongs any temperature difference between components of the system, and in many cases the phenomenon of "hot" solvent–"cold" solute arises. One solution is to couple different components separately to the heath bath, but the problem of unequal energy distribution among components may still remain.

3.3.5. Pressure Control

Just like one wants to control the temperature, one may also want to control the pressure in a simulation. This allows checking the behavior of the system to be explored as a function of pressure, enabling one to study, e.g., conformational changes induced by ultra-high pressure conditions. These types of studies are being applied for instance to study enzymatic reactivity for industrial applications, or in the study of the properties of, e.g., Prion Proteins and their conformational behavior. Pressure fluctuations are generally much more pronounced than temperature fluctuations, since the pressure is related to the *virial*, which is obtained as the product of the positions and the derivative of the potential energy function. This product is much more sensitive to the variations in position than the internal energy, which brings about bigger pressure fluctuations.

As a pressure control algorithm, we will present the Berendsen one. It is based on the same philosophy as the temperature control. This algorithm rescales the coordinates and the box vectors every step with a matrix μ, which has the effect of a first-order relaxation of the pressure toward a given reference pressure P_0.

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_P} \qquad [14]$$

The scaling matrix μ is given by:

$$\mu_{ij} = \delta_{ij} - \frac{\Delta t}{3\tau_P}\beta_{ij}\{P_{0ij} - P_{ij}(t)\} \qquad [15]$$

Here β is the isothermal compressibility of the system. In most cases this will be a diagonal matrix, with equal elements on the diagonal, the value of which is generally not known. In general, it is sufficient to consider a rough estimate of the value of β since this value only affects the non-critical time constant of the pressure relaxation without affecting the average pressure itself. For water at 1 atm and 300 K this value is $\beta = 4.6 \times 10^{-10}$ Pa^{-1} $= 4.6 \times 10^{-5}$ Bar^{-1}. The scaling can be done isotropically or anisotropically depending on the type of system being simulated. For instance for the simulation of a globular protein, when we are only interested in fluctuations, etc., isotropic scaling is fine. For system with interfaces or with conformational changes in one direction mainly, anisotropic scaling is best. In general, in this type of algorithm, the volume of the simulation box is changed (scaled) according to the coupling protocol.

Other methods are being used nowadays for pressure and temperature coupling, but they will not be discussed herein, as they are out of the scopes of the paragraph.

3.3.6. Summary

This chapter has introduced the basic concepts of molecular dynamics simulations, giving a brief description of the basic quantities accessible, the capabilities of the technique, etc. In general, the algorithms and the concepts used are quite simple and not particularly mathematically involved (except for Ewald summation for electrostatics). However, one should always keep in mind that all-atom MD simulations on realistic systems are extremely computationally intensive and that one should always choose the most appropriate model for the description of the system under exam. The wrong choice of initial condition, or of the electrostatic systems, etc., will certainly have serious consequences on the final results. One should never use MD simulations as a simple black box.

3.4. Simplifying the Conformational Search Problem. The Coarse-Grained Monte Carlo for Protein and Peptide Simulations

Several methods and techniques have been developed over the years to overcome the time limitations of all-atom MD. A widely used strategy is to renounce to follow the complete conformational pathways in favor of identifying the conformations of lowest free energy through various optimization techniques including simulated annealing, multicanonical methods, genetic algorithms, etc. (20–27). Another commonly employed route is to simplify not only the description of the solvent but also of the protein structure itself. Accordingly, several coarse-grained models have been introduced where the protein is described as a chain of linked beads which interact through suitable effective potentials. The drastic structural simplification entailed by these models allows exploration of a portion of the conformational space vastly larger than in all-atom MD. However, the limitations of both the structural representations and of the energy functions clearly prevent reproduction of the finer features of the conformational-recognition processes. Nevertheless they have proved valuable for capturing and analyzing various general aspects such as the calorimetric co-operativity, the relation between native-state topology and folding rates, the identification of folding nuclei, and the modeling of functional motions, etc. (28–39). These observations suggested the approach we introduced previously for the study of protein folding, and that we extend here to the problem of peptide self-organization, which combines a coarse-grained Monte Carlo (MC) search with all-atom molecular dynamics (MD) simulations in explicit solvent. The effect of the coarse-grained part is to simplify the protein's energy landscape so as to identify efficiently physically meaningful starting conformations for the MD. After bringing these structures into the realm of all-atom representations, explicit solvent MD is used to introduce back the fine

chemical details which are ultimately responsible for driving the evolution toward the native state. The link between the two structural representations is a fine-graining algorithm which allows reliable reconstruction of the full atomic detail of the protein using a library of previously generated protein fragments. The scheme adopted here, therefore, aims at extending the reach of ordinary MD simulations by using simple physico-chemical criteria to identify the marginally compact starting conformations in which partial formation of intermolecular stabilizing interactions has taken place. Thus, the approach has a different spirit from that followed in other two-stage approaches, notably the pioneering work of Vieth et al. (40) and Liwo et al. (41) where the fine-graining step was aimed at perturbing or refining the coarse-grained structures which minimized a given energy functional.

3.4.1. The Coarse-Grained Monte Carlo Method

The preliminary Monte Carlo exploration of the configurational space is designed to be as efficient as possible. To this end, we have adopted a simplified representation of peptide conformations (**Fig. 4.4**). Most of the procedures adopted to coarse grain the microscopic degrees of freedom of proteins and peptides substitute a whole amino acid with a small number of effective centroids. We have followed this choice and used effective C_α and C_β centroids for each amino acid (with the exception of *Gly* which lacks the C_β centroid). This choice

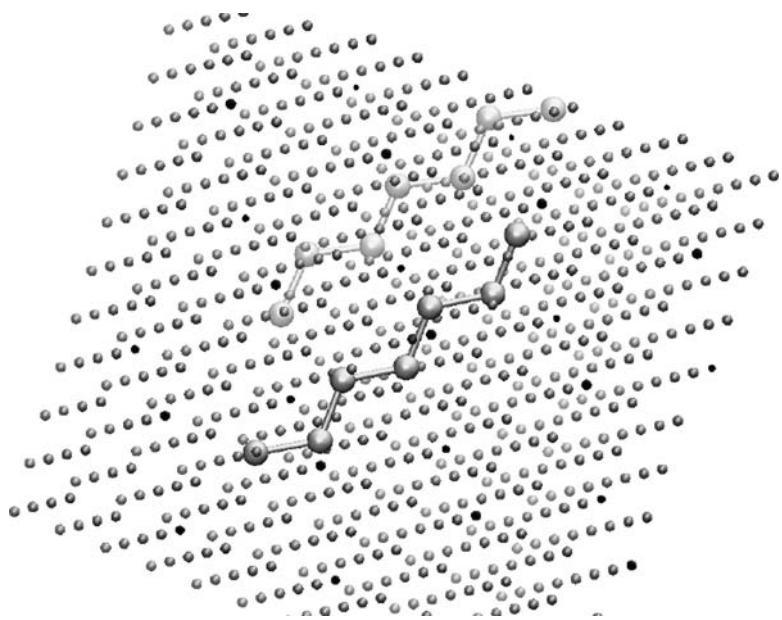


Fig. 4.4. An example of the 3D-grid used for MC simulation. In *blue* are showed the allowed position for peptide (in bigger balls and sticks) movement.

brings about a drastic simplification for the structure of the Hamiltonian which will involve only interactions between the centroids. In particular we regard the C_α centroids as the centers of the effective interactions among contacting residues. The C_β ones play, instead, a passive role being used to capture the steric hindrance of the side chains. In order to minimize the dimensionality of conformational space only the C_α degrees of freedom where considered and the effective C_βs were constructed according to the Park and Levitt rule (42).

In the extension of this approach to the case of self-organization of multiple peptide chains, the single peptide chains are represented in their extended structure, with no internal flexibility. Motions in the 3D space will involve rotations and translations of chains (**Fig. 4.4**). The dynamics of multiple chains in configuration space is carried out using a Monte Carlo technique. At each attempted Monte Carlo step the current configuration of the system (C_α-trace) is distorted by means of rotations and translations of one or more than one chain on the points of a 3D grid. These types of moves are employed since they do not alter the separation of consecutive C_αs (bond length). The newly generated configuration is then accepted/rejected according to the standard Metropolis rule. The system energy function comprises the following terms:

$$H = H_{steric} + H_{pairwise} \qquad [16]$$

The first term, H_{steric} is used to enforce some basic steric constraints on the conformations generated with the MC procedure. These knowledge-based constraints involve mainly two-body terms. In particular, no two centroids of distinct residues, i and j (be they C_α or C_β) are allowed to come at a distance smaller than 3.0 Å.

The second energy term, H_{pairwise} is used, instead, to capture the effective interactions among pairs of sufficiently closed residues. The strength of the contact interaction between two residues i and j (if only if belonging to different chain) is modulated with a sigmoidal function, $f(r_{i,j})$, which depends on the separation of the corresponding C_α centroids, $r_{i,j}$, and has a point of inflection at 6.5 Å,

$$f(r_{i,j}) = \frac{1}{2} + \frac{1}{2}\tanh(6.5 - r_{i,j}) \qquad [17]$$

Therefore, H_{pairwise} can be written as:

$$H_{pairwise} = \sum_{i,j}{}' \varepsilon(S_i, S_j) f(r_{i,j}) \qquad [18]$$

where the prime denotes that the summation runs over non-consecutive residues, S_i denotes the type of residue at the sequence position i, and ε is the symmetric matrix describing the strength of

the effective residue–residue interactions, the sum run over only residues that belong to different chain. Various criteria have been employed to extract tables of effective potentials. Here we recoursed to the one developed by Kolinsky, Godzik, and Skolnick (43), which proved useful to identify the correct fold of short proteins (44).

The 3D space in which multiple peptide copies move is discretized by means of a grid representation. The grid is cubic and its points are separated by 3 Å. This makes the exploration of the otherwise continuous space separating different copies easier and more efficient. In particular, at the beginning of each simulation, peptide chains are randomly positioned in a cubic box, such that peptide–peptide separation is bigger than the length of a single extended chain. This should avoid biases toward particular configurations of the system. Randomly selected peptides are then randomly translated such that the center of mass (CoM) occupies a new grid point. At the same time single chains are rotated in 3D space using the algebra of quaternions.

The configurational space sampling is based on a simulated annealing Metropolis Monte Carlo algorithm. As mentioned above, all the protein chain configurations generated through the Monte Carlo procedure have a simplified structural representation essentially described by the C_α degrees of freedom. These coarse-grained structures need to be reconstructed with all the atomic detail before they can be processed in ordinary all-atom molecular dynamics schemes. The problem of reconstructing reliably a protein's atomic detail starting from a coarse-grained representation has been addressed previously. Several studies have appeared which address the problem of reconstructing, with full atomic detail, a protein's backbone and/or side chains starting from very limited structural information (45–50). In the present study we have used a novel knowledge-based strategy that is simple and whose accuracy is not inferior to more sophisticated techniques. This strategy was developed by Micheletti (29, 51). We now illustrate this reconstruction scheme. For reasons of space our description will be not complete, but schematic.

The algorithm that reconstructs the full atomic detail of a given C_α trace is based on the use of a library of protein fragments built from about 100 NMR gapless structures taken from the PDB select list (52, 53). These structures were parsed into template fragments of four consecutive residues retaining the whole atomic detail of the main chain and of the side chains of the middle residues. For each set of four consecutive C_αs in the C_α trace one finds the best superimposable fragment in the library and assigns the central peptide plane to the protein to be reconstructed after an optimal roto-translation (50) the first and last peptide planes are treated separately. Finally, the side chain of any given residue, R, is

obtained by first considering only the set of template fragments where one of the middle residues, R', is of the same type as R. Next, after aligning (in sequence) R and R', the side chain of R is assigned (again after an optimal roto-translation) from the fragment providing the best superposition with the reconstructed backbone. Large deviations may be encountered in correspondence of charged residues, whose side chain orientation is heavily influenced by the local electrostatic (obviously not captured in our simple reconstruction scheme). This is, however, a minor problem in this context, since the reconstructed structure is solvated, energy-minimized, and equilibrated before starting the MD run.

3.5. Molecular Docking

In this chapter work, we will make use of molecular docking programs and algorithms. In molecular docking we attempt to predict the structure of an intermolecular complex between two (or more) molecules. In general an algorithm generates an ensemble of possible structures for the complex which have to be scored by some type of scoring function. We will make use of the Autodock suite of programs. A complete description of Autodock can be found elsewhere (7). It can be summarized as follows: a 2.5 Å spaced grid is positioned around the receptor, and the ligand is forced to move on the grid points. The interaction energy between the ligand and the protein is calculated at every grid point and involves the summation of a 6–12 Lennard–Jones term, the electrostatic term calculated according to the Mehler–Solmajer model, and a term accounting for the flexibility of ligand side chains. The conformational scoring is carried out using the Lamarckian Genetic Algorithm. More detailed treatments of docking algorithms, scoring, etc., can be found elsewhere and are not within the scopes of the present work.

4. Rational Development of Shepherdin

4.1. Introduction Molecular Determinants of Cancer

4.1.1. Survivin and Hsp90

Cancer development is associated with the presence of multiple molecular abnormalities and genetic instabilities which may elude the identification of one single, disease-driving oncogene. As a consequence, pathways that intersect multiple essential functions of tumor cells may provide wide therapeutic opportunities. Two of these intersecting pathways are the control of cell death (*apoptosis*) and the adaptive response to environmental insults and cellular stress stimuli mediated by the class of proteins named molecular chaperones.

Apoptosis plays an important physiological role in several processes, ranging from embryonic development to maintenance of adult tissue homeostasis **(Fig. 4.5)**.

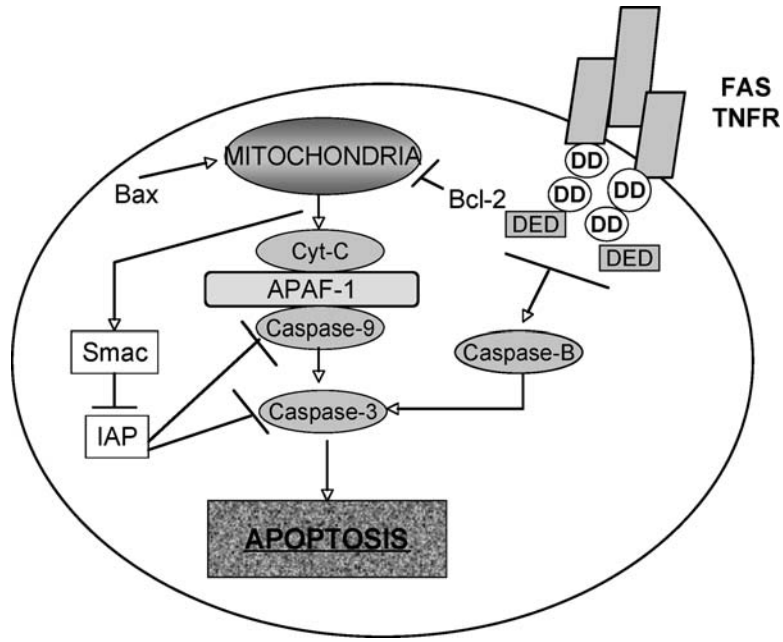


Fig. 4.5. ''Extrinsic'' and ''intrinsic'' apoptotic pathways in human cells.

This evolutionary conserved genetic programme of cell death is characterized by unique biochemical and morphological features that distinguish it from necrosis (54). Defects in the physiological pathways for apoptosis contribute to many types of cancers. Moreover, impairment of apoptotic cell death might negatively affect the response of cancer cells to chemotherapy and irradiation and lead to treatment resistance. Thus far, two major pathways of apoptosis have been identified. The *"extrinsic"* pathway is triggered by the binding of ligands to extracellular membrane receptors (death receptors) which belong to the tumor necrosis factor-family and leads to activation of caspase-8. Caspases (cysteine proteases that cleave after aspartic acid) are constitutively inactive zymogens (pro-caspases) that can trans-process each other, producing fully active proteases and resulting in a proteolytic cascade. The *intrinsic* apoptotic pathway involves mitochondria, which respond to pro-apoptotic signals by releasing cytochrome c. Cytochrome c binds and activates the apoptotic protease activating factor-1 (Apaf-1), causing assembly of a multiprotein caspase-activating complex (apoptosome) and leading to activation of caspase-9 and initiation of a protease cascade **(Fig. 4.5**).

This pathway is primarily governed by proteins of the Bcl-2 family, which include anti-apoptotic molecules (Bcl-2, Bcl-xl, Mcl-1) and pro-apoptotic molecules (Bax, Bak, Bcl-xs, Bad, and Bid) able to differentially affect mitochondrial homeostasis and

cytochrome c release. Moreover, other proteins belonging to the inhibitors of apoptosis protein (IAP) family, including ML-IAP, XIAP, cIAP1, cIAP2, NIAP, apollon, and survivin, are able to block a common step downstream of mitochondrial cytochrome c release by inhibiting terminal effector caspase-3 and caspase-7, and interfering with caspase-9 activity and processing (4). Survivin, in particular, has been shown to play a particularly important role in the inhibition of apoptosis by its high levels of up-regulation in a wide series of cancer cells. Moreover, up-regulation of survivin has also been noted in other rapidly developing and differentiating cells, such as embryos. Its levels are however generally undetectable in normal cells and tissues. With regards to the precise role of survivin in cell death, it is still controversial whether this protein inhibits caspases through direct binding, like other IAPs, or indirectly, via the intervention of other proteins. What is clear in all cases is that in order to carry out its functions within cells, survivin has to be correctly folded. Correct folding of survivin to its active structure is attained through interaction with the chaperone protein heat shock protein 90 (Hsp90). This interaction identifies the crossing point of the apoptotic pathways with adaptive responses to cellular stresses and external stimuli.

Following exposure to environmental insults, in fact, the cells in most tissues dramatically increase the production of a small group of proteins that are collectively known as "heat-shock" or stress proteins. Many groups over the past 30 years have shown that these heat shock proteins (HSPs) and their close, constitutively expressed relatives are actually molecular "chaperones" that guard against illicit or promiscuous interactions between other proteins. Their basal levels facilitate normal protein folding and guard the proteome from the dangers of mis-folding and aggregation (55).

Their increased expression in tissues that are subjected to various proteotoxic stressors (including heat, heavy metals, hypoxia, and acidosis) is an adaptive response that enhances cell survival. Both functions are needed in tumors. So, the increased expression of chaperone proteins that is observed in many tumor types undoubtedly reflects the efforts of malignant cells to maintain homeostasis in a hostile environment. However, in addition to facilitating the survival of tumor cells within their stressful microenvironments, other evidence indicates that chaperone proteins also allow tumor cells to tolerate alterations from within. Mutations in crucial signaling molecules that would otherwise be lethal are not only tolerated but actually help to drive oncogenesis (56). In this sense, at the phenotypic level, chaperones seem to serve as biochemical buffers for the numerous genetic lesions that are characteristic of most human cancers. As our understanding of the roles that chaperone proteins have in

initiating and maintaining transformed phenotypes increases, so too does our interest in the pharmacological modulation of chaperone function for the treatment of cancer and other diseases (57, 58). Over the past decade, *HSP90*, an ATPase-directed chaperone, has been identified as a target for its importance in the development of important proteins: the repertoire of proteins which are folded by Hsp90 (defined *client proteins*) is restricted mainly to growth-regulatory and signaling molecules, especially kinases and transcription factors, which may contribute to tumor cell maintenance (59, 60). Several small-molecule drugs that target the molecular chaperone have been identified as potential anticancer agents. These drugs have the unusual ability of disrupting the activity of numerous receptors, kinases, and transcription factors that are known to be involved in oncogenesis.

By realizing that survivin is a Hsp90 client protein with a critical role in cancer development, we identified the survivin–Hsp90 complex as a target for the disabling of multiple intersecting signaling pathways in tumors. Disrupting and preventing the formation of this complex by inhibiting the chaperone function of Hsp90 may provide wider therapeutic opportunities in the selective treatment of cancer cells. The activity of Hsp90 has been inhibited by blocking its ATPase activity by a class of antibiotics prototypically exemplified by geldanamycin (GA). Derivatives of GA are now in clinical trials.

In this chapter, we report the computational/theoretical structure-based design and characterization of *shepherdin*, a novel peptidomimetic antagonist of the complex between Hsp90 and survivin. Using synthetic peptidyl mimicry, we recently identified a survivin sequence K79–K90, which blocks the interaction between survivin and Hsp90, in vitro (3). By surface plasmon resonance, the survivin K79–K90 peptide-bound Hsp90 with high affinity ($K_D \pm \text{SEM} = 8.38 \times 10^{-8} \pm 3.5^{-9}$ M) at increasing ligand concentrations (0.1–10 μM). Conversely, a mutant C84→A peptide (see below) did not bind Hsp90, and exhibited a 10-fold increased off-rate as compared with the wild-type sequence, as shown by surface plasmon resonance. To narrow the minimal interacting region in K79–K90, we tested variant peptides, and found that the sequence I74–L87 retained the ability to bind Hsp90 in vivo with its associated chaperone Hsp70, but not Hsp27, by affinity chromatography of B lymphoma Raji cell extracts (**Fig. 4.6**).

Therefore, the minimal survivin sequence K79–L87 was named shepherdin for its binding to the "shepherding" chaperone Hsp90. To test whether shepherdin disrupted the survivin–Hsp90 complex in a cellular context, we reconstituted this interaction in reticulocyte extracts, which have been used before to assemble functional Hsp90 complexes (61). Recombinant survivin added to untreated reticulocyte extracts formed a complex with

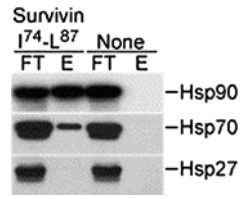


Fig. 4.6. Identification of shepherdin: Affinity chromatography. S100 Raji cell extracts were fractionated over control resin or immobilized survivin peptide I74–L87. Flow-through (FT) or eluted (E) material was analysed by Western blotting, from (2).

endogenous Hsp90, as detected by coimmunoprecipitation and Western blotting **(Fig. 4.7)**. In contrast, a control IgG did not immunoprecipitate survivin from reticulocyte extracts **(Fig. 4.7)**. Preincubation of reticulocyte extracts with shepherdin completely inhibited the association of recombinant survivin with Hsp90, whereas a scrambled peptide had minimal effect **(Fig. 4.7)**.

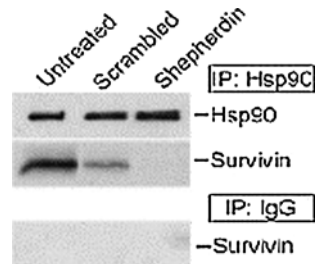


Fig. 4.7. Inhibition of survivin–Hsp90 interaction. Aliquots of rabbit reticulocyte extracts were *left* untreated or incubated with shepherdin or scrambled peptide, mixed with recombinant survivin, and immunoprecipitated with control IgG or an antibody to Hsp90. The immune complexes were analysed by Western blotting.

Next, we asked whether shepherdin interfered with ATP binding to Hsp90 (62). Addition of a scrambled peptide to recombinant N-domain of Hsp90 (residues 1–272) did not affect binding to ATP–sepharose, as shown by affinity chromatography (**Fig. 4.8**). In contrast, shepherdin blocked the interaction between N-Hsp90 and immobilized ATP, which was also fully competed out by addition of soluble ATP (**Fig. 4.8**).

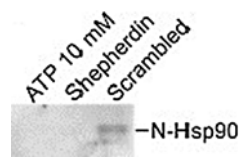


Fig. 4.8. Peptide competition of Hsp90–ATP binding. Recombinant Hsp90 N-domain was mixed with soluble ATP, shepherdin, or scrambled peptide and incubated with γ-phosphate-linked ATP–sepharose, and bound material was eluted in 5% SDS and analysed by Western blotting.

To map the requisites of shepherdin binding to Hsp90, we generated mutant peptides with single amino acid substitutions. Mutagenesis of H80, S81, S82, G83, and C84 abolished shepherdin binding to Hsp90, whereas substitution of K79, A85, F86, or L87 had no effect **(Fig. 4.9**).

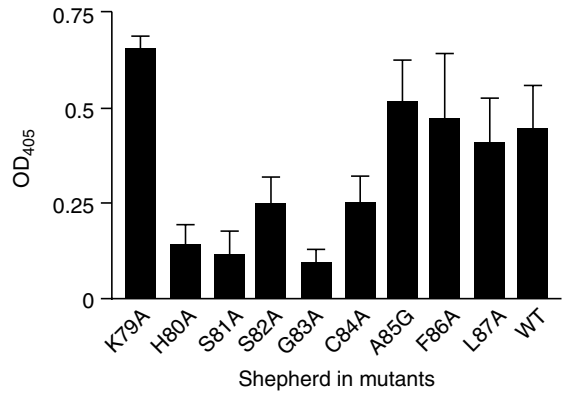


Fig. 4.9. Peptide mutagenesis. Wild-type or mutant shepherdin peptides were immobilized on plastic microtiter plates and Hsp90 binding was detected by ELISA. Data are the mean ± SD of two independent experiments. WT, wild-type shepherdin.

For its potent and broad antitumor activity, selectivity of action in tumor cells versus normal tissues, and inhibition of tumor growth in vivo without toxicity, shepherdin (*K79–L87*, *KHSSGCAFL*) and its retroinverso analog shepherdin-RV may offer a promising approach for rational cancer therapy. We will then show how we could identify the minimum sequence of shepherdin, labeled shepherdin[79–83] (*K79–G83*, *KHSSG*) required for activity in acute leukemia cancer cells. The structures of these peptides are studied by means of long time-scale MD simulations in explicit water. Subsequently, the dominant structures are docked to Hsp90, and the resulting complexes are also relaxed by means of long time-scale MD simulations to identify at equilibrium the dominant interactions responsible for binding. Finally, we will describe the use of the information developed in this part to identify a new non-peptidic small molecule which represents the prototype for a new class of compounds which can selectively inhibit Hsp90s chaperone activity. Computational and theoretical results will be benchmarked by experimental validations in vitro and in vivo.

4.1.2. Simulation Setup

Peptide Molecular Dynamics simulations. The peptides that have been simulated comprised the survivin sequence K79–K90 (*KHSSGCAFLIVK*) (3), the survivin sequence K79–L87

(*KHSSGCAFL*, shepherdin) and shepherdin-RV (*LFACGSSHK*, all D-aminoacids), and the minimal peptide K79–G83 (*KHSSG*). The C- and N- termini of each peptide were capped to avoid electrostatic artifacts due to the attractions between free opposite charges at the C- and N-termini. The side chain of Lys (K79 in shepherdin and shepherdin-min and *K87* in shepherdin-RV) was considered to be protonated, bearing a net charge of +1. The charge state of each peptide molecule is consistent with the solution conditions of the experiments. Each peptide was solvated with water in a periodic truncated octahedron, large enough to contain the peptide and 0.9 nm of solvent on all sides. All solvent molecules within 0.15 nm of any peptide atom were removed. Two Cl^- counterions were added to the *K79–K90* system, while one Cl^- counterion was added to shepherdin, shepherdin-RV, and shepherdin-min to ensure electroneutrality of the solution. Different initial conditions and conformations were used for the different peptide systems studied. For peptide *K79–K90*, the starting structure was totally extended to avoid biases in the conformational search. The simulation (production run) was of 100 ns. Shepherdin was subjected to two 100 ns long molecular dynamics (MD) simulations (for a total of 200 ns), starting from either the totally extended conformation (all backbone dihedrals set to 180°) or the conformation the peptide has in the survivin crystal structure (5). Shepherdin-RV, being the most active agent in vivo, was simulated for longer time-scales by combining five different simulations starting with different initial velocities on the atoms of the peptide and the solvent, obtained from a Maxwellian velocity distribution at the desired temperature of 300 K. The initial conformation is in all cases completely extended. Four simulations were 100 ns long, while one was 53 ns long for a total time of 453 ns. Shepherdin[79–83] was simulated for 200 ns. The set-up of each simulation is summarized in **Table 4.1**. Each system was initially energy minimized with a steepest descent method for 1000 steps. In all simulations the temperature was maintained close to the intended value of 300 K by weak coupling to an external temperature bath (63) with a coupling constant of 0.1 ps. The peptide and the rest of the system were coupled separately to the temperature bath. The *GROMOS96* force field (16) was used. The simple point charge (SPC) (64) water model was used. The LINCS algorithm (65) was used to constrain all bond lengths. For the water molecules the SETTLE algorithm (66) was used. A dielectric permittivity, $\varepsilon = 1$, and a time step of 2 fs were used. A cut-off was used for the calculation of the non-bonded van der Waals interactions. The cut-off radius was set to 0.9 nm. The calculation of electrostatic forces utilized the PME implementation of the Ewald summation method. All atoms were given an initial velocity obtained from a Maxwellian distribution at the desired initial temperature of 300 K. In each simulation, the density of the

Table 4. 1
Summary of MD simulation calculated for "Shepherdin" peptide

System	Shepherdin-RV					Shepherdin	
Simulation n°	1	2	3	4	5	1	2
Time (ns)	100	100	100	53	94	100	100
Description of Initial Conformation	Estesa	Estesa	Estesa	Estesa	Estesa	Estesa	Conformazione in Survivin
Temperature (K)	300	300	300	300	300	300	300
Number of Solvent Molecules	2842	2842	2486	2486	2838	2645	1355
System	**Shepherdin-RV C84A**					**Shepherdin-RV C84A**	
Simulation n°	1		2			1	2
Time (ns)	160		100			130	100
Description of Initial Conformation	Shepherdin-RV		Extended			Shepherdin-RV	Extended
Temperature (K)	300		300			300	300
Number of Solvent Molecules	755		2866			671	2876

system was adjusted performing the first equilibration runs at NPT condition by weak coupling to a bath of constant pressure (P_0 = 1 bar, coupling time $\tau_P = 0.5$ ps) (63). All simulations, starting from the appropriate peptide geometry, were equilibrated by 50 ps of MD runs with position restraints on the peptide to allow relaxation of the solvent molecules. These first equilibration runs were followed by other 50 ps runs without position restraints on the peptide. The production runs using NVT conditions, after equilibration, were 50 ns long for all of the complexes. All the MD runs and the analysis of the trajectories were performed using the GROMACS software package (67). Conformational cluster analysis of the combined trajectories for shepherdin (200 ns) or shepherdin-RV (453 ns) and shepherdin-min was performed using the method described in Daura et al. (68): count number of neighbors using a cut-off of 0.15 nm root mean square deviation (RMSD) between the optimal backbone superposition of different structures, take structure with largest number of neighbors with all its neighbors as cluster and eliminate it from the pool of clusters. This procedure is repeated for the remaining structures in the pool. The most

populated clusters, corresponding to the most visited structures in the MD simulations, for shepherdin and shepherdin-RV comprised β-hairpin type of structures (**Fig. 4.10**).

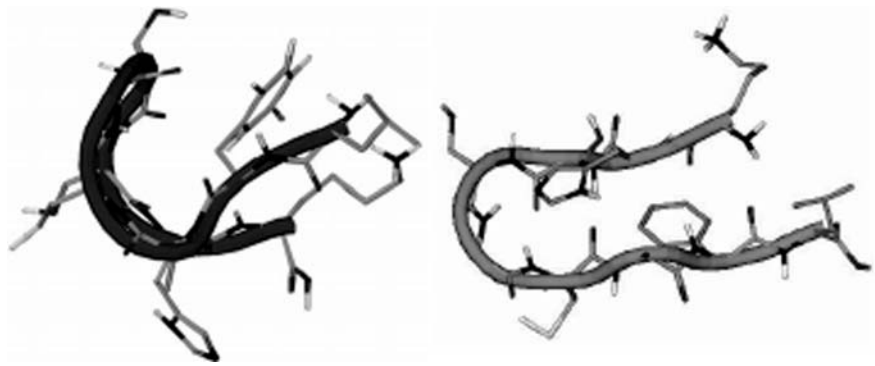


Fig. 4.10. On the *left* to *right* the most representative structure for shepherdin-RV and shepherdin.

The most populated structural cluster for shepherdin[79–83] was in contrast characterized by an extended structure. The representative structures (dominant structures of the peptide in solution) of the most populated cluster were used for docking experiments on Hsp90.

Single-point mutants of shepherdin-RV are produced with the Swiss Pdb Viewer program (69), and MD simulations on the mutants are carried out following the same protocol as described in the previous paragraph. Simulated sequences are:

LEU(87)-PHE-ALA-CYS-GLY-SER-SER-**HIS**-LYS(79)
LEU(87)-PHE-ALA-**CYS**-GLY-SER-SER-HIS-LYS(79)

Docking procedure. The β-hairpin structures of shepherdin and shepherdin-RV, and the extended structure of shepherdin-min were subjected to blind docking experiments on the N-terminal domain of *Hsp90* using the program Autodock (7). The crystal structure of the protein was taken from the protein data bank (pdb code 1YET). The original X-ray structure contains the ligand geldanamycin (GA), which was removed from the active site to yield the apo-open form of Hsp90. To test the viability of the docking procedure, and its ability to reproduce the experimental structure, a blind docking procedure was initially run on GA. The ligand was removed from Hsp90 and a docking experiment was run with no information on the position of the binding site. The minimum free energy structure determined in this experiment is exactly superimposable to the X-ray derived structure described (70). The docking procedure used can be summarized as follows. Mass-centered grid maps were generated with 0.25 Å spacing by

the program Autogrid for the whole Hsp90 protein target. Lennard–Jones parameters 12–10 and 12–6 (default parameters in the program package) were used for modeling hydrogen bonding and van der Waals interactions, respectively. The distance dependent dielectric permittivity of Mehler and Solmajer (71) was used for the calculation of the electrostatic grid maps. The Lamarckian genetic algorithm (LGA) and the pseudo-Solis and West methods were applied for minimization using default parameters. The number of generations was set to 25 million in all runs, and the stopping criterion was, therefore, defined by the total number of energy evaluations. Random starting positions on the entire protein surface, random orientations, and torsions (flexible ligand only) were used for the ligand. For shepherdin 100 different runs were performed with the parameters described above. For shepherdin-RV, a total of 350 runs were performed. The results of docking runs were analyzed by the clustering procedure described by Hetenyi et al. (72) were classified by a two-step procedure. First, the docked conformations of the ligand peptides were listed in increasing energy order. The structure of the complex that corresponded to the global minimum energy was used as the starting point of the first refinement molecular dynamics run. Second, the ligand conformation with the lowest energy was used as a reference, and all conformations with a center of mass to center of mass distance of less than 3 Å from the reference were taken to belong to the first class. After a ligand was assigned to a class, it was not used again for other classes. The process was then repeated for all hitherto unclassified conformations until all conformations were put in a class (**Fig. 4.11**).

The representative structure of the most populated class was then used for the second molecular dynamics refinement run. The refinement molecular dynamics runs of the complexes obtained after the Autodock runs were each 70 ns long.

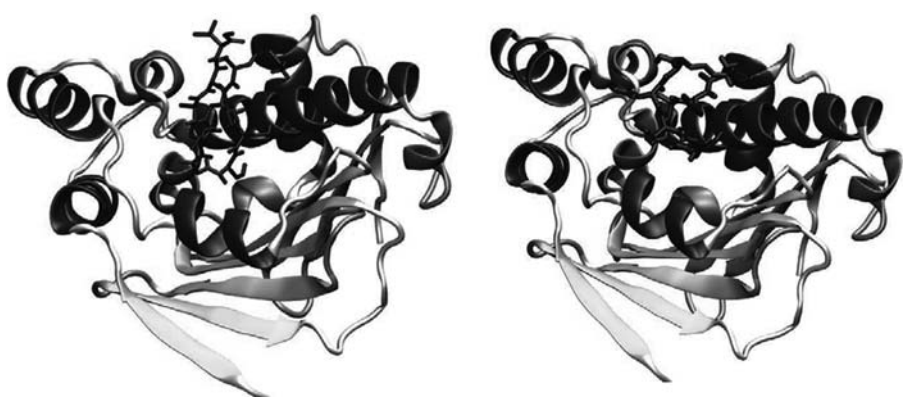


Fig. 4.11. *Right*, the energy minima from docking simulation for shepherdin and shepherdin-RV, on *left*.

4.1.3. Results

Shepherdin and shepherdin-RV. The peptides of the shepherdin series were identified as described in the last paragraph of the Introduction to this chapter. The modeling study of the peptide *K79–L87* named shepherdin and its retroinverso version *L87–K79* (all D aminoacid) shepherdin-RV start with a long molecular dynamic simulation with the target to identify the characteristic conformations of these peptides in solution. Analysis of the trajectories predicts that both shepherdin and shepherdin-RV have a dominant configuration characterized by a turn involving *S82–G83* in shepherdin and *G83–S84* in shepherdin-RV, and overall β-hairpin geometry (**Fig. 4.12**). The most populated conformation of shepherdin-RV shows a higher degree of compactness, with the aromatic ring of *F80* packing on the turn region (**Fig. 4.12**). The representative β-hairpin conformations of both peptides were subjected to multiple docking experiments on Hsp90 using the AutoDock program package (7). In all cases, the peptides were predicted to dock into the ATP binding site of Hsp90 (**Figs**. **4.11** and **4.12**). The geometry of the final complex is highly correlated with that of the complex between Hsp90 and GA (70), with the turn region of the peptides closely tracing the ansa ring backbone of GA **(Fig. 4.14**).

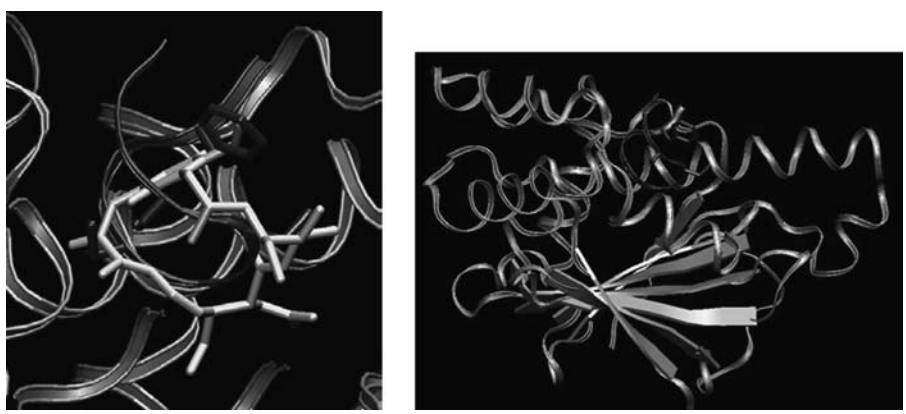


Fig. 4.12. On the *left* orthogonal views of shepherdin-RV (ribbon) in complex with Hsp90 as obtained by docking simulations. The peptide is overlaid onto the crystal structure of GA (licorice representation in light gray) in complex with Hsp90. On the right views of the dominant structure of shepherdin-RV (ribbon) as obtained by MD simulations are shown overlaid onto the crystal structure of Hsp90.

Shepherdin and shepherdin-RV make 13 and 18 predicted hydrogen bonds with the ATP pocket of Hsp90, respectively, involving the side chains of *H80, S81, S82*, the carbonyl group of *G83*, and the side chains of *K87* and *C82* (shepherdin-RV). Except for *D93*, the complementary residues of Hsp90 predicted to make contact with shepherdin and/or shepherdin-RV largely overlap with amino acids implicated in GA binding (70), including *S113*,

which has been recently shown to contribute to stepwise accessibility of the ATP pocket of *Hsp90* to GA (73). Shepherdin and shepherdin-RV are predicted to assume more extended conformations than GA in the Hsp90 pocket (**Figs. 4.11** and **4.12**), and bury a solvent accessible surface of 498 and 546 $Å^2$, respectively, as opposed to 402 $Å^2$ buried by GA.

To check these structural predictions, and validate experimentally that shepherdin engaged Hsp90 differently from GA, we introduced targeted mutations in the ATP pocket of Hsp90, and tested their effect on shepherdin binding. Individual substitution of N51, S52, D102, or S113 in the N-domain of Hsp90 reduced binding to shepherdin by 20–60%, whereas mutagenesis of "GA-specific" D93 had no effect, and a scrambled peptide did not bind wild-type or mutant Hsp90 (**Fig. 4.13**). Next, we compared the direct binding of shepherdin to recombinant N or C domain of Hsp90. Shepherdin associated with the N-domain Hsp90 in a concentration-dependent and saturable manner (**Fig. 4.9**). In contrast, no specific binding of shepherdin to the C domain of Hsp90 was demonstrated, as compared with scrambled peptide.

Simulations of shepherdin-RV mutants. To investigate the impact of single point mutation on the structure–activity relationship properties of shepherdin-RV, and to shed more light on the

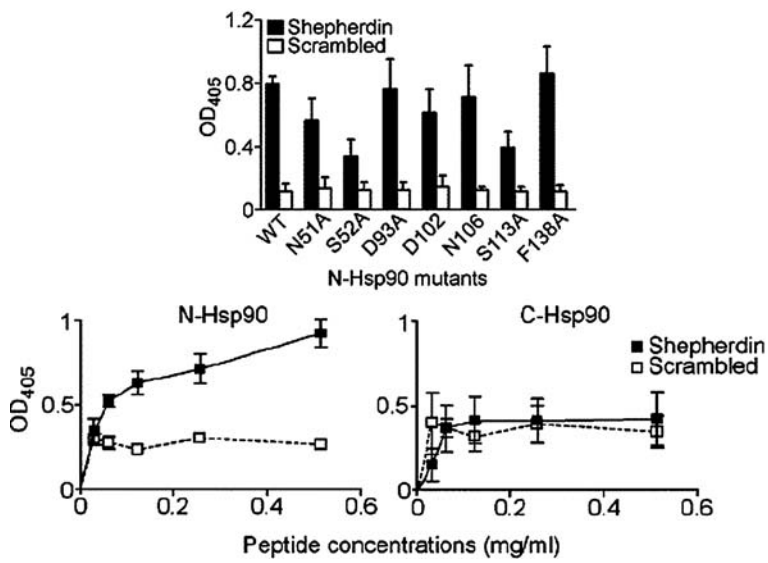


Fig. 4.13. *Top panel*, Hsp90 mutagenesis. The indicated mutants of the N-domain of Hsp90 were tested for binding to shepherdin or scrambled peptide. Data are the mean ± SEM of three independent experiments. WT, wild-type N-Hsp90. *Bottom line*, domain-specific interactions. Increasing concentrations of shepherdin or scrambled peptide were immobilized on microtiter wells and incubated with recombinant Hsp90 N or C domain, and protein binding was detected with domain-specific antibodies by ELISA. Data are the mean ± SEM of four independent experiments.

determinants of the interaction between the peptide and Hsp90, two mutants peptides (*H80A* and *C84A)* were simulated with long time-scale all-atom MD simulations. A total of four simulations (two runs for each mutant) are calculated for the two peptide mutants. Two different initial conformations were used: one completely extended and the second one from the dominant shepherdin-RV conformation found in the previous runs, and subjected to mutation (*see* **Table 4.1**). The purpose of the first simulation is to identify the characteristic conformations of these peptides in solution and the stability of the shepherdin-RV β-hairpin conformation after the mutation (**Fig. 4.14**).

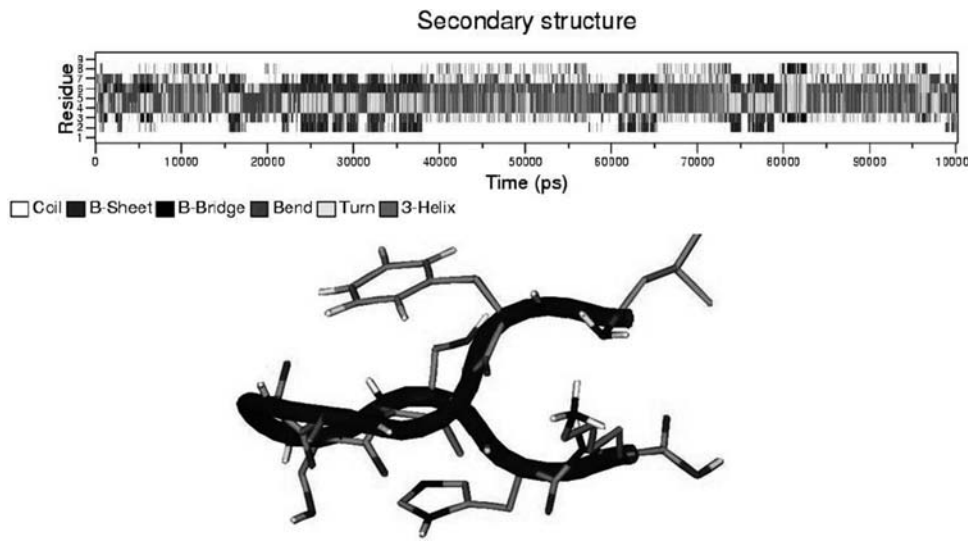


Fig. 4.14. *Top panel*, the behavior of secondary structure for shepherdin-RV; *bottom line*, the representative structure for the same simulation.

In the case of the *C84A* mutant, simulations suggest that the mutation dramatically decreases the tendency of the peptide to form a stable hairpin like structure. In the 100 ns time span from the completely extended conformation neither the analysis of the time evolution of secondary structure (*see* **Section 5.2** for more details, and (15)) nor the structural cluster analysis is able to identify a hairpin conformation similar to that observed for the original sequence (**Fig. 4.15**).

The second simulation, from a preformed hairpin structure, shows that the mutant peptide retains the hairpin conformation for about 10 ns and after that the turn geometry changes for a long period, before complete loss of the initial conformation. An analogous behavior is observed for *H80A*mutant peptide. Both the analysis of the time evolution of secondary structure and the

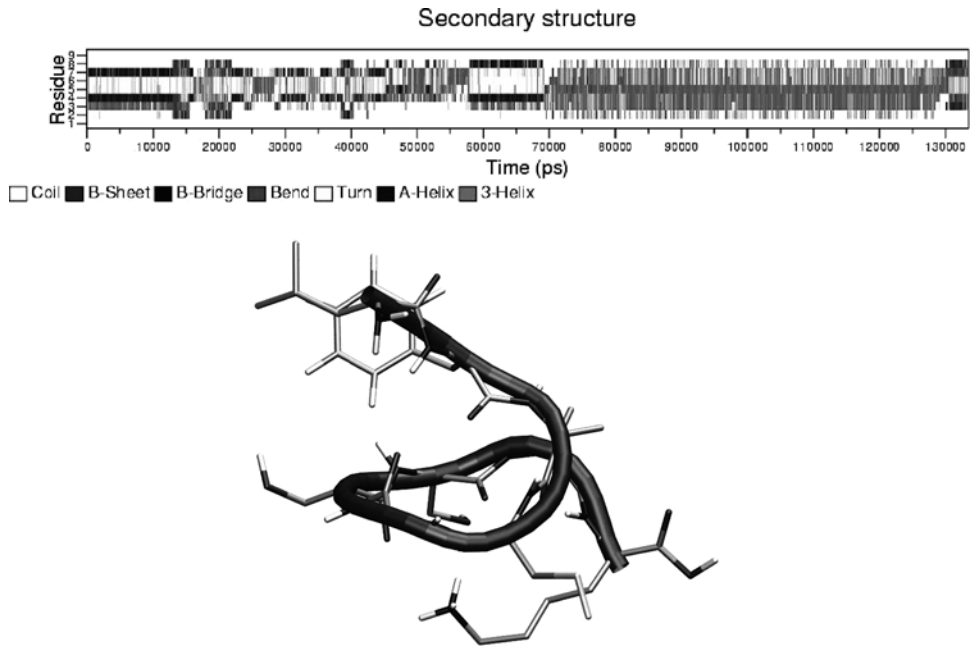


Fig. 4.15. *Top panel*, the behavior of secondary structure for shepherdin-RV mutation C84A; *bottom line*, the representative structure for the same simulation.

structural clustering suggest that the hairpin is not the dominant conformation in solution, despite being present for a smaller percentage. These results clearly suggest that the bent, hairpin like conformation of shepherdin is a fundamental determinant for recognition with the active site of Hsp90. It is worth noting at this point, that this type of conformation for the same sequence is also present in the native structure of shepherdin, suggesting a certain level of structural pre-organization for this sequence, optimized for binding to Hsp90.

Shepherdin[79–83]. The combination of theoretical analysis and experimental verification described in the previous paragraphs suggest that the minimal motif necessary for recognition should contain the HSSG sequence. Based on this considerations, a new short peptide, with sequence *KHSSG*spanning residues 79–83 of survivin was synthesized and named *shepherdin[79–83].*

To understand possible structure–activity relationships of this peptide and of its interaction with Hsp90, shepherdin[79–83] was simulated in isolation in explicit water. The peptide did not populate any preferred ordered secondary structure, and so the hydrophilic side chains and the backbone carboxyl and amino groups tended to maximize their interactions with the surrounding water solvent. Cluster analysis of the 200 ns simulations determined that the main conformational family of shepherdin[79–83] was characterized by a slight bend geometry involving residues His-80, Ser-

81, and Ser-82. Docking experiments on Hsp90 with this geometry predicted that shepherdin[79–83] bound to the ATP pocket of Hsp90 (**Fig. 4.16**). Two different orientations of shepherdin[79–83] were observed: one that corresponded to the global free-energy minimum structure of the shepherdin[79–83]–Hsp90 complex and one that represented the most frequently obtained structure after statistical clustering of all the structures studied during the Autodock simulations. The sites of contact between Hsp90 and shepherdin[79–83] in either configuration overlapped (**Fig. 4.14**). In the global free-energy minimum configuration (**Fig. 4.16**), the side chain of His-80 in shepherdin[79–83] made hydrophobic contacts with Ile-96 and hydrogen bonded with Gly-97 in Hsp90, the side chain of Ser-81 in shepherdin[79–83] hydrogen bonded with the side chains of Asp-102 and Asn-106 in Hsp90, and Ser-82 in shepherdin[79–83] hydrogen bonded with Asn-51 and Phe-138 in Hsp90. In the most frequently obtained shepherdin configuration (**Fig. 4.16**), His-80 in shepherdin[79–83] formed a hydrophobic interaction with Ile-96 in Hsp90 but was also involved in a new hydrogen bonding interaction with the side chain of Asp-54 in Hsp90; Ser-81 interacted with Asp-93 and Asn-106 in Hsp90, and Ser-82 interacted with Asn-106 and Asp-102 in Hsp90. Consistent

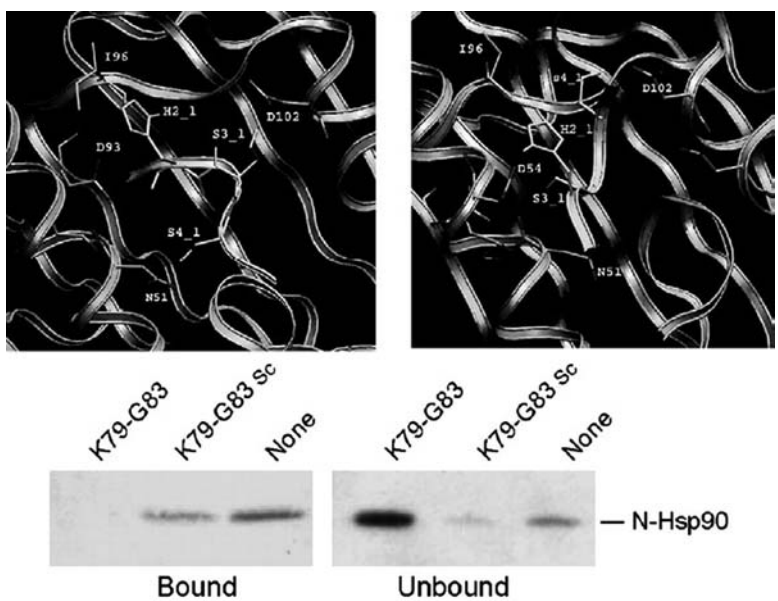


Fig. 4.16. *Top panel*, molecular dynamics simulations. Orthogonal views of the global free-energy minimum structure of the shepherdin[79–83]–Hsp90 complex (*left*) and the most populated conformational family of all structures of shepherdin[79–83] in complex with Hsp90 (*right*) were obtained from docking simulation studies. *Bottom line*, inhibition of Hsp90–ATP binding. Recombinant Hsp90 N-domain peptides (N-Hsp90) were mixed with soluble ATP, shepherdin[79–83], or its corresponding scrambled (Sc) peptide and incubated with γ-phosphate–linked ATP–sepharose for 2 h at 4°C. Bound or unbound material was then eluted in 5% sodium dodecyl sulfate and examined by Western blot analysis with an antibody to the N-domain of Hsp90. Bands were visualized by chemiluminescence.

with these molecular dynamics predictions, in biochemical experiments, shepherdin[79–83] efficiently displaced ATP binding from the N-domain of recombinant Hsp90, whereas the scrambled peptide was ineffective (**Fig. 4.16**).

4.2. Discussion

In this study, we used structure-based rational studies to identify and characterize shepherdin, a novel anticancer peptidomimetic modeled on the survivin–Hsp90 binding interface (3). All theoretical predictions were subjected to experimental verification and the activities of the peptides were evaluated both in vitro and in vivo, in a large multidisciplinary effort. Shepherdin engages the ATP pocket of Hsp90 with unique binding characteristics, destabilizes survivin several additional client proteins, and causes massive killing of tumor cells by apoptotic and non-apoptotic mechanisms (**Fig. 4.17**).

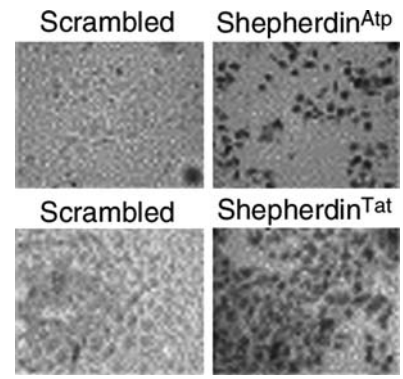


Fig. 4.17. Loss of plasma membrane integrity. HeLa cells were incubated with shepherdin^{Atp} (150 μM), shepherdin^{Tat} (50 μM), or cell-permeable scrambled peptides at identical concentrations, stained for Trypan blue exclusion after 4 h, and analyzed by light microscopy.

Shepherdin is selective in its antitumor activity, and does not affect the viability of normal cells or tissues, including human hematopoietic progenitors (**Fig. 4.17**).

When administered in vivo, shepherdin is safe and well tolerated, and inhibits growth of different tumor cell types without systemic or organ toxicity (**Figs. 4.17** and **4.18**). Taken together, these features may make shepherdin an attractive lead prodrug for "targeted" cancer therapy (**Fig. 4.19**).

Although initially designed as a high-affinity ($K_D \sim 80$ nM) inhibitor of the survivin–Hsp90 interaction, the data presented here suggest that shepherdin may function as a more global antagonist of Hsp90 chaperone activity. This conclusion is based on the structure–function analysis of shepherdin, and in particular its ability to expansively engage the chaperone ATP pocket,

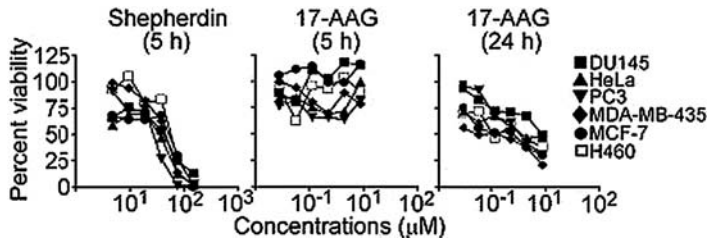


Fig. 4.18. Comparison with 17-AAG. The indicated tumor cell lines were treated with shepherdin^{Atp} or 17-AAG and analyzed for cell viability after 5 or 24 h by MTT.

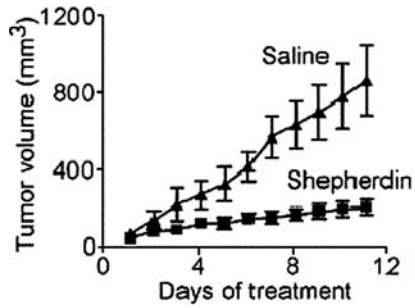


Fig. 4.19. Kinetics of prostate cancer xenograft growth. SCID/beige mice carrying human PC3 xenograft tumors were treated with daily i.p. injections of saline or cell permeable shepherdin-RV (50 mg/kg) for the indicated time intervals (5/6 animals/group). Tumor volume was measured with a caliper. Data are expressed as mean ± SEM.

compete for the Hsp90–ATP complex, and destabilize multiple Hsp90 client proteins in addition to survivin, in vivo. Because of these features, shepherdin appears ideally suited to interfere with the periodicity of Hsp90 ATPase cycles, by directly preventing ATP binding (70), and/or by competing with cochaperone recruitment, especially that of $p50^{cdc37}$, which is required for ATPase activity and shares overlapping binding contacts with shepherdin (74). In this context, the simultaneous destabilization of survivin levels (75), combined with the acute collapse of Hsp90 function (76), would be expected to cause a general breakdown of multiple cell proliferation and cell survival pathways in tumor cells, suitable for therapeutic exploitation (77).

Consistent with this model, a brief exposure of disparate tumor cell lines to cell-permeable variants of shepherdin resulted in massive and complete cell killing via activation of non-apoptotic and apoptotic mechanisms, the latter involving mitochondrial dysfunction, i.e., permeability transition. The concentrations of shepherdin needed to achieve complete tumor cell killing ($IC_{50} \sim 25$–75 μM, depending on the cell penetrating sequence), are in line with those of previous studies

using cell-permeable carrier sequences (78, 79)), and reflect the efficiency of intracellular penetration rather than the affinity of the mimetic for the target. The ability of shepherdin to trigger both apoptotic and non-apoptotic cell death may explain its broad antitumor activity, which indistinguishably affected tumor cell types of different derivations, regardless of their proliferative condition (80), p53 status (81), or overexpression of potent survival signals, e.g., Bcl-2 (82), all conditions that typically compromise the efficacy of conventional or "targeted" anticancer therapy (83). In addition, shepherdin kills tumor cells far more rapidly and more potently than other Hsp90 antagonists currently in the clinic, e.g., 17-AAG (84). As demonstrated by affinity chromatography, this higher antitumor efficacy of shepherdin can not be explained with nonspecific recognition of multiple cellular proteins, and does not involve an independent specificity for Hsp70, a regulator of mitochondrial apoptosis and caspase-independent cell death (85). Conversely, it seems plausible that by virtue of its unique binding interface with Hsp90, shepherdin may disrupt additional, and potentially as yet unrecognized, cell survival functions of the chaperone that are not efficiently compromised by GA or 17-AAG.

When tested as an anticancer agent in tumor models, shepherdin was selective and well tolerated, sparing normal cells, preserving colony-forming ability of purified human hematopoietic progenitors, and causing no organ or systemic toxicity after prolonged administration in vivo. This desirable selectivity may involve the differential overexpression of survivin and Hsp90 in tumors as opposed to normal tissues (75, 76)), as well as qualitative changes in Hsp90, which exhibits a ~100-fold higher affinity for binding ATP pocket antagonists in tumor versus normal cells. Previously demonstrated for 17-AAG (86), a similar paradigm can be extended to shepherdin, which bound avidly to Hsp90 in lymphoma cells, but not at all in normal human mononuclear cells.

Experimental tests with the minimal sequence of five residues also confirm theoretical hypotheses. The results of MD simulations on the structure of shepherdin[79–83] and of its complex with Hsp90 were challenged with competition experiments by use of enzyme-linked immunosorbent assay (ELISA). Apoptosis, Hsp90 client protein expression, and mitochondrial dysfunction were evaluated in acute myeloid leukemia (AML) types (myeloblastic, monocytic, and chronic myelogenous leukemia in blast crisis), patient-derived blasts, and normal mononuclear cells. Effects of shepherdin[79–83] on tumor growth were evaluated in AML xenograft tumors in mice ($n = 6$). Organ tissues were examined histologically. Taken together, these results showed that shepherdin[79–83] bound to Hsp90, inhibited formation of the survivin–Hsp90 complex, and competed with ATP binding to Hsp90. Cell-permeable

shepherdin [79–83] induced rapid (within 30 min) and complete (with concentrations inducing 50% cell death of 24–35 μM) killing of AML types and blasts, but it did not affect normal mononuclear cells. Shepherdin[79–83] made contact with unique residues in the ATP pocket of Hsp90 (Ile-96, Asp-102, and Phe-138), did not increase Hsp70 levels in AML cells, disrupted mitochondrial function within 2 min of treatment, and eliminated the expression of Hsp90 client proteins. Shepherdin[79–83] abolished growth of AML xenograft tumors (mean of control group = 1698 mm3 and mean of treated group = 232 mm3; difference = 1466 mm3, 95% confidence interval = 505.8 to 2426; $P = 0.008$) without systemic or organ toxicity and inhibited Hsp90 function in vivo. The combination of these data show that shepherdin[79–83] is a novel Hsp90 inhibitor with a unique mechanism of anticancer activity.

In summary, shepherdin has the molecular features of both an inhibitor of a critical protein–protein interaction in tumor cells, e.g., survivin–Hsp90 (3), and an enzymatic antagonist of Hsp90 ATPase cycles (76). Because of these combined features, plus its considerably higher potency compared to other Hsp90 inhibitors, e.g., 17-AAG, shepherdin may provide a potent and selective new anticancer agent in humans, consistent with the use of peptidomimetics in targeted cancer therapy (87). In addition, we narrowed the shepherdin binding interface to a short stretch of amino acids between H80 and C84 in the survivin sequence. Previously, mutagenesis of H80 (88) or C84 (89) resulted in dominant negative phenotypes with mitotic defects and induction of apoptosis in tumor cells, thus further underscoring their critical roles in survivin function. This small cluster of residues may thus provide a manageable platform for further derivatization of shepherdin, as well as for chemical screenings to identify shepherdin-like small molecules with enhanced, "targeted" anticancer activity in humans.

5. Rational Identification of a New Non-peptidic Anticancer Lead, Based on Peptide–Hsp90 Recognition

5.1. Introduction

As we stated in the previous paragraph, cancer therapy now aims at disabling oncogenic pathways that are selectively operative in tumor cells, so as to spare normal tissues and limit side effects in humans. This "targeted therapy" relies on a better understanding of cancer genes, particularly those implicated in tumor cell proliferation and survival (90). Accordingly, targeted inhibition of the Bcr-Abl kinase with small molecule antagonists has produced dramatic clinical responses in malignancies driven by this oncogene (91–93). However, such approaches may not be immediately available for the majority of tumors where multiple molecular abnormalities and genetic instabilities may elude the

identification of one single, disease-driving oncogene (90). Conversely, pathways that intersect multiple essential functions of tumor cells may provide wider therapeutic opportunities. A prime target for this strategy is the heat shock protein 90 (Hsp90), a molecular chaperone that oversees the correct conformational development of polypeptides and protein refolding through sequential ATPase cycles and stepwise recruitment of co-chaperones. This adaptive pathway contributes to the cellular stress response to environmental threats including heat, heavy metal poisoning, hypoxia etc., and is extensively exploited in cancer, where Hsp90 ATPase activity is up-regulated ~100-fold (59). The repertoire of Hsp90 client proteins is restricted mainly to growth-regulatory and signaling molecules, especially kinases and transcription factors, which may contribute to tumor cell maintenance (59, 60). Therefore, targeted suppression of Hsp90 ATPase activity with a small molecule inhibitor, the benzoquinone ansamycin antibiotic 17-allylamino-17-demethoxygeldanamycin (17-AAG), has shown promising anticancer activity in preclinical models, and has recently completed safety evaluation in humans (77, 94). One Hsp90 client protein with critical roles in tumor cell proliferation and cell viability is survivin, an inhibitor of apoptosis (IAP) protein selectively overexpressed in cancer (3, 75). Accordingly, targeting the survivin–Hsp90 complex may provide a strategy to simultaneously disable multiple signaling pathways in tumors, and a peptidomimetic antagonist of this interaction structurally different from 17-AAG, shepherdin, inhibited the chaperone activity and exhibited potent and selective anticancer activity in preclinical models (2).

In this study, we have used shepherdin-RV (*LFACGSSHK*, all D-amino acids) as a scaffold to rationally identify low-molecular-weight compounds that may act as structurally novel Hsp90 antagonists. We built a 3D pharmacophore to screen a database of non-peptidic structures and we identified a novel antagonist of Hsp90 chaperone function with promising anticancer activity. Identifying non-peptidic leads has several medicinal chemistry advantages over the use of peptides: reduced immunogenicity, better bioavailability, and lower degradation rates in the cellular environment. The results presented here open the possibility to expand the molecular diversity space of Hsp90 antagonists.

5.2. Materials and Methods

5.2.1. Simulation Setup, Docking Experiments, and Molecular Dynamics (MD) Refinement of Hsp90–Shepherdin Complex

MD simulations of shepherdin in isolation in solution were described in *(2)*. The docking procedure can be summarized as follows: the representative of the main conformational clustering for shepherdin structures (β-hairpin) was subjected to blind docking experiments on the putative N-terminal Hsp90 receptor using the program AutoDock (7). The crystal structure of the protein was taken from the protein data bank (pdb code

1YET.pdb) (70). The original X-ray structure contains the ligand GA, which was removed from the active site to yield the apo-open form of Hsp90.

Mass-centered grid maps were generated with 0.25 Å spacing by the program Autogrid for the whole Hsp90 protein target. Lennard–Jones parameters 12-10 and 12-6 (the default parameters in the program package) were used for modeling H-bonding and van der Waals interactions, respectively. The distance-dependent dielectric permittivity of Mehler and Solmajer was used for calculation of the electrostatic grid maps (95). The Lamarckian genetic algorithm (LGA) and the pseudo-Solis and West methods were applied for minimization using default parameters. The number of generations was set to 25 million in all runs, and the stopping criterion was, therefore, defined by the total number of energy evaluations. Random starting positions on the entire protein surface, random orientations, and torsions (flexible ligand only) were used for the ligand. A total of 350 runs were performed. At the end of the docking runs, the conformations of the ligand peptide were listed in increasing energy order. Subsequently, the ligand conformation with the lowest energy was used as the reference, and all conformations with a center of mass to center of mass distance of <2.5 Å from the reference were assigned to the first cluster. Once a conformation was assigned to a cluster, it was not used again for other, energetically less favorable clusters. Then the process was repeated for all hitherto unclassified conformations until all conformations were put in a cluster, thus identifying 49 structural clusters corresponding to different free-energy values for the complex. Most of the docked structures shared common conformational characteristics, which are prototypically represented by the structure of the global minimum of the complex. This structure was chosen for further studies via all-atom, explicit solvent MD simulations.

The shepherdin–Hsp90 complex was first solvated with water in a periodic truncated octahedron, large enough to contain the peptide and 0.9 nm of solvent on all sides. The protonation and charge states of the side chains of the ligand and the receptor were chosen to be consistent with the solution conditions of the experiments: NH groups were considered as having a $+1$ charge and carboxylic groups as having a -1 charge. The system was found to have a total charge of -8. All solvent molecules within 0.15 nm of any peptide atom were removed. Eight Na^+ counterions were added to the system. Two different sets of initial velocities obtained from a Maxwellian velocity distribution at the desired temperature of 300 K were used to yield two different production runs of 54 and 73 ns, respectively.

In each case, the system was initially energy minimized with a steepest descent method for 1000 steps. In all simulations the temperature was maintained close to the intended value of 300 K

by weak coupling to an external temperature bath with a coupling constant of 0.1 ps. The Hsp90/shepherdin complex and the rest of the system were coupled separately to the temperature bath (63). The GROMOS96 force field (16, 17) and the simple point charge (SPC) (64) water model were used. The LINCS algorithm (65) was applied to constrain all bond lengths. For the water molecules the SETTLE algorithm (66) was used. A dielectric permittivity, $\varepsilon = 1$, and a time step of 2 fs were used. A cut-off was taken for the calculation of the non-bonded van der Waals interactions. The cut-off radius was set at 0.9 nm. The calculation of electrostatic forces utilized the PME implementation of the Ewald summation method (18). In each simulation, the density of the system was adjusted by performing the first equilibration runs in *NPT* condition by weak coupling to a bath of constant pressure (P_0= 1 bar, coupling time τ_P= 0.5 ps) (63). All simulations were equilibrated by 50 ps of MD runs with position restraints on the protein and ligand to allow relaxation of the solvent molecules. These first equilibration runs were followed by another 50 ps runs without position restraints on the solute The production runs using *NVT* conditions, after equilibration, were 54 and 73 ns long. All the MD runs and the analysis of the trajectories were performed using the GROMACS software package (67). Configurations of the receptor–ligand complex were saved every 4 ps for subsequent statistical analysis.

Conformational cluster analysis of the two trajectories (67) was performed using the method described by Daura et al. (96): count number of neighbors using a cut-off of 0.20 nm RMSD between the optimal backbone superposition of different structures, take structure with largest number of neighbors with all its neighbors as cluster, and eliminate it from the pool of clusters. This procedure is repeated for the remaining structures in the pool. The most populated cluster, representative of the most visited structures in the MD simulations, for the Hsp90/shepherdin complex was used as a template for pharmacophore design. The relative distances and orientations (dihedral angles) among the different groups were evaluated as averages over the whole trajectories. For dihedral calculations, the angles were defined as the solid angles in 3D space determined by the vectors that unite the O–H or S–H atoms in Ser or Cys, or by the vector normal to the imidazole plane of histidine.

5.2.2. Pharmacophore Generation

Three different pharmacophore models were built and labeled PHARM1, PHARM2 and PHARM3 based on the results of MD simulations. The conformation of shepherdin and the orientations of its side chain functional groups in the most populated structural cluster from MD trajectories of the complex were used as structural template. The distributions of dihedral values (**Table 4.2**) and distances (**Table 4.3**) among critical functionalities were used

Table 4. 2
Orientations of pharmacophoric groups. Minimum and maximum values of dihedral angles between functional groups used for PHARM1. The definitions of dihedral are reported in materials and methods

Functional groups	Minimum dihedral value (Degrees)	Maximum dihedral value (Degrees)
Ser84_γOH–Ser85_γOH	260.7	290.7
Ser84_γOH–Cys82_SH	247.0	277.0
Ser85_γOH–Cys82_SH	184.0	214.0
Cys82_SH–Imidaz_His86	27.2	57.2
Ser84_γOH–Imidaz_His86	133.7	163.7
Ser85_γOH–Imidaz_His86	43.9	73.9

Table 4. 3
Distances between pharmacophoric groups. Average distances and standard deviation from MD simulations used to define distances between functional groups in PHARM1

Functional Groups	Average Distance (nm)
Ser84_γOH–Ser85_γOH	0.71 ± 0.09
Ser84_γOH–Cys82_SH	0.90 ± 0.12
Ser85_γOH–Cys82_SH	0.63 ± 0.15
Cys82_SH–Imidaz_His86	0.77 ± 0.21
Ser84_γOH–Imidaz_His86	1.11 ± 0.09
Ser85_γOH–Imidaz_His86	0.52 ± 0.14

to define upper and lower boundaries for geometric constraints. The HypoGen Module of the Catalyst® program from Accelrys (97)was used for this purpose. For PHARM1 a four-point model was created by assembling 3 H-bond donors mapped over the two Ser γ -OH and the Cys SH groups, and one imidazole moiety mapped over the imidazole ring of His of shepherdin. The torsional and distance restraints reported in **Tables 4.1** and **4.2** were added to restrict the database search. Furthermore, we imposed

location constraints (the volume in which the functions can reside), specifying the radius of the spheres according to Catalyst's defaults. PHARM2 and 3 were constructed in a similar way, augmenting PHARM1 with an aromatic function centered on the position of the Phe80 benzene ring and a hydrophobic function centered on the S atom of Cys82 (PHARM2), or by the presence of a positive charge mapped on the position of the ammonium group of Lys87 (PHARM3). These were used as queries for a search in the NCI_3D database using the database search module of catalyst.

5.2.3. Small Molecule – Hsp90 Docking and MD Refinement of the Complex

The small molecule lead, 5-aminoimidazole-4-carboxamide-1-β-D-ribofuranoside (98) (AICAR), identified as active through the procedure described above and experimental tests in different tumor and normal cell lines, was docked on Hsp90 using the Autodock program and the same protocol as described for shepherdin. The clustering procedure identified one dominant set of conformations for the AICAR/Hsp90 complex, including the free-energy minimum structure. The latter structure was used as the starting structure for subsequent MD refinement of the complex in explicit water. MD was carried out once again following the procedure described above for the shepherdin/Hsp90 complex for a total of 100 ns. The resulting trajectory was analyzed in terms of the main interactions (hydrogen bonding, favorable contacts, etc.) taking place during the simulation time. This information can be used to define possible structure–activity relationships for the AICAR/Hsp90 complex with the aim to design derivatives with improved activities and pharmacological properties.

5.2.4. Recombinant Proteins and Binding Studies

The N-domain (residues 1–272) or C-domain (residues 629–732) of Hsp90 were expressed as recombinant fusion proteins and purified after removal of the GST frame by thrombin cleavage, as described previously. Binding of AICAR to recombinant Hsp90 fragments was carried out by ELISA. Briefly, increasing concentrations of AICAR (Sigma-Aldrich, St. Louis, MO, USA; 0.02–5 mM) were immobilized on plastic microtiter wells, blocked in 3% gelatin and further mixed with recombinant N- or C-domain of Hsp90 (1 μg/ml) for 2 h at 310 K (37°C). After washes, binding of the various Hsp90 fragments to AICAR-coated wells was detected using domain-specific antibodies to Hsp90, and quantified by absorbance at 405 nm. In another series of experiments, AICAR (5 mg/ml) was coupled to Sepharose beads (1 ml) and used to fractionate HeLa cell extracts prepared in 20 mM Tris, pH 7.2, 0.5% deoxycholate, 1% Triton X-100, 0.1% SDS, 150 mM NaCl, 1 mM EDTA plus protease inhibitors. After washes, bound proteins were eluted in 0.1 M glycine, pH 2.5, immediately

buffered to neutrality, separated on SDS polyacrylamide gels and analyzed by Western blotting. Fractionation of HeLa cell extracts over uncoupled sepharose beads was used as control.

5.2.5. Cell Lines and Cultures

Androgen-independent adenocarcinoma (DU145), cervical carcinoma (HeLa), and melanoma (JR8) cell lines were grown in RPMI-1640 medium (BioWhittaker, Verviers, Belgium) supplemented with 10% fetal bovine serum. The normal human lung fibroblast cell line (WI38) was grown in DMEM medium (BioWhittaker) supplemented with 5% fetal bovine serum. All cell lines were maintained as monolayers at 310 K (37°C) in a 5% CO_2 humidified atmosphere in air.

5.2.6. Growth Inhibition and Apoptosis Assays

After harvesting in the logarithmic growth phase, cells were seeded in six-well plates for 24 h and then exposed to various concentrations of AICAR or 17-AAG for 72 h at 310 K (37°C) in a 5% CO_2 humidified atmosphere. At the end of treatment, cells were washed twice with PBS, trypsinized, and counted in a particle counter (Coulter Counter, Coulter Electronics, Luton, UK). Each experimental point was run in triplicate. The results were expressed as the number of cells in treated samples compared with control samples.

For determination of apoptosis, cells were fixed in 70% (v/v) ethanol, stained with a solution containing 50 μg/ml propidium iodide, 10 mg/ml RNase, and 0.1% (v/v) Tween20 for 30 min at room temperature and analyzed with a FACScan flow cytometer (Becton Dickinson, Sunnyvale, CA, USA). A flow-cytometric sub-$G_{0/1}$ peak was detected on DNA plots using the CellQuest software according to the Modfit model (Becton Dickinson). In parallel, an aliquot of the propidium iodide-stained cell suspension was spotted onto slides and assessed for typical apoptotic nuclear morphology (nuclear shrinkage, condensation, and fragmentation) under a fluorescence microscope with appropriate filter combinations. Apoptotic cells were determined by two independent observers scoring at least 500 cells in each sample.

Caspase-3 and caspase-9 activities were measured as the ability to cleave the specific substrates *N*-acetyl-Leu-Glu-His-Asp-pNA (LEHD-pNA) and *N*-acetyl-Asp-Glu-Val-Asp-pNA (DEVD-pNA) by means of the APOPCYTO/caspase-9 and APOPCYTO/caspase-3 assay kits (Medical & Biological Laboratories, Naka-ku Nagoya, Japan) according to the manufacturer's instructions.

5.2.7. Analysis of Hsp90 Chaperone Function

To monitor changes in Hsp90 client proteins, cells were harvested and solubilized in lysis buffer (0.01% NP40, 10 mM Tris [pH 7.5], 50 mM KCl, 5 mM $MgCl_2$, 2 mM DTT, 20% glycerol plus protease inhibitors) plus four pulses on a sonicator (10 s each) at 50 J/W-s, alternated by 30 s intervals on ice, and analyzed by Western blotting primary antibodies specific for

survivin (Novus Biologicals, Littleton, UK), Hsp90, Hsp70, CDK-6, Raf-1, Neu (Santa Cruz Biotechnology, Santa Cruz, CA, USA) and Akt (Cell Signaling Technology, Danvers, MA, USA). Alternatively, lysates were immunoprecipitated with a rabbit polyclonal Hsp90 antibody (Santa Cruz Biotechnology) precoupled to 8 μl of a 50% slurry of protein-G-agarose beads for 1 h at 4°C. After washes the immunoprecipitates were analyzed for telomerase activity by telomeric repeat amplification protocol (TRAP) assay using the TRAPeze kit (Intergen Co., Oxford, UK). Each reaction product was amplified in the presence of a 36-bp internal TRAP assay standard (ITAS). Quantitative analysis was performed with the ImageQuanT software (Molecular Dynamics) as previously described *(99)*.

5.3. Results

5.3.1. Characterization of Hsp90/Shepherdin Binding Interface

The dominant conformations of shepherdin in solution were investigated through all-atom, explicit solvent MD simulations for a total time span of 400 ns. Statistical cluster analysis showed that shepherdin displays one main conformation, characterized by the presence of a turn involving residues G83–S84 and an overall hairpin geometry (*see* (2) for details). The remaining clusters were mainly extended conformations, with the peptide backbone groups involved in hydrogen bonding with water.

The most populated conformation was subjected to multiple blind docking experiments on Hsp90 using the Autodock program (7). In all cases, the peptide was predicted to bind within the ATP binding site of Hsp90. Analysis of the blind docking results through the procedure described by Hetenyi et al.(72). showed that low energy poses are all closely correlated with one another, with an RMSD from the global minimum structure lower than 2.5 Å. Control docking experiments were conducted with the extended structures representative of other clusters, but in those cases it was not possible to univocally identify any particular binding site on Hsp90.

The free-energy minimum structure of the Hsp90/shepherdin complex was then subjected to two long, 54 and 73 ns, all-atom MD simulations. Analysis of the statistical and time-dependent distribution of the interactions between functional groups of the ligand and of the chaperone was carried out to develop pharmacophore models, keeping into account the motional and flexibility properties of both the ligand and the receptor. Shepherdin partially reoriented to increase the total number of stabilizing contacts with the ATP binding pocket of Hsp90 (**Fig. 4.20**).

Attention was focused in particular on the analysis of hydrogen-bonding, hydrophobic/aromatic, and charge–charge interactions, as these represent the most common types of intermolecular forces determining host/guest recognition in drugs. The functional groups of shepherdin involved in direct or water-mediated

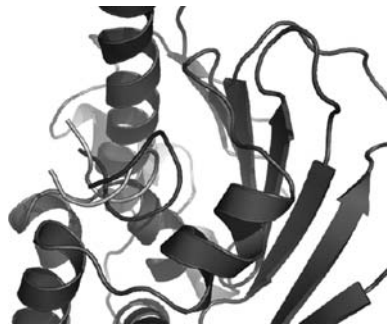


Fig. 4.20. Structure of the MD-refined docked complex between shepherdin and the N-domain of Hsp90. Initial (*light gray*) and final (*dark gray*) structure of shepherdin within the ATP-binding pocket of Hsp90 with two different orientations.

hydrogen bonds with the binding pocket of Hsp90 included the γ-OH functions of Ser84, Ser85 and the imidazole ring of His86 (**Fig. 4.21**). The latter, in particular, could satisfy hydrogen-bonding conditions being involved in interactions both as an acceptor (Nε atom) and as a donor (Nδ-H functional group) (**Fig. 4.21**). The remaining hydrogen-bonding group on shepherdin, Cys82,

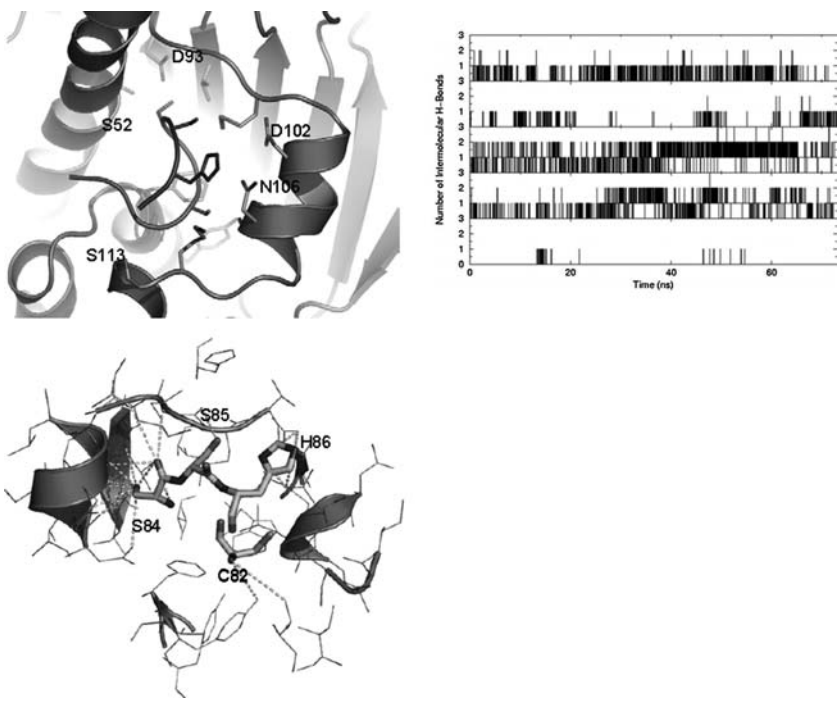


Fig. 4.21. *Top panel* on *right*, side chain groups of Hsp90 responsible for the main interactions with shepherdin. *Top panel* on *left*, close-up of the groups of shepherdin responsible for stabilizing interactions with Hsp90. *Bottom line*, time evolution of the H-bonding interactions between Hsp90 and shepherdin groups (from *top* to *bottom*) *g*-OH of S84, *g*-OH of S85, Ne of H86, Nd∼H of H86, and SH of C82.

was involved to a lesser extent in intermolecular H-bonding interactions; however, its presence was shown experimentally to be necessary to ensure binding. Moreover, it displayed hydrophobic interactions with the side chains of Hsp90 Leu108 and the alkyl part of Asn109. Cys82 is also important for preserving the hairpin structure: mutations to Ala on the isolated peptide lead to loss of ordered hairpin structure (2).

To define the presence of hydrophobic/aromatic interactions, the contacts involving the side chains of Phe80 and His86 were monitored during simulation. Shepherdin Phe80 was found to be in contact mainly with the charged/polar side chains of Lys59, Asn52 and Asn55 on Hsp90, while shepherdin His86 was not significantly involved in hydrophobic/aromatic contacts with Hsp90 residues. Finally, the role of the positively charged ammonium group on the side chain of shepherdin Lys87 was found to be only marginally involved in interactions with the backbone carbonyl oxygens of Hsp90 residues Phe135 and Gly136, being mostly exposed to the water solvent during MD simulations.

5.3.2. Pharmacophoric Hypotheses and Small Molecule Identification

Three different pharmacophore models were built and labeled PHARM1, PHARM2 and PHARM3 based on the results of MD simulations. The conformation of shepherdin and the orientations of its side chain functional groups in the most populated structural cluster from MD trajectories of the complex were used as structural template (**Figs. 4.20** and **4.21**). The distributions of dihedral values (**Table 4.2**) and distances (**Table 4.3**) among critical functionalities were used to define upper and lower boundaries for geometric constraints.

PHARM1 (**Fig. 4.22**) consisted of four pharmacophoric points: three H-bonding donor functionalities mapped over the side chain OH group of Ser84, Ser85 and the SH group of the Cys82, plus one imidazole ring moiety mapped on the position of

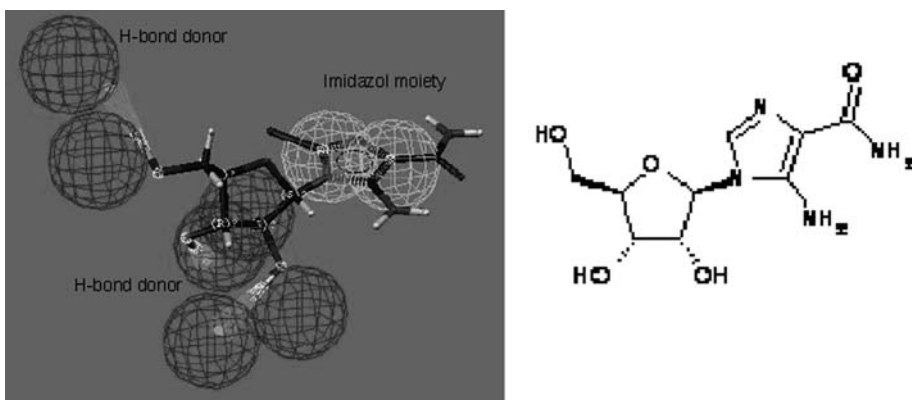


Fig. 4.22. The pharmacophoric hypochapter mapped on the 3D structure of AICAR and its molecular structure (5-aminoimidazole-4-carboxamide-1-β-D-ribofuranoside, AICAR).

the corresponding ring of His86 (**Fig. 4.25**). Each imidazole atom was allowed to bear any substituent or be a bridgehead in the presence of a fused ring. PHARM2 consisted of two H-bonding donor groups corresponding to the γ-OH of Ser84 and Ser85, one aromatic function centered on the position of the Phe80 benzene ring, and one hydrophobic function centered on the S atom of Cys82. PHARM3 had the same properties as PHARM1, augmented by the presence of a positive charge mapped on the position of the ammonium group of Lys87.

The three models described above were used as queries for a search of the NCI_3D database of molecules (containing approximately 160,000 compounds). The search with PHARM1 yielded 73 compounds, the search with PHARM2 yielded 42 compounds, while PHARM3 gave no hits. In experimental tests, the molecules corresponding to hits of PHARM2 proved to be extremely insoluble due to the presence of aromatic–hydrophobic groups and had thus no tumor-cell-killing effect. The search with PHARM1 yielded, among others, 20 hits reminiscent of the class of known purine-based inhibitors of the ATP-binding pocket of Hsp90 (100, 101). Interestingly, one of the non-purine-based hits that was found to be effective in experiments, AICAR (98)(**Fig. 4.22**), was not previously known to interfere with Hsp90 chaperone functions and was characterized by a novel molecular structure among Hsp90 antagonists.

5.3.3. Molecular Interactions Between AICAR and Hsp90: Molecular Dynamics Simulation Results

The structure of the complex between AICAR and Hsp90, obtained with the docking procedure described in **Section 5.2**, was studied in solution by MD simulations in explicit water solvent. The 100 ns time length reported for this simulation represents, to the best of our knowledge, the first case where such timescales have been accessed and used for the study of systems of this degree of complexity. Analysis of the trajectory shows that AICAR docks into the ATP-binding pocket of Hsp90 (**Fig. 4.23**)and that it is involved in several direct hydrogen-bonding contacts with the residues of the Hsp90 ATP-binding pocket. The five-member sugar ring fits nicely into a pocket flanked by negatively charged and hydrophobic regions (**Fig. 4.23**). The OH group in position 2 on the ribose points into a deep hydrophobic channel defined by the side chains of Leu48, Met98, Val136, Phe138, and Val186.

In this binding mode AICAR is also involved in direct hydrogen-bonding contacts with the side chain of Asp93, through the OH groups in positions 3 and 5 of the ribose ring (**Fig. 4.24**). Interestingly, the interaction with Asp93 and is characteristic of the interaction of the chaperone with the anticancer molecule radicicol and some of its derivatives (102). The hydroxyl group in position 3 is also involved as a hydrogen bond acceptor with the side chain of Ser52. The imidazole moiety points toward a negatively charged region of the Hsp90 ATP-binding side,

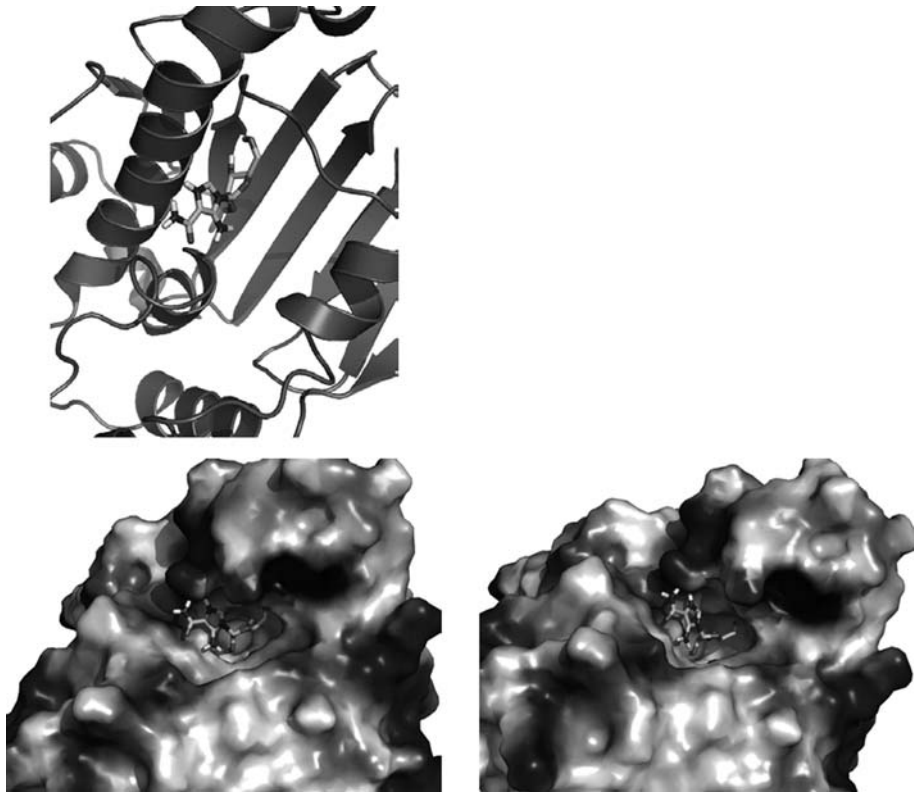


Fig. 4.23. *Top panel*, structure of the MD-refined docked complex between AICAR and the N-domain of Hsp90. The protein is represented by its secondary structure. *Bottom line*, structure of the MD-refined docked complex between AICAR and the N-domain of Hsp90. The ligand is represented by sticks, while the protein is represented by its Van der Waals surface, shaded according to its electrostatic potential: *black* indicates regions of negative charge, *white* indicates hydrophobic regions, and *gray* indicates positively charged regions.

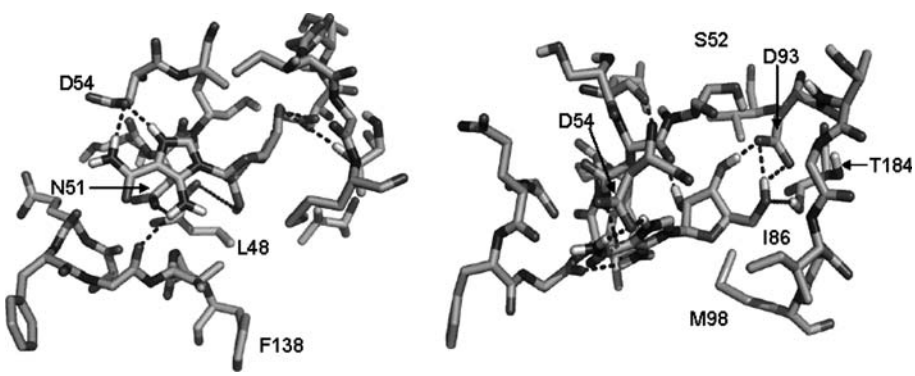


Fig. 4.24. Two different orientations of the Hsp90 binding site with AICAR bound. The main interactions between AICAR and Hsp90 residues are highlighted.

characterized by the presence of Asp54 and Asn51. The acidic hydrogens on the aromatic nitrogen atom in position 3 and on the carboxamide at position 4 of the imidazole are directly involved in stabilizing interactions with the carboxylate group of Asp54, while the NH_2 group on imidazole position 5 acts as a hydrogen bonding acceptor with respect to the side chain amide group of Asn51 and donates one hydrogen bond to the carbonyl group of Gly125.

Finally, it is worth noting that the regions of the protein which display the highest fluctuations and undergo secondary structure rearrangements include residues 65–75 and 105–112 and are directly in contact with the active site region of the protein. The latter adapts dynamically to the presence of the ligand, which in turn finds the optimal sterical and chemical fitting within the ATP binding pocket. These observations are consistent with the results of recent X-ray studies of Hsp90 in complex with different ligands reported by Wright et al.(103).

Molecular interactions between AICAR and Hsp90: experimental tests. ELISA showed that AICAR bound the recombinant Hsp90 N-domain in a specific and saturable manner, whereas no specific binding of AICAR to the C-terminus region of Hsp90 was demonstrated (**Fig. 4.25**). To further investigate the physical interaction between AICAR and Hsp90 in vivo, affinity chromatography experiments were carried out. Fractionation of HeLa cell extracts over an AICAR–Sepharose column resulted in the isolation of Hsp90 by Western blotting of eluted fractions (**Fig. 4.25**). In contrast, no specific association of Hsp90 with a control Sepharose column was observed (**Fig. 4.25**).

5.3.4. AICAR Induces Tumor Cell Death

A 72-h exposure to different concentrations (from 31.25 to 250 μM) of AICAR induced a dose-dependent decline in cell proliferation in various human tumor cell lines tested (**Fig. 4.26**).

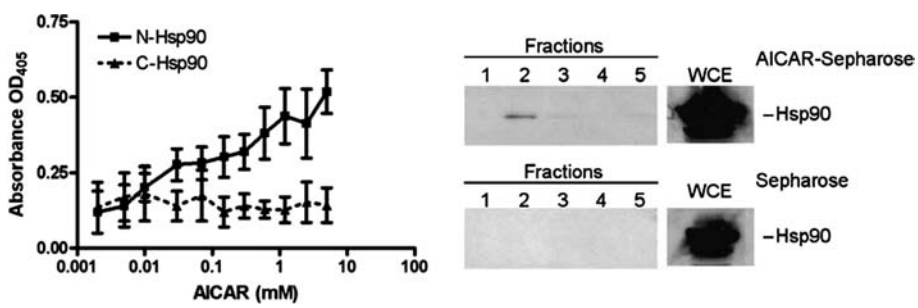


Fig. 4.25. *Right*, ELISA. Microtiter wells were coated with the indicated increasing concentrations of AICAR, and incubated with recombinant N- or C-domain of Hsp90. Binding was determined using domain-specific antibodies to Hsp90 and quantified by absorbance at OD_{405}. Data are the mean ± SEM of three independent experiments. *Left*, affinity chromatography. AICAR was coupled to sepharose and used to fractionate HeLa cell extracts. Eluted fractions were analyzed by Western blotting. A sepharose column was used as control.

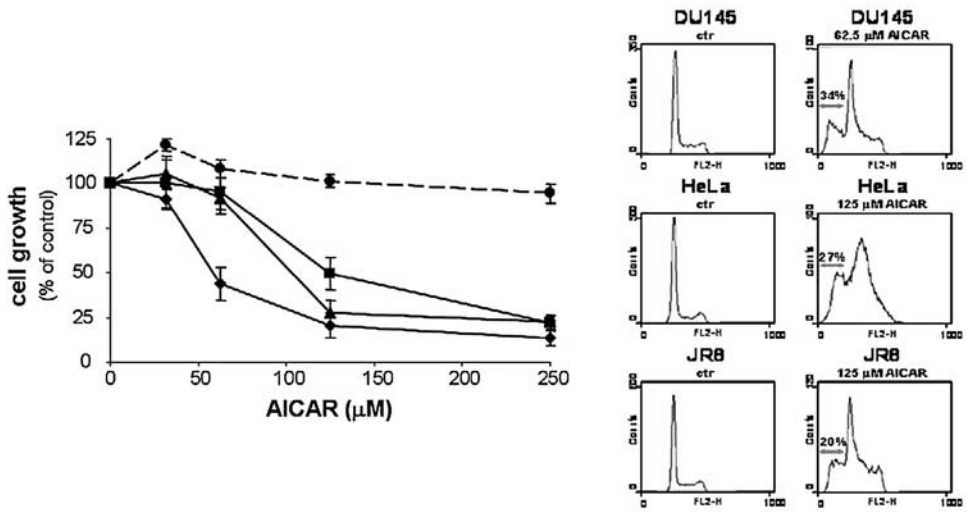


Fig. 4.26. Inhibition of cell proliferation in DU145 (◆), HeLa (■), JR8 (▲), and WI38 (•) cells as evaluated by cell counting after a 72-h exposure to AICAR. Values represent the mean (± SD) of three independent experiments. Right Panel DNA content analysis. DU145, HeLa and JR8 cells were treated with AICAR or 17-AAG for 72 h, harvested and analyzed for DNA content by flow cytometry. The percentage of the apoptotic sub-$G_{0/1}$ population is reported in each histogram.

Specifically, the 50% inhibitory concentration (IC_{50}) values, as calculated from the growth curves, were 59.0 ± 8.5 μM, 105.0 ± 7.7 μM and 126.0 ± 10.3 μM for the DU145, JR8 and HeLa cell lines, respectively, **Table 4.4**. In contrast, identical concentrations of AICAR did not reduce the growth of WI38 normal human lung fibroblasts (**Figs**. **4.26** and **4.27**). For comparative purposes, the antiproliferative effect of 17-AAG was also evaluated and IC_{50} values ranging from 53.3 ± 4.1 to 108.2 ± 19.3 nM were observed for the different cell lines, **Table 4.4**. In tumor cell lines, AICAR induced the activation of apoptosis as demonstrated by the appearance of an apoptotic sub-$G_{0/1}$ peak after a 72-h exposure to the drug (**Fig. 4.28**).

Table 4.4
Cytotoxic activity of AICAR and 17-AAG in tumor cell lines. Drug activity is expressed in terms of the concentration able to inhibit cell growth by 50% (IC_{50}). Values represent the mean ± SD of three independent experiments

Cell line	AICAR IC_{50} (μM)	17-AAG IC_{50} (nM)
DU145	59.0 ± 8.5	68.3 ± 5.8
HeLa	126.0 ± 10.3	108.2 ± 19.3
JR8	105.0 ± 7.7	53.3 ± 4.9

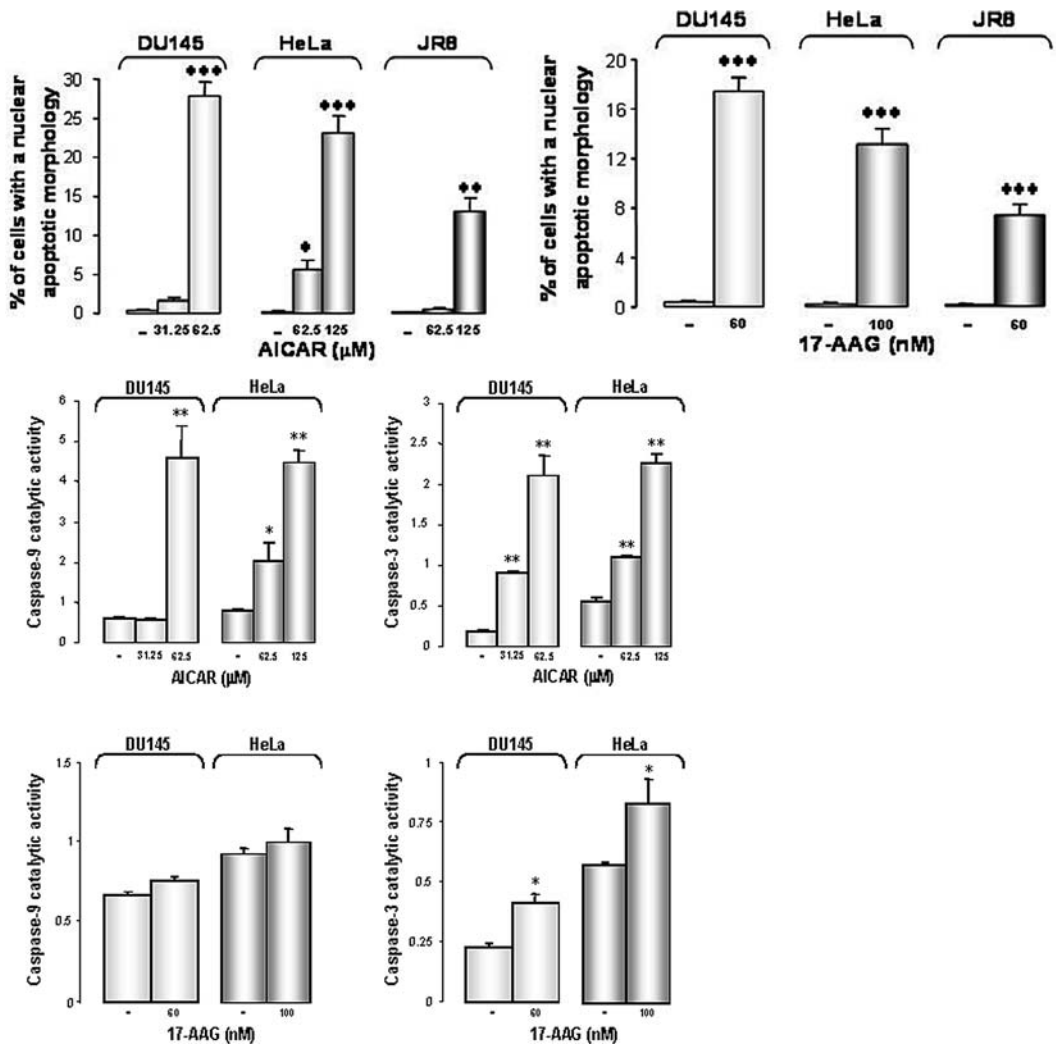


Fig. 4.27. *Top panel*, fluorescence microscopy analysis. The percentage of cells with an apoptotic morphology was assessed by fluorescence microscopy after exposure of cells to AICAR or 17-AAG for 72 h. Data represent the mean values (± SD) of three independent experiments. ***, $P<0.01$, **, $P<0.02$, *, $P<0.05$, Student's *t* test. *Bottom line*, caspase activation. Caspase-9 and caspase-3 catalytic activity were determined by hydrolysis of the fluorogenic substrate LEHD-pNA and DEVD-pNA, respectively, in cells exposed to AICAR for 72 h. Data represent the mean values (± SD) of three independent experiments. **, $P<0.01$, *, $P<0.05$, Student's *t* test.

Such a peak, which was not appreciable in untreated cells, accounted for 34.0 ± 2.3% of the overall cell population in DU145 cells treated with 62.5 μM AICAR, and for 27.0 ± 1.4% and 20.0 ± 6.1%, respectively, in HeLa and JR8 cells treated with the highest (125 μM) drug concentration. Genuine induction of tumor cell apoptosis by AICAR was also confirmed by a significant increase in the percentage of cells with an apoptotic nuclear morphology, as detected by fluorescence microscopy after cell staining with propidium iodide (**Fig. 4.27**).

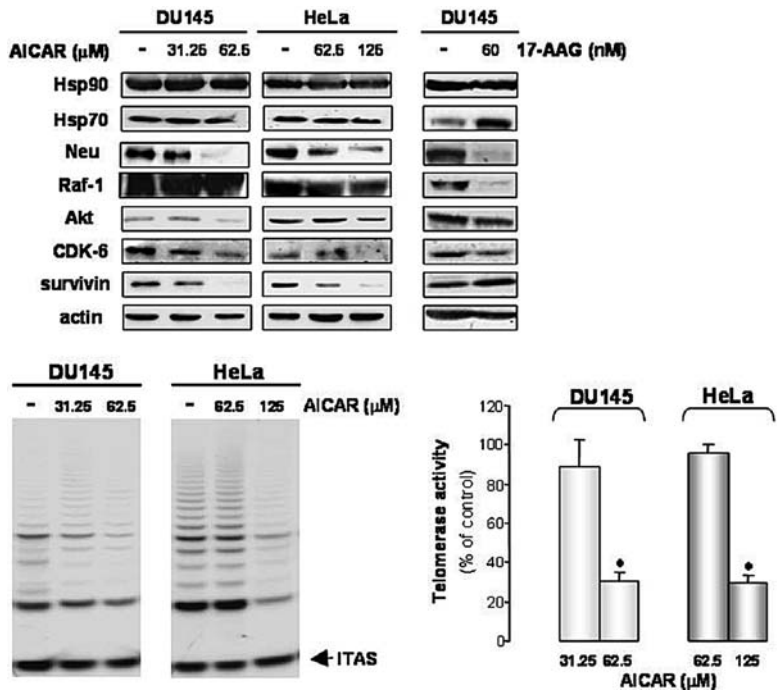


Fig. 4.28. *Top panel*, loss of Hsp90 client proteins in cancer cells treated with AICAR or 17-AAG for 72 h and analyzed by Western blotting. *Bottom line*, telomerase activity. Hsp90 was immunoprecipitated from AICAR-treated DU145 and HeLa cells, and telomerase activity was determined by TRAP assay. ITAS, internal amplification standard. Data represent the mean values (± SD) of three independent experiments. *, $P<0.01$, Student's *t* test.

In these experiments, spontaneous apoptosis was negligible (0.15–0.44% in different cell lines), whereas AICAR treatment increased the apoptotic rate in a concentration-dependent manner. Maximum peaks of 28 ± 1.7%, 23 ± 2.4% and 13 ± 1.8% were observed in DU145, HeLa and JR8, respectively, following exposure to AICAR concentrations around the IC_{50} values of each cell line. Exposure to equitoxic ($\sim IC_{50}$) 17-AAG concentrations resulted in an apoptotic rate of 15.9 ± 2.0%, 10.8 ± 1.5% and 12.0 ± 1.9% in DU145, HeLa and JR8, respectively (**Fig. 4.27**). AICAR-induced tumor cell apoptosis was accompanied by a significant increase in capase-9 and capase-3 catalytic activity, as assessed by the in vitro hydrolysis of specific fluorogenic substrates. Although to a lesser extent, significant activation of caspase-3 was also observed in DU145 and HeLa cells following 17-AAG exposure. Conversely, a negligible drug effect on caspase-9 activity was recorded (**Fig. 4.27**).

5.3.5. AICAR Inhibits Hsp90 Chaperone Function

We next asked whether the cytotoxic effect exerted by AICAR was due to the breakdown of multiple cell survival pathways as a consequence of destabilization of Hsp90 client proteins. AICAR-treated

DU145 and HeLa cells exhibited reduced expression of the Hsp90 client proteins Neu, Akt, and CDK-6 and almost complete loss of survivin by Western blotting (**Fig. 4.28**). In contrast, AICAR did not affect Hsp90 and Hsp70 expression (**Fig. 4.28**). Exposure of DU145 cells to 17-AAG resulted in almost complete abrogation of the Hsp90 client protein Raf-1 and a marked decline in Neu. Conversely, no appreciable inhibition of survivin expression was detected. 17-AAG did not modify the Hsp90 expression level but markedly increased Hsp70 protein abundance (**Fig. 4.28**). In addition, immune complexes precipitated from extracts of AICAR-treated tumor cells revealed a 70% reduction in telomerase activity, which requires Hsp90 chaperone function (104), as compared with control incubation reactions (**Fig. 4.28**).

5.4. Discussion

In this study, we used a structure- and dynamics-based rational design to identify the non-peptidic small molecule AICAR as a structurally novel inhibitor of Hsp90. The compound was selected to engage the ATP-binding pocket of the N-terminal domain of Hsp90, with binding and functional properties that mimic those of the peptidic antagonist of the survivin–Hsp90 complex, shepherdin (11). Accordingly, AICAR bound the Hsp90 N-domain, destabilized multiple Hsp90 client proteins in vivo, including survivin, and exhibited broad antiproliferative activity in multiple tumor cell lines, although at higher concentrations than those required to obtain the same growth inhibitory effect with 17-AAG, with induction of apoptosis and inhibition of cell proliferation. Reminiscent of the selective anticancer activity of shepherdin, AICAR did not affect proliferation of normal human fibroblasts.

Hsp90 is generally viewed as a "druggable" target for rational cancer therapy because of its role at the crossroads of multiple signaling pathways associated with cell proliferation and cell viability, and its elevated ATPase activity in tumors as opposed to normal tissues, which may provide a favorable therapeutic window (59, 105). This has prompted considerable interest in the development of Hsp90 antagonists for clinical use, a strategy that relied largely on the macrocyclic antibiotic geldanamycin (GA) and its 17-allyl derivative, 17-AAG, as template structures (70, 74, 105, 106). Conversely, the pharmacophore strategy presented here, using the shepherdin peptidomimetic as a lead structure, uncovered an unexpected degree of molecular and structural heterogeneity for Hsp90 inhibitors, and identified AICAR as a novel antagonist of the chaperone ATPase activity with promising anticancer activity. It is worth noting that the molecular structure of AICAR is totally uncorrelated with that of ansamycins. The former features a ribose ring linked to a heterocyclic imidazole function mimicking the aromatic portion of His86 and provides easy access to several functional groups for further chemical modification, while the latter is generally characterized by an aliphatic

macrocycle linking non-adjacent positions of an aromatic ring. Similarly to shepherdin (2), preliminary docking studies predicted AICAR to engage the ATPase pocket of Hsp90 differently from GA, and amino acids that contact GA were predicted by simulation analysis to be marginally involved, or not involved at all in shepherdin or AICAR docking to Hsp90. Strengthening the rationale of the proposed shepherdin-based screening strategy, which takes into account the motional and flexibility properties of the complex with Hsp90 via all-atom MD, previous site-directed mutagenesis experiments confirmed the essential roles of the two Ser, His, and Cys residues in shepherdin that make up the pharmacophore used here. Moreover, a minimal shepherdin *KHSSG* sequence was sufficient to retain the ability to bind Hsp90 and exhibit selective anticancer activity (8). Models generated based only on the initial semi-rigid docking results may not properly recapitulate flexible regions in Hsp90 required for inhibitor binding, and failed to identify the groups essential for recognition. In this context, the results of the AICAR/Hsp90 interaction studies presented here are indicative of the structural determinants for the activity of AICAR as an Hsp90 inhibitor. In particular, AICAR is involved in a direct contact with Asp93 through two hydroxyl groups on the furanoside, suggesting that modifications of these residues with other polar groups may actually strengthen the molecular recognition interaction. The same considerations hold for the substituents on the imidazole ring. Analysis of the structure of the complex after long time-scale MD refinement also suggests other modifications aimed at improving the hydrophobic contribution to the free-energy of association: the hydroxyl group in position 2 of the furanoside points directly toward a hydrophobic channel which can be occupied by a hydrophobic moiety.

In previous studies, AICAR has been proposed as a pharmacological activator of AMP-activated protein kinase (AMPK), with potential anti-inflammatory properties (107, 108). The recently described anticancer activity of AICAR was attributed to this molecular recognition, and in particular to inhibition of PI3 kinase-Akt signaling with increased expression of cyclin-dependent kinase inhibitors, p21 and p27, and p53 (109). Our data suggest an additional function of AICAR as a novel and structurally diverse Hsp90 antagonist, and its anticancer activity reflects inhibition of the chaperone function with destabilization of multiple Hsp90 client proteins, including Neu, CDK6, Akt, survivin, and telomerase. It is worthy of note that the molecular mechanisms by which AICAR exerts its activity are not superimposable to those of 17-AAG. In fact, we failed to evidence an inhibitory effect on survivin expression after cell exposure to a cytotoxic concentration (IC_{50}) of 17-AAG. This finding is in accord with previous data from our group showing a decline of survivin protein expression in prostate cancer cells only starting from a very high 17-AAG

concentration (>850 nM) (110). Moreover, at variance with 17-AAG, AICAR did not modify the expression of Hsp70 and Raf-1 proteins. It has to be stressed that our data do not exclude the interference of AICAR with the AMPK pathway, but suggest its possible parallel interaction in the cell with a different biological pathway exploited in cancer. Most importantly, the cell killing effect of AICAR is selective, as the drug spares normal cells. The strategy proposed here can be extended to the investigation of other targets involved in cancer development and to the identification of possible new uses for already known drugs. In summary, we have presented a novel, simple and general design strategy based on the combination of docking and all-atom MD refinement to identify the binding determinants of complex formation while keeping flexibility into account. This protocol was applied to study the interaction between a novel peptidomimetic antagonist of the survivin–Hsp90 complex, shepherdin, and Hsp90. Our approach allowed us to explore the molecular diversity space of Hsp90 inhibitors and to rationally identify a new structural antagonist, AICAR. The ability to inhibit the chaperone function, destabilize multiple client proteins in vivo, and exhibit selective anticancer activity makes AICAR a viable lead structure for the further development and refinement of novel Hsp90 antagonists structurally different from 17-AAG.

6. Peptide Self-Organization

6.1. Introduction

Molecular self-organization and self-assembly are processes by which Nature builds complex 3D multicomponent structures with well-defined functions starting from simple building blocks such as oligonucleotides, lipids or peptides. In the case of peptides and proteins, in particular, these phenomena have been the focus of intense research in recent years, after the realization that self-assembly is at the basis of the aggregation processes leading to the formation of oligomers and of insoluble amyloid fibrils, whose deposition is considered a hallmark of many types of human diseases. These involve the most diffused neurodegenerative disorders such as the Alzheimer's disease, Parkinson's, Creutzfeld–Jakobs Syndrome, and many others.

However, the ability to form ordered amyloid aggregates is not restricted to disease-related sequences. A large number of non-pathogenic polypeptides have been shown to form ordered fibrils under particular solvent, temperature, and pH conditions (1). In spite of the high sequence diversity, many proteins and peptides aggregate into a common cross-β-sheet structure and the resulting fibrils show remarkable ultrastructural and biophysical similarity.

Analysis of the deposits using electron and atomic force microscopy demonstrate the existence of fibrils with a diameter of 7–10 nm and a length of several microns. An intriguing point is the finding that very short peptides, as short as tetrapeptides, can form typical fibrils with the canonical features of amyloid fibrils. These fibrillar assemblies belong naturally to the realm of the nanoscale, and the study of their properties and mechanisms of formation may be the source of inspiration for the development of ordered, rationally designed nanostructures with potentially interesting applications in biotechnology and other fields ranging from material sciences to tissue engineering, from molecular electronics to molecular medicine and drug delivery. As a consequence, this area is extremely promising for many future chemical applications.

The richness of shape, chemical, and physical properties provided in particular by peptides (as well as other biological polymers) offers in fact a unique opportunity for the development of ordered supra-molecular nanostructures. An outstanding example of the use of molecular self-assembly and nanotechnology in biomedicine is given for instance in a paper by Ellis-Behnke et al.(111), where the authors report that a designed self-assembling peptide nanofiber scaffold creates in an animal model a permissive environment for axons not only to regenerate through the site of an acute injury but also to knit the brain tissues together.

Considering the importance that the peptide self-assembly process holds for vast communities in the scientific and technological worlds, much effort is being undertaken to unveil the fine details of the mechanism by which peptides aggregate in well-ordered structures. The high degree of complexity of aggregation itself has actually hampered the experimental atomic-level characterization of the essential factors of the process. Understanding the physico-chemical determinants underlying peptide self-assembly at the atomic level is in fact a fundamental step in view of the rational design of new nanostructures for targeted applications or of new drugs able to interfere with the harmful amyloid formation process. This requires the combined effort of computational–theoretical and experimental approaches. Herein, we will first describe several recent studies from the computational and theoretical fields that have contributed to improve our understanding of the principles governing the organization of self-assembling peptide systems; then, we will present the development and testing on an experimentally characterized model of a new protocol for the simulation of the initial phases of self-aggregation developed in our laboratory.

6.2. Computer Simulations of Peptide Self-Assembly

Theoretical and computational methods play two main, strictly interconnected, roles in the study of peptide self-organization: one is the development of a framework for understanding the mechanisms leading to the formation of ordered aggregates, and

the other is to help and drive the design of new sequences with selected properties for nanobiotechnological applications. In this context, our attention will be concentrated on (1) the study of the correlations between the conformational properties of the monomeric constituents of the aggregate and their tendencies to form stable polymeric assemblies and the formation of oligomeric species considered as the initial building blocks for the subsequent fibril growth and (2) on models of the final organization of whole fibrils which allowed to investigate the effects on stability of possible different arrangements, the effects of sequence mutations on a certain supra-molecular structure or the conformational dynamics in the aggregate.

6.3. Simulations of the Properties of Monomers and the Formation of Oligomers

The study of the dynamics and conformational properties of peptides in relation to their tendency to form aggregating species stemmed from the experience obtained from the study of protein and peptide folding over the last decade. The results of these simulations can, in many cases, be compared directly with experiments. Simulations can thus be used to shed light on the conformational transitions that trigger the self-assembly of otherwise soluble peptides. In this context, long time-scale molecular dynamics (MD) simulations have been used to monitor the dynamics of transition from the soluble ideal α-helical to aggregating β-hairpin conformations of pathogenic peptides such as the H1 peptide from the prion protein and the Aβ(12–28) fragment from the whole length Aβ Alzheimer's peptide. The results suggest a common α to β transition mechanism characterized by an initial unfolding of the α-helix, followed by the formation of bent conformations and a final convergence to ordered in register β-hairpin conformations. In spite of the different sequences, the β-hairpins exhibit the presence of a peculiar pattern of exposed hydrophobic side chains, in particular in the turn regions, suggesting a common aggregation mechanism (112). Extensive MD simulations have also been used to probe the folding properties and the effect of mutations on the aggregation propensities of different Aβ stretches (Aβ(25–35), Aβ(21–30), etc.) or polyglutamine (113–115). Calculations show that native sequences have a higher tendency than their mutants to populate conformations prone to self-assembly and to the formation of amyloids, and suggest that the structure and stability of the intermediates can be targeted in drug design. Moreover, these results provide clear sequence–structure–mechanism relationships for peptide aggregation through monomer simulations, an aspect of fundamental importance for the rational design and modification of self-assembling sequences.

The effect of the solvent environment on the aggregation propensities of model peptides has also been targeted by MD simulations (116, 117). In general, the presence of apolar solvents favors helical non-aggregating conformations, and suggest the

possibility to tune the conformational determinants of the process by simple "solvent-engineering," in a very similar way to what has been done in the past with enzymatic reactivity.

Compared to the conformational analysis of monomers, the direct simulation of peptide aggregation, self-organization and the formation of initial oligomers is still an extreme challenge for all-atom models as simulations must handle many chains and find ways to follow the dynamics during some of the critical steps. However, the hurdle can be passed successfully by neglecting the finest atomic details in favor of models that catch only the essential physical properties of the system under exam. The use of a more coarse-grained level of representation of the sequences decreases the number of particles, degrees of freedom and interactions to compute and results in a definite gain in the time required to generate physically meaningful configurations. Peptides could be represented, for instance, as chains with all backbone atoms included while side chains could be modeled by one bead with a certain van der Waals radius and geometrical parameters with respect to the backbone. Within this theoretical framework, Derreumaux and co-workers run simulations of different copies of the amyloidogenic *KFFE* sequence and showed that while the monomer is disordered, the presence of four copies of the peptide in the simulation box could induce the formation of a variety of oligomeric states with fully antiparallel or mixed parallel–antiparallel configurations. Increasing the number of chains to a minimum of six peptides, they could observe the others formation of double layer β-sheet assemblies with a 10 Å distance between two layers and a 4.5 Å distance between to peptide in one layer, consistent with experimental X-ray data. These methods were also applied to study the di- and trimerization of Aβ(16–22). The results predicted the formation of in-register antiparallel sheets, consistently with NMR data (118). The oligomerization of Aβ(16–22) into mainly antiparallel β-sheets was also observed by Favrin (119) and co-workers using a sequence-based atomic model with an effective potential based on hydrogen bonds and hydrophobic attractions and unbiased Monte Carlo simulations for sampling.

All the observations on the oligomerization of the peptides reported above point to hydrophobic and stacking interactions as the determining factors in defining the relative orientation, directionality, and the supra-molecular ordering in self-organizing peptide constructs appeared. Moreover, all the simulations showed the significant presence of a number of conformations with parallel geometries, helping explain the shift from antiparallel to parallel structures when the peptides are coupled to an octanoyl moiety. This is a clear example where the use of computer simulations can help in rationalizing experimental observations on the self-assembly behavior of peptides and their modified analogs (120, 121).

The fundamental role of stacking interactions in determining the structure and self-assembly behavior of short peptides emerged also in simulations employing more refined all-atom models. Gsponer et al., for instance, studied the evolution of a system of three copies of the *GNNQQNY* peptide using MD with a simplified implicit solvent representation. The stacking of the tyrosine rings and hydrogen bonds between amide groups favored the parallel β-sheet arrangement over the antiparallel one.

The significant role of stacking interactions was also observed in simulations of the *NFGAIL* sequence in explicit solvent, in conditions mimicking extremely high peptide concentrations to speed up the aggregation process. At two different temperatures, the formation of flat ellipsoid shaped clusters of peptides was induced by the presence of edge-to-face or face-to-face stacking interactions involving the aromatic moieties of the Phe residue. Experimental mutation of Phe to Ala completely abolished the peptide's ability to self-aggregate and produce fibrils.

6.4. Simulations of Preformed Supra-molecular Structures

While many crucial aspects of peptide self-assembly could be observed by using simplified or even all-atom approaches on small system consisting of a few copies of the monomers, the long time-scales for ordered aggregation are still out of reach for computer simulations of systems consisting of tens or hundreds of copies of monomers with explicit solvent representations. However, experimental data (X-ray, solid state NMR, FT-IR, etc.) already provide interesting insight on the final organization of the fibril which can be used to build reliable starting structures for MD simulations. As a consequence, it is possible to investigate the effects on the stability of a certain assembly of different hypothetic arrangements (122) or the effects of sequence mutations on a certain supra-molecular conformation. This approach was applied for instance by Nussinov and co-workers to several peptides of unrelated origin such as the 113–120 stretch from the Syrian Hamster Prion, the *h*IAPP$_{22–27}$ and hIAPP$_{22–29}$ and several peptides from the Alzheimer's Aβ, showing how the final organization of the fibrils depends in a specific way on the protein sequence. In this context, de la Paz et al. carried out several explicit solvent MD simulations on a series of single point mutants of the de novo designed amyloidogenic *STVIIE* peptide, starting from different initial conformations of a preformed polymeric six-stranded β-sheet. The results provided evidence for the influence of a small number of site-specific hydrophobic interactions on the packing and stabilization of the aggregates as well as on the interplay between side chain interactions and the net charge of the molecule on the strand arrangement of polymeric β-sheets. Importantly, this MD analysis also shed light into the origin of the position dependence on mutation of β-sheet polymerization in

agreement with several experimental observations. Caflisch and co-workers recently extended this approach to identify possible aggregating stretches from folded protein sequences.

An interesting application of MD simulations to the study of the collective properties of large-scale aggregates is represented by the study of Daura and co-workers on a preformed model of a fibril of the Siv gp32 fusion peptide. Starting from a non-twisted model they observed that the parallel β-sheet aggregates spontaneously adopted a helical suprastructure in the simulations. A dynamic equilibrium was observed involving partial unwinding and rewinding of the suprastructure. The supra-molecular structure observed could be described as a left-handed twisted ribbon with saddle-like curvature. The chirality of the constituent chains rendered this structure more thermodynamically stable than a helical ribbon. This was the first example in which the spontaneous twisting of a fibril into an *equilibrium* twisted-ribbon suprastructure was observed and provided a very interesting view of the dynamic nature of self-assembled aggregates.

Summarizing, simulations, and MD in particular, can already be applied to detect critical interactions for peptide self-organization, and to investigate the dynamic conformational behavior of the resulting assemblies. Understanding and being able to predict and control both these aspects is fundamental for the design of new supra-molecular entities with tailored properties for application in both nanotechnology and nanobiology.

6.5. Combining Coarse Grained and All-Atom Simulations for the Ab Initio Simulation of Peptide Aggregation

The self-organization of multiple peptide chains into ordered aggregates is an inherently very complex problem: the free-energy landscape that a system composed of different chains has to explore is enormously vast and so is the number of degrees of freedom that should be considered to completely characterize that landscape. Moreover, the study of aggregation and self-organization of peptides is aggravated by the difficulties in experimentally characterizing the final fibrils. Since amyloid fibrils are insoluble and do not form crystals, high-resolution atomic structures are still lacking for most sequences and our understanding of fibril structure and stability comes essentially from X-ray fibril diffraction, negative stain electron microscopy, atomic force spectroscopy, and, more recently, solid-state nuclear magnetic resonance (NMR) spectroscopy.

Fiber diffraction patterns of aggregates show the presence of strong reflections at 4.8 Å and at 10–11 Å resolution, indicative of the cross-β-spine structural motif typical of amyloid fibrils. The consensus emerging from these studies is that all fibrillogenic peptides aggregate into a common cross-β-sheet structure with the β-strands perpendicular to the fiber axis. A complete structural definition of these aggregates is, however, still highly debated. Over the years, several models have been proposed, which are

essentially variations on the common cross-β-spine theme. Moreover, there is mounting evidence that the toxicity of the fatal neurodegenerative diseases may be caused by the intermediate oligomers in addition to the mature fibrils (123–126). It is therefore important to characterize the early steps of oligomer formation at the atomic level. Because these structures are metastable and short-lived, experimental data are difficult to obtain (127, 128).

The use of model peptides whose solid aggregates recapitulate most of the features associated with amyloid fibers has proved to be an important strategy to try and overcome these difficulties. In particular, impressive results have been obtained in the structural characterization of the peptide *GNNQQNY* derived from the prion-determining domain of yeast protein Sup35. The solid state of this peptide shows all the physical properties associated with amyloid fibrils. Most importantly, this is the only peptide whose atomic-level high-resolution structure in the fibrillar state has been reported to date (129). The structure has revealed unique features along with expected properties. The final fibril shows a hierarchical organization based on different types of stabilizing interactions. The basic unit of the assembly is a pair of β-sheets separated by a dry interface. Each β-sheet is formed by parallel strands, which are orthogonal to the fiber axis (*see* **Fig. 4.29**).

In addition to the classical hydrogen bonds between backbone atoms, the β-sheets are stabilized by hydrogen bonds between polar side chains that follow the pattern denoted as polar zipper. At the dry interface, the side chains of residues N2, Q4, and N6 are tightly interdigitated with the corresponding residues of the opposing sheet (steric zipper) (**Fig. 4.29**). This complementarity makes this motif unique.

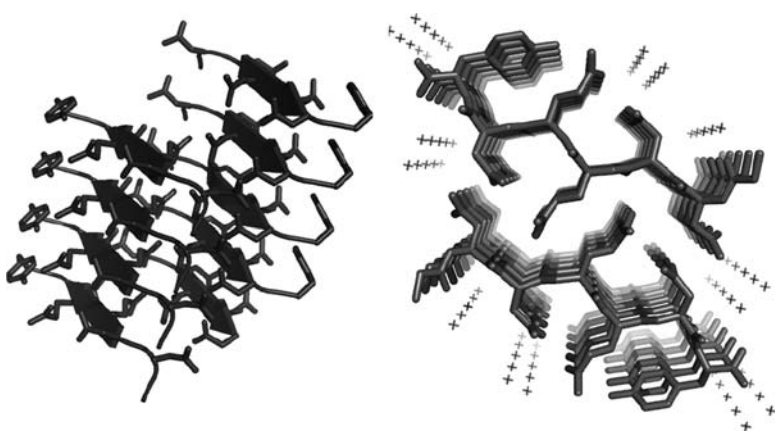


Fig. 4.29. On *left*, the pair-of-sheets structure, showing the backbone of each beta-strand as an arrow, with side chains protruding. On *right*, the steric zipper. This is a close-up view of a pair of GNNQQNY molecules, from (1).

All of these properties make this small system an optimal model for testing aggregation studies and molecular simulations in particular.

The questions we ask here, are:

(1) Is it possible to reproduce the hierarchical organization observed in the fibril with molecular simulations starting from unbiased initial configurations ?

(2) Is it possible to detect in molecular simulations the stabilizing patterns observed experimentally?

(3) Is there a correlation between the sets of structures determined computationally and the experimental one?

Predicting the aggregation paths and the detailed atomic structures of the intermediates and of the initial stable structures is a difficult task for computer simulations. The typical time-scales for the formation of fibrils range from hours to several days even to weeks and thus are several orders of magnitude in time beyond what can be obtained by current all-atom simulations. It is also difficult to determine the minimal model that reproduces the essential structural and energetic features of peptides and allows an extensive search of conformational space. A variety of chain representation and energy models and methods including molecular dynamics (MD), discontinuous MD, replica exchange MD simulations, and Monte Carlo-based methods (MC) have been applied to amyloid-forming peptides, with variable degrees of success.

Starting from these considerations, in the present study, we aim at improving the reach of aggregation simulations by combining the extensive sampling properties of coarse-grained methods to the refinement abilities of all-atom MD simulations. To achieve this goal, we first run a preliminary, simplified, Monte Carlo exploration of the free-energy landscape. If the free-energy landscape associated to the simplified MC dynamics retained all the relevant features of the true one, we could expect that the all-atom MD simulations started from the sampled configurations are expected to have significant advantages over, for example, those starting from fully random placements of peptide positions which can be affected by significant lag-phases (130). The advantage could, however, be more conspicuous if the MC procedure allowed picking structures from the transition state ensemble of the aggregation process, in which case the all-atom dynamics would progress toward models of the final aggregates in a time-scale much shorter than the typical fibril formation time. Thus, the expectation is that the two-stage MC–MD trajectories are "time-advanced" with respect to those started, for example, from configurations consisting of random placements of the single chains or from subjectively chosen preformed oligomers. Simulations were run on several systems consisting of different numbers

of copies of the *GNNQQNY* peptide, starting from random initial configurations. Each set of simulations contained 2–8 copies of the peptide. For each system three different simulations are run, each starting from the representative structure of the three most populated structural clusters identified at the transition temperature of the coarse-grained MC analysis (*vide infra*). Simulations are labeled according to the number of chains in the simulation box (2,3,4,. . .,7) followed by a "_c" (Chains), the cluster number (1,2, or 3) and the temperature at which the simulation is run. So, for instance, the simulation of four peptide chains, starting from the representative of the second-most populated cluster at 300 K will be labeled 2c_2_300.

6.6. Results

The preliminary simplification of the free-energy landscape operated by the MC approach is achieved through a coarse-graining of the structural degrees of freedom of each peptide chain composing the system. In particular, the protein is described in terms of its C_α trace and of the associated C_β centroids. This schematic description is accompanied by a simplification of the energy functional (see the Monte Carlo section of the Methods chapter) that incorporates effective pairwise aminoacid–aminoacid interactions, and information to avoid steric clashes. The use of this type of simplified interaction potential has already proven to be effective in the description of the folding dynamics of small proteins with non-trivial tertiary structure features (11, 131, 132). Since the folding process is an *intramolecular* recognition process governed by non-local interactions between aminoacids which are of the same physical nature as the interactions governing the *intermolecular* aggregation processes, it is natural to extend the use of the effective potential and simplified functional to the case of peptide self-organization. In this framework, the thermodynamics of the system consisting of multiple copies of the aggregating peptide (from two to eight copies) can be characterized by a simulated annealing MC evolution involving rotational and translational moves of the peptides on a discrete grid with a 3 Å distance between grid points. At this simulation stage, the peptides are considered as rigid chains in extended conformation.

The average energy and radius of gyration (R_g) as a function of temperature of the coarse-grained system are visible in **Fig. 4.30**.

As the temperature is decreased, each system undergoes a collapse represented by a sudden (cooperative-like) decrease in both the system's energy and radius of gyration. In correspondence of the collapse temperature, T_c, the specific heat exhibits a peak which reflects the large energy fluctuations typically associated to a phase transition: the multiple chain system is in fact undergoing a transition from a totally disordered condition to an ensemble of aggregate structures with a certain degree of

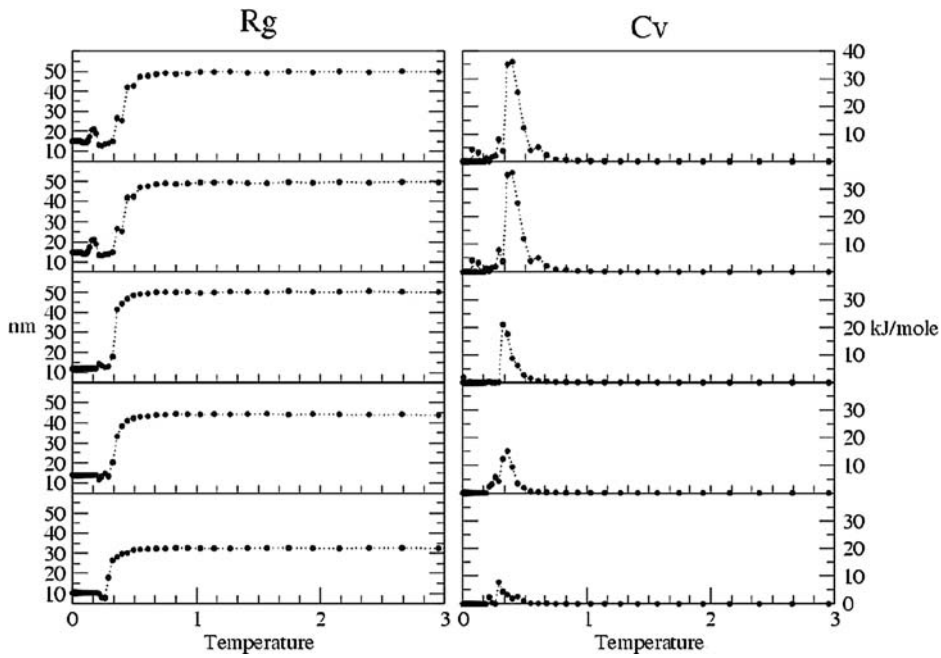


Fig. 4.30. On the *right* the radius of gyration, on the *left* for the same system the behavior of the specific heat with MC temperature. From *bottom* to *top* the graphics refer to 2,3,4 chain MC simulation and the last two refer to the seeding MC simulation with 3,4 free chain, respectively.

compactness. The latter ensemble is characterized by the coexistence of polymeric structures with one to two residues per chain involved in interchain contacts, with other structures in which the number of intermolecular interactions is higher and the chains lie in the same plane with either an antiparallel- or parallel-like orientations. The polymeric ensembles containing different numbers of peptide copies at the T_c are thus poised to collapse into compact structures with non-trivial 3D organizations. This aspect resembles closely what was already observed in the case of peptide folding, underlining once more the fact that peptide folding and aggregation are intimately connected. As a consequence, the aggregated structures sampled at the T_c with the MC algorithm represent attractive candidates for investigating the dynamics of formation of oligomeric species and their structural properties. Moreover, the conformational variability of the aggregates at T_c is such that significantly different structures can be picked.

The structures at T_c were grouped in families of similar conformations using the conformational clustering algorithm developed by Daura et al. The representative structures of the three most populated clusters from each of the MC runs were then selected as starting structures for the all-atom MD runs. These structures had to be first reconstructed in atomic detail

according to the procedure described in the Section 3.4. In the following we will describe the different simulations involving different numbers of peptides.

Two chains. The cluster analysis of the all-atom MD trajectories of the simulations with two chains in the simulation box immediately shows interesting features. In the 50 ns all-atom simulation starting from the representative of the first cluster at 300 K, 2c_1_300 K, the two peptides form a stable antiparallel structure, with the amidic side chains interdigitated at the interface between the two peptides with remarkable shape complementarity. In this particular arrangement, the interface between the two peptides is not occupied by water molecules. This structure closely resembles the one of the dry interface, with two antiparallel β-sheets in the same plane held together by van der Waals contacts between sterically complementary side chains that exclude water molecules from the environment (**Fig. 4.31**).

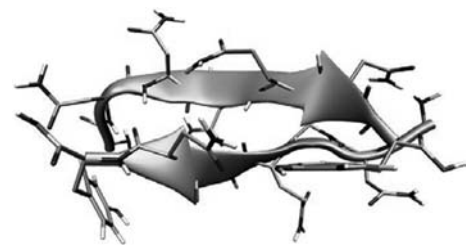


Fig. 4.31. On *left* the most sampled structure in MD simulation. On the *right* the experimental structure from (1), reproduced.

In the trajectory started from the representative structure of the second cluster determined at the T_c, 2c_2_300 K, the formation of a stable parallel β-sheet is evident immediately after the first few nanoseconds of simulation. This structural motif is stabilized by in register backbone–backbone H-bonding interactions. Additional stabilization and ordering is provided by side chain–side chain interactions between residues facing each other in the parallel arrangement: these involve not only the formation of stable hydrogen bonds between N and Q amidic groups but also packing interactions between the aromatic rings of the terminal tyrosines (**Fig. 4.32**).

The initial structure of the trajectory starting from the representative structure of the third cluster, 2c_3_300 K, is characterized by the two terminal Y residues forming a packing interaction, with the remaining part of each peptide chain pointing into the solvent without additional interchain contacts. Interestingly, this type of Y–Y packing is also observed in the crystal structure of the fibril, with the role of connecting different interdigitated dimeric

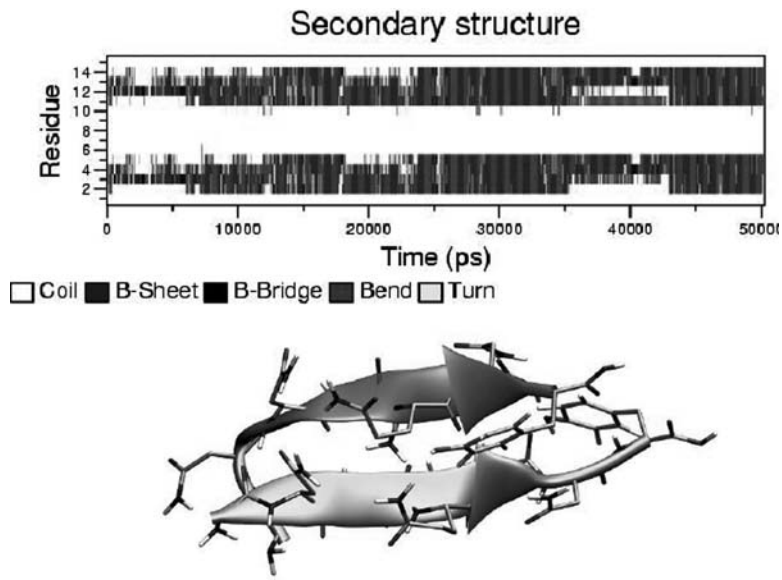


Fig. 4.32. On the *top* of the panel the secondary structure behavior with time for the simulation 2c_2_300 K. *Bottom line*, on the *right* the most sampled structure in MD simulation and on the *left* the experimental structure, from (1).

complexes. In the all-atom MD simulations, however, this starting structure is not stable and evolves toward the antiparallel arrangement, already observed for simulation 2c_1_300 K.

Control simulations were run at the temperature of 360 K to check the stability of the structures obtained in the above reported all-atom simulations at 300 K. The antiparallel arrangement observed in simulations 2c_3_300 K and 2c_1_300 K is disrupted at around 5 ns. The two chains explore an ensemble of bent conformations and, at around 21 ns, the two chains relax into a parallel arrangement that is disrupted and recovered during the rest of the simulation.

In contrast, running a simulation at 360 K starting from the final stable parallel conformation of simulation 2c_2_300 K does not cause any significant changes in terms of structure or relative organization of the *strands*, suggesting that the parallel arrangement is particularly stable for this sequence, in **Fig. 4.33**.

Three chains. Increasing the number of chains in the simulation box to three determines a different and more varied conformational scenario. In the simulation started from the first cluster obtained at the T_c of the MC run, labeled 3c_1_300 K, one chain is juxtaposed to a preformed dimer similar to that observed in simulation 2c_1_300 K. The preformed dimer is held together by interactions between the side chains forming the dry interface and the third chain binds parallel to one of the previous two.

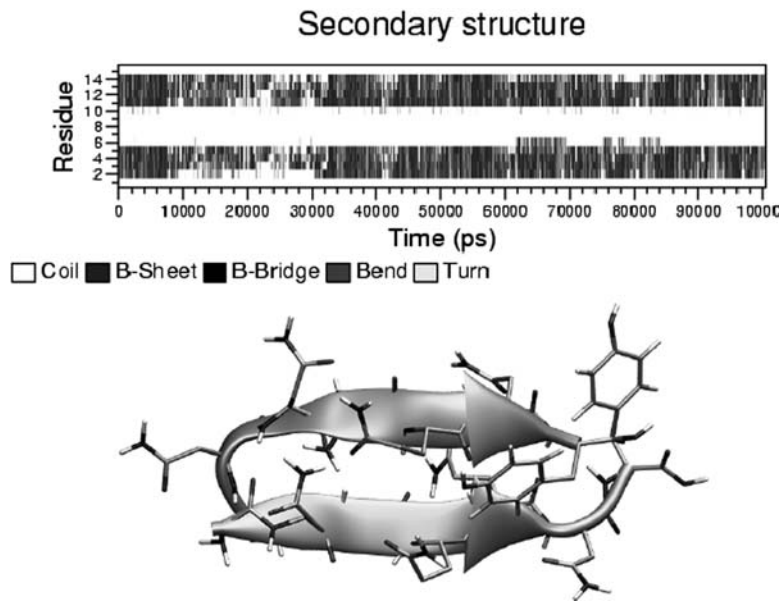


Fig. 4.33. *Top panel*, the behavior of the secondary structure with time for the simulation 2c_2_360 K. *Bottom line*, the most sampled structure during MD simulation.

Simulation 3c_2_300 K starts from a dimer where the dry-interface is formed, and the remaining chain is far in space. The latter, however, is eventually able to dock on the preformed complex forming a parallel structure with one of the two prestructured strands.

Finally, simulation 3c_3_300 K starts with a preformed dimer with the two tyrosines packing with their aromatic rings, the remaining part of the chains pointing toward the same direction in space and the third chain separated by about 50 Å from the other two. Interestingly, the evolution in the all-atom MD simulation at 300 K leads to the formation of a parallel trimer with a significant degree of packing of the three Y residues and backbone–backbone interchain hydrogen bonds stabilizing the parallel conformation (**Fig. 4.34**). Heating this structure to high temperatures (360 K) does not change the situation, once again suggesting an enhanced stability of the parallel conformations for oligomeric complexes.

Four chains. The simulations involving four different peptide chains are run to investigate the dynamics of formation of the fibril nucleus, as identified by Eisenberg and co-workers. Similarly to the previous cases, three possible starting configurations are identified from the cluster analysis of the T_c structural ensemble in the coarse-grained MC trajectory. The representative structure of the first cluster shows a "micelle-like" arrangement for the four aggregated chains: the four Y residues point toward the center of the

Fig. 4.34. On *left*, the most sampled structure for the simulation 3c_1_300 K, on *right* for the simulation 3c_2_300 K.

micelle forming a hydrophobiccore with their aromatic rings, while the rest of the peptide points outward into the solvent. During the 50 ns all-atom MD simulation at 300 K, 4c_1_300 K, this construct reorganizes to form a stable parallel β-sheet involving two central chains. The remaining two chains remain basically disordered.

The simulation involving the representative of the second cluster, 4c_2_300 K, shows the formation of a stable, parallel β-sheet involving four strands (**Fig. 4.35**). It is worth underlying that this structure is totally uncorrelated with the starting one. The latter is characterized by the presence of two dimeric complexes, contacting through the aromatic side chains of the tyrosines. Each of the dimeric complexes is initially devoid of ordered secondary structure at the beginning of the simulation.

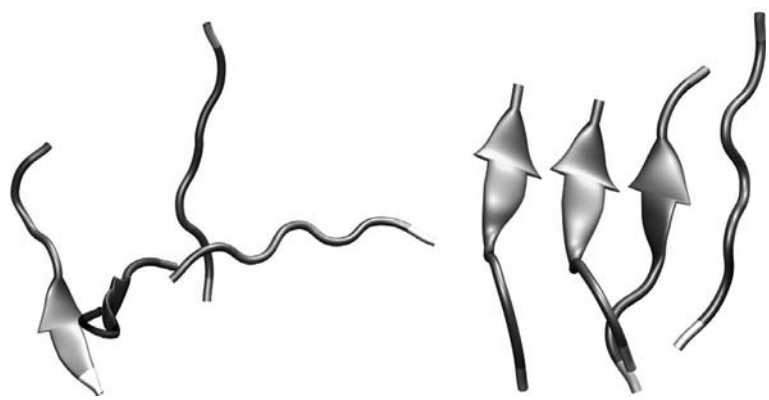


Fig. 4.35. On *left*, the initial conformation for the simulation 4c_2_300 K; on *right*, the structure after 50 ns of MD simulation.

Very interestingly, however, we could observe the rapid alignment of the two chains in each dimer to form a parallel, in-register β-sheet. During the simulation, the two β-sheets evolve spontaneously to form a flat parallel β-sheet involving all the four

chains present in the simulation box. This structure is remarkably stable at 300 K and at higher temperatures (330 K, 360 K). The rapid formation and the stability of this supra-molecular complex is in substantial agreement with the suggestion by Eisenberg and coworkers that the four chains may represent a stable nucleus for fibril formation.

The results obtained in the simulation starting from the third structure is similar to that described for 4c_2_300 K, and will not be commented further.

Seeded aggregation. The stability of the four stranded, parallel β-sheet combined with the fact that it forms rapidly even on the MC–MD time-scales suggested additional simulations in which this complex could be used as a preformed seed. Two MC simulations were thus run on two systems consisting of the following:

1. The preformed four stranded, parallel β-sheet plus three free chains.
2. The preformed four stranded, parallel β-sheet plus four free chains.

In both cases we only reconstructed and simulated at the all-atom level the representative structure of the most populated cluster. This was done to save on computational time, considering the dimensions of the boxes in the presence of such a high number of free and unconstrained chains (133).

In the all-atom MD simulation starting from the first ensemble of structures, labeled 4_plus_3, the preformed four stranded β-sheet remains stable during the whole simulation time. What is most interesting to observe is that all the three remaining peptides dock on the preformed plane in a antiparallel orientation with respect to the seed, and parallel to one another. This arrangement is very similar to the one observed in the crystal structure in the interface between the two β-sheet layers running parallel to the fibril axis (**Figs**. **4.36** and **4.29**).

It is worth underlying at this point, that in the preliminary coarse-grained MC run the additional three peptides are completely free to translate and rotate in space, the only constraint being on the structure of the seed. In the subsequent all-atom MD, *none* of the peptides is subjected to any constraint. As a consequence the whole system is free to move and reorganize.

The same considerations hold for the second simulation started from a "seeded" system, containing the four stranded β-sheet plus the four free peptides. The results of the all-atom, unconstrained MD run show that one of the free peptides adds up to the seed structure in a parallel fashion. Two of the three remaining peptides pack on the same side of the plane of this complex, with an antiparallel orientation with respect to the five stranded β-sheet.

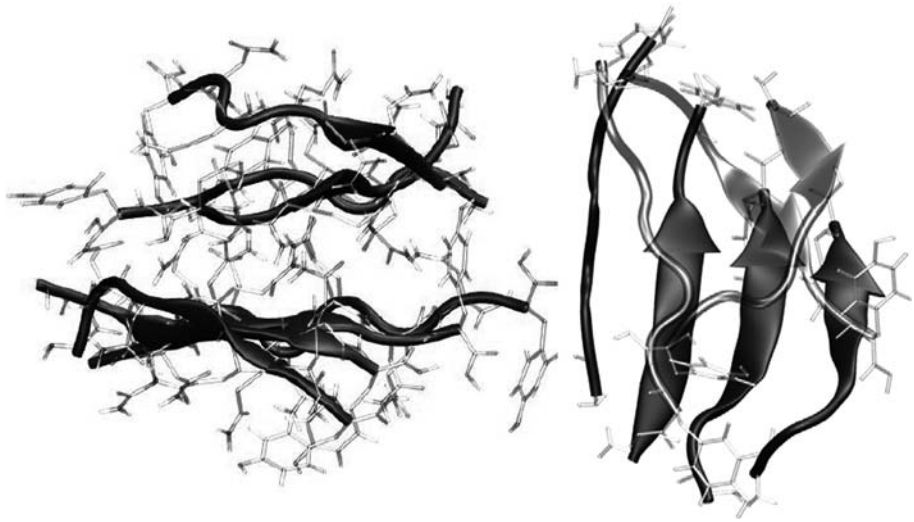


Fig. 4.36. Two images that show the results for the simulation 4_plus_3. On the *left* is show the interdigitation between the side chains of the two pairs of sheet. In *gray* are represented the two sheet. On the *right* the beta sheet and the C- and N-terminals are reported as sticks.

6.7. Discussion

In this paragraph, we have presented the results of multiple simulations on the first stages of the ordered self-organization of the peptide *GNNQQNY*. The availability of high-resolution structures of the final fibrillar form makes this system an optimal model to test aggregation simulations. Our present study aims at improving the reach of ordinary all-atom MD simulations by recoursing to a preliminary, simplified, Monte Carlo exploration of the free-energy landscape. If the free-energy landscape associated to the simplified MC dynamics retained all the relevant features of the true one, we could expect that the all-atom MD simulations started from the sampled conformations are expected to have significant advantages over, for example, those starting from fully extended peptide configurations in random orientations, which can be affected by insurmountable (on the MD time-scale) lag-phases.

The actual quantification of the expected time-advancement would be transparent and straightforward if the aggregation process could be characterized in terms of one (or more) reaction coordinates. Unfortunately, all the typical and intuitive order-parameters (such as R_g, RMSD, buried hydrophobic surface, etc.) appear inadequate to this scope due to their wide degree of fluctuation and diversity connected with the high number of possible conformations that a forming oligomer can achieve. Therefore, the benefits of the present strategy can be ascertained only through the comparison of the degree of order, β-sheet content, 3D organization, and side chain interaction patterns of the explored trajectories comparing them to the experimentally

determined structure of the *GNNQQNY* fibril. The final results of our simulations are consistent with the experimental structures and with the mechanistic considerations on fibril formation based on these structures.

The analysis of our mixed *MC–MD* trajectories suggests a hierarchical picture of the assembly process, similar to what Eisenberg and coworkers suggested based on structural analysis of the fibril. Parallel β-sheets form easily in many instances and are in general not disrupted by the high-temperature conditions. Moreover, they are mostly characterized by an in-register organization of interchain hydrogen bonds. Preformed parallel β-sheets can act as templates in the aggregation process to accelerate fiber formation.

Once the first seed is formed, more strands can pack on it mainly through van der Waals and H-bonding interactions involving the N and Q side chains, giving rise to the second level of organization. Optimal packing is achieved when the peptides forming the second plane are antiparallel to the first one. This mechanism leads to the formation of the dry interface, with exclusion of water molecules from the space between the two layers. We observe that the pattern of interactions in the second stage is less specific than the one determined by the formation of in-register hydrogen bonds observed in the first level, as can be expected considering the flexibility of the amidic side chains of the interdigitating residues. The non-specificity of van der Waals interactions in the dry interface leads to polymorphism in the structures of the starting aggregates, which can also be reflected in the amyloid fibril polymorphism observed for several sequences and strictly dependent on the solution conditions (134, 135).

The third level of organization would involve pair-of-sheets structures interacting to form the fibril, but at this level, our simulations can not reach such a degree of complexity. However, the docking of multiple preformed double layer structures with a dry interface would represent a natural consequence of the first two levels of events.

It is also interesting to observe that the ensemble of structures observed is extremely dependent on the number of molecules simulated, especially in the cases of two, three or four peptide chains in the simulation box. In the case of three peptides, in particular, we could observe a high diversity of aggregate conformations consistent with the fact that ordering three peptide molecules can determine a substantial kinetic barrier to fibril formation. In the case of four peptides in solution, we observe the easy formation of two stable parallel β-sheets, which can subsequently dock as two rigid preordered bodies to form the four stranded nucleus. Qualitatively speaking, the entropic costs of this mechanism are much lower than those of the previous one.

The results of our simulations suggest that the overall aggregation process starts with the initial formation of an unordered compact structure, which could be the rate limiting step of the process. This part of the mechanism is caught by the preliminary coarse-grained MC investigation. Hydrophobic interactions, reproduced by the residue specific statistical potential, could play a key role at the beginning of the process by disfavoring organized aggregates formation. Reorganization of this initial aggregate with the formation of more specific H-bonding and cross-β-spine interactions can eventually drive the process to reach the ordered state typical of amyloid fibrils.

To summarize, we have followed the dynamical evolution of the aggregation process of the *GNNQQNY* peptide starting from several configurations obtained by a coarse exploration of the system energy landscape. The simulated trajectories include notable instances where ordered β-sheet rich aggregated structures of the peptides are reached and maintained. Furthermore, the significant secondary content and organization found in all simulations resulted in interesting dynamical evolutions that convey valuable information on the aggregation process, shedding light on different possible organization mechanisms. The protocol used here has been kept as general and unbiased as possible and may be further extended, for example, at the level of the coarse-grained model and/or of the selection criteria of the starting structures. Therefore, the proposed strategy ought to be applicable to other instances of short proteins with significant advantages over the use of only MD simulations. To our knowledge, our results represent the first example of all-atom, explicit solvent simulation of the self-organization of multiple peptide chains without the inclusion of whatsoever information/constraint at the beginning of the calculations.

7. Conclusions

The two biological processes studied during this chapter are quite different but connected by the underlying theme of protein folding and molecular recognition. All biological molecules, such as enzymes and proteins need the correct fold for their function. *Anomalies in 3D shapes may cause diseases such as Alzheimer's, Parkinson's while too many correct folded molecules may cause diseases as cancer cell.*

In the first part of this chapter it has been possible to find by molecular dynamics the most important solution conformation of a peptide named shepherdin, with demonstrated anticancer

activity. As shown from MD simulation and from experimental data this conformation is the most active one for the molecular recognition by Hsp90 chaperonin binding site. The statistical analysis of interaction between Hsp90 and shepherdin based on all-atom MD simulation of the complex, allows us to derive a pharmacophore, which defines the minimum functional and geometrical requirements that a molecule must have to be a good ligand. These binding and geometrical constraints are found in a molecule known as AICAR. As demonstrated by experimental data this molecule preserves all features of shepherdin as inhibitor of Hsp90 chaperoning function.

The first biological processes studied has as topic an enzyme, Hsp90, a supervisor of folding of several proteins involved in survival and proliferation of the cell, such as talomerase and IAP's protein family. An up-regulation of this enzyme, a hallmark of tumor cells (3), causes an abnormal level of such proteins blocking the apoptosis signal, so that the cancel cell can spread without any control (4).

The second biological process have the opposite characteristic, mis-folded proteins can aggregate in very stable and ordered structures that cause diseases. The mis-folded proteins usually aggregate with a common structural motifs called 'cross-beta' spine.

The studies reported in this chapter are focused on the understanding of the driving forces of molecular recognition that make possible these phenomena in this two biological processes. The molecular modeling techniques used as showed are suitable methods for this research.

In the second part of this chapter the first result of a promising method to study the initial stage of the formation of the fibril structures is shown. This method stems from a theoretical framework developed for folding studies (12). As for protein folding, for aggregation of the mis-folded proteins, MD simulation alone can not be used to follow these biological process. A simplification of the energy landscape of this phenomena by simplification of the representation model of the peptide and the interaction energy function can cross the MD-lag time. Without any experimental data used as constrains for MC and for MD simulation we can reproduce with good agreement the experimental data obtained by Eisenberg (129) for the system *GNNQQNY*.

References

1. Eisenberg, D.; Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Sambashivan, S.; Ivanova, M. I.; Madsen, A. Ø.; Riekel, C. (2006) The Structural Biology of Protein Aggregation Diseases: Fundamental Questions and Some Answers. *Acc. Chem. Res* **39**, 568–575.
2. Plescia, J.; Salz, W.; Xia, F.; Pennati, M.; Zaffaroni, N.; Daidone, M. G.; Meli, M.; Dohi, T.; Fortugno, P.; Nefedova, Y.; Gabrilovich, D. I.; Colombo, G.; Altieri, D. C. (2005) Rational design of Shepherdin, a novel anticancer agent. *Cancer Cell* **7**, 457–467.

3. Fortugno, P.; Beltrami, E.; Plescia, J.; Fontana, J.; Pradhan, D.; Marchisio, P. C.; Sessa, W. C.; Altieri, D. C. (2003) Regulation of survivin function by Hsp90. *Proc. Natl. Acad. Sci. USA* **100**, 13791–13796.
4. Zaffaroni, N.; Daidone, M. G. (2002) Survivin expression and resistance to anticancer treatments: perspectives for new therapeutic interventions. *Drug Resist. Updat.* **5**, 65–72
5. Verdecia, M. A., Huang, H., Dutil, E., Kaiser, D. A., Hunter, T., and Noel, J. P. (2000) Structure of the human anti-apoptotic protein survivin reveals a dimeric arrangement. *Nat. Struct. Biol.* **7**, 602–608.
6. Lindahl, E.; Hess, B.; van-der-Spoel, D. (2001) Gromacs 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Mod.* **7**, 306–317.
7. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. (1998) Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function. *J. Comp. Chem.* **19**, 1639–1662.
8. Gyurkocza, B.; Plescia, J.; Raskett, G. M.; Garlick, D. S.; Lowry, P. A.; Carter, B. Z.; Andreeff, M.; Meli, M.; Colombo, G.; Altieri, D. C. (2006) Antileukemic activity of Shepherdin and molecular diversity of hsp90 inhibitors. *J. Natl. Cancer Inst.* **98**, 1068–1077.
9. Meli, M.; Pennati, M.; Curto, M.; Daidone, M. G.; Plescia, J.; Toba, S.; Altieri, D. C.; N., Z.; Colombo, G. (2006) Small-Molecule targeting of HSP90 chaperone function: rational identification of a new anti-cancer lead. *J. Med. Chem.* **49**, 7721–30.
10. *Accounts of Chemical Research* (2006) 39, Special issue on aggregation.
11. de-la-Paz, M. L.; de-Mori, G. M. S.; Serrano, L.; Colombo, G. (2005) Sequence dependence of amyloid fibril formation: Insights from molecular dynamics simulations. *J. Mol. Biol.* **349**, 583–596.
12. De Mori, G. M. S.; Micheletti, C.; Colombo, G. (2004) All-atom folding simulations of the villin headpiece from stochastically selected coarse-grained structures. *J. Phys. Chem. B* **108**, 12267–12270.
13. Meli, M.; Morra, G.; Colombo, G. (2008) Investigating the mechanism of peptide aggregation: insights from mixed monte carlo-molecular dynamics simulations. *Biophys. J.* **94**, 4414–26.
14. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; Gunsteren, W. F. v.; Mark, A. E. (1999) Peptide folding: when simulation meets experiment. *Angew. Chemie Intl. Ed.* **38**, 236–240.
15. Kabsch, W.; Sander, C. (1983) Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **22**, 2576–2637.
16. van Gunsteren, W. F.; Daura, X.; Mark, A. E. (1998) GROMOS Force Field. *Encyclopedia Comput. Chem.* **2**, 1211–1216.
17. van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hunenberger, P. H.; Kruger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. (1996) Biomolecular Simulation: The GROMOS96 Manual and User Guide. In vdf Hochschulverlag, ETH Zurich: Switzerland..
18. Darden, T.; York, D.; Pedersen, L. (1993) Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092.
19. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. (1995) A smooth particle mesh Ewald potential. *J. Chem. Phys.***103**, 8577–8592.
20. Unger, R.; Moult, J. (1993) Genetic algorithm for protein folding simulation. *J. Mol. Biol.* **231**, 75–81.
21. Rensburg, E. J. v.; Orlandini, E.; Whittington, S. (1996) Monte-Carlo study of the interacting self-avoiding walk in three dimension. *J. Stat. Phys.***82**, 155–181.
22. Lee, J.; Scheraga, H.; Rackovsky, S. (1997) New optimization method for conformational energy calculations on polypeptides: conformational space annealing. *J. Comput. Chem.* **18**, 1222–1232.
23. Irbäck, A.; Potthast, F. (1995) Studies of an off-lattice model for protein folding: sequence dependence and improved sampling at finite temperature. *J. Chem. Phys.***103**, 10298–10305.
24. Hansmann, U.; Okamoto, Y. (1999) New Monte Carlo algorithms for protein folding. *Curr. Opin. Struct. Biol.* **9**, 177–183.
25. Scheraga, H.; Hao, M. (1999) Entropy sampling Monte Carlo for polypeptides and proteins. *Adv. Chem. Phys.* **105**, 243–272.
26. Arkin, H.; Yarkin, F.; Celik, T.; Berg, B.; Meirovitch, H. (2002) Multicanonical simulations of some peptides. *Comp. Phys. Comm.* **147**, 600–603.
27. Garcia, A.; Onuchic, J. (2003) Folding a protein in a computer: an atomic description of the folding/unfolding of protein. *Proc Natl Acad Sci USA* **100**, 13898–13903.

28. Kolinski, A.; Godzik, A.; Skolnick, J. (1998) Monte Carlo studies of the thermodynamics and kinetics of reduced protein models: Application to small helical, b, and a, b proteins. *J. Chem. Phys.* **108**, 2608–2617.
29. Micheletti, C.; Banavar, J. R.; Maritan, A.; Seno, F. (1999) Protein structures and optimal folding from a geometrical variational principle. *Phys. Rev. Lett.* **82**, 3372–3375.
30. Baker, D. (2000) A surprising simplicity to protein folding. *Nature* **405**, 39–42.
31. Clementi, C.; Nymeyer, H.; Onuchic, J. N. (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and en-route intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **298**, 937–953.
32. Klimov, D. K.; Thirumalai, D. (2000) Native topology determines force-induced unfolding pathways in globular proteins. *J. Chem. Phys.* **113**, 8319–8328.
33. Cieplak, M.; Hoang, T. X. (2000) Scaling of folding properties in go models of proteins. *J. Biol. Phys.* **26**, 273–294.
34. Hao, H. M.; Scheraga, H. A. (1997) On foldable protein-like models: a statistical-mechanical study with Monte Carlo simulations. *Physica A* **244**, 124–145.
35. Tiana, G.; Broglia, R. A. (2001) Statistical Analysis of Native Contact Formation in the Folding of Designed Model Proteins. *J. Chem. Phys.* **114**, 2503.
36. Lazaridis, T.; Karplus, M. (1997) New view of protein folding reconciled with the old through multiple unfolding trajectories. *Science* **278**, 1928–1931.
37. Bahar, I.; Erman, B.; Jernigan, R. L.; Atilgan, A. R.; Covell, D. G. (1999) Collective Motions in HIV-1 Reverse Transcriptase: Examination of Flexibility and Enzyme Function. *J. Mol. Biol.* **285**, 1023–1037.
38. Hoang, T. X.; Cieplak, M. (2000) Molecular dynamics of folding of secondary structures in go-type models of proteins. *J. Chem. Phys.* **112**, 6851–6862.
39. Favrin, G.; Irback, A.; Wallin, S. (2002) Folding of a small helical protein using hydrogen bonds and hydrophobicity forces. *Proteins* **47**, 99–105.
40. Vieth, M.; Kolinski, A.; Brooks, C.; Skolnick, J. (1994) Prediction of the folding pathways and structure of the gcn4 leucine zipper. *J. Mol. Biol.* **237**, 361–367.
41. Liwo, A.; Pincus, M.; Wawak, R.; Rackowsky, S.; Scheraga, H. (1993) Prediction of protein conformation on the basis of a search for compact structures; test on avian pancreatic polypeptide. *Protein Sci.* **2**, 1715–1731.
42. Park, B.; Levitt, M. (1996) Energy functions that discriminate x-ray and near-native folds from well-constructed decoys. *Proteins* **258**, 367–392.
43. Kolinski, A.; Galazka, W.; Skolnick, J. (1993) A general method for the prediction of the three dimensional structure and folding pathway of globular proteins: Application to designed helical proteins. *J.* Chem. Phys. **98**(9), 7420–7433.
44. Kolinsky, A.; Skolnick, J. (1994) Monte Carlo simulations of protein folding. I. Lattice model and interaction scheme. *Proteins* **18**, 338–352.
45. Correa, P. (1990) The building of protein structures from a-carbon coordinates. *Proteins* **7**, 366–377.
46. Rey, A.; Skolnick, J. (1991) Efficient algorithm for the reconstruction of a protein backbone from the a-carbon coordinates. *J. Comput. Chem.* **13**, 443–456.
47. Holm, L.; Sander, C. (1991) Database algorithm for generating protein backbone and side-chain coordinates from a c trace, application to model building and detection of coordinate errors. *J. Mol. Biol.* **218**, 183–194.
48. Bower, M.; Cohen, F.; Dunbrak, R. (1997) Prediction of protein side-chain rotamers from a backbone-dependent rotamer library. *J. Mol. Biol.* **267**, 1268–1282.
49. Kazmierkiewicz, R.; Liwo, A.; Scheraga, H. (2003) Addition of side chains to a known backbone with defined side-chain centroids. *Biophys. Chem.* **100**, 261–280.
50. Kabsch, W. (1978) A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A*, **34**, 828–829.
51. de-Mori, G. M. S.; Colombo, G.; Micheletti, C. (2005) Study of the villin headpiece folding dynamics by combining coarse-grained Monte Carlo evolution and all-atom molecular dynamics. *Proteins-struct. funct. bioinform.* **58**, 459 471.
52. Hobohm, U.; Scharf, M.; Schneider, R.; Sander, C. (1992) Selection of a representative set of structures from the brookhaven protein data bank. *Protein Sci.* **1**, 409–417.
53. Hobohm, U.; Sander, C. (1992) Enlarged representative set of protein structures. *Protein Sci.* **2**, 522.

54. Hengartner, M. O. (2000) The biochemistry of apoptosis. *Nature* **107**, 770–776.
55. Whitesell, L.; Lindquist, S. (2005) Hsp90 and the chaperoning of cancer. *Nat. Rev. Cancer* **5**, 761–772.
56. Takayama, S.; Reed, J. C.; Homma, S. (2003) Heat-shock proteins as regulators of apoptosis. *Oncogene* **22**, 9041–9047.
57. Dymock, B. W.; Drysdale, M. J.; McDonald, E.; Workman, P. (2004) Inhibitors of HSP90 and other chaperones for the treatment of cancer. *Expert. Opin. Ther. Patents* **14**, 837–847.
58. Morimoto, R. I.; Santoro, M. G. (1998) Stress-inducible responses and heat shock proteins: new pharmacologic targets for cytoprotection. *Nat. Biotechnol.* **16**, 833–838.
59. Whitesell, L.; Lindquist, S. (2005) Hsp90 and the chaperoning of cancer. *Nat. Rev. Cancer* **5**, 761–772.
60. Young, J. C.; Moarefi, I.; Hartl, F. U. (2001) Hsp90: A specialized but essential protein-folding tool. *J. Cell. Biol.* **154**, 267–273.
61. Morishima, Y.; Murphy, P. J.; Li, D. P.; Sanchez, E. R.; Pratt, W. B. (2000) Stepwise assembly of a glucocorticoid receptor hsp90 heterocomplex resolves two sequential ATP-dependent events involving first hsp70 and then hsp90 in opening of the steroid binding pocket. *J. Biol. Chem.* **275**, 18054–18060.
62. Marcu, M. G.; Chadli, A.; Bouhouche, I.; Catelli, M.; Neckers, L. M. (2000) The heat shock protein 90 antagonist novobiocin interacts with a previously unrecognized ATP-binding domain in the carboxyl terminus of the chaperone. *J. Biol. Chem.* **275**, 37181–37186.
63. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684.
64. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. (1987) The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271.
65. Hess, B.; Bekker, H.; Fraaije, J. G. E. M.; Berendsen, H. J. C. (1997) A linear constraint solver for molecular simulations. *J. Comp. Chem.* **18**, 1463–1472.
66. Miyamoto, S.; Kollman, P. A. (1992) SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water models. *J. Comp. Chem.* **13**, 952–962.
67. Spoel, D. v. d.; Lindahl, E.; Hess, B.; Buuren, A. R. v.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Feenstra, K. A.; Drunen, R. v.; Berendsen, H. J. C. (2004) *Gromacs User Manual version 3.2.* www.gromacs.org.
68. Daura, X.; Antes, I.; van Gunsteren, W. F.; Thiel, W.; Mark, A. E. (1999) The effect of motional averaging on the calculation of NMR-derived structural properties. *Proteins: Struct. Funct. Genet.* **36**, 542–555.
69. Guex, N.; Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723.
70. Stebbins, C. E.; Russo, A. A.; Schneider, C.; Rosen, N.; Hartl, F. U.; Pavletich, N. P. (1997) Crystal Structure of Hsp90–Geldanamycin complex: Targeting of a protein chaperone by an antitumor agent. *Cell* **89**, 239–250.
71. Solmajer, T.; Mehler, E. L. (1991) Electrostatic Screening in Molecular Dynamics simulations. *Protein Eng.* **4**, 911.
72. Hetenyi, C.; van der Spoel, D. (2002) Efficient docking of peptides to proteins without prior knowledge of the binding site. *Prot. Sci.* **11**, 1729–1737.
73. Lee, Y. S.; Marcu, M. G.; Neckers, L. (2004) Quantum chemical calculations and mutational analysis suggest heat shock protein 90 catalyzes trans–cis isomerization of geldanamycin. *Chem. Biol.* **11**, 991–998.
74. Roe, S. M.; Ali, M. M.; Meyer, P.; Vaughan, C. K.; Panaretou, B.; Piper, P. W.; Prodromou, C.; Pearl, L. H. (2004) The Mechanism of Hsp90 regulation by the protein kinase-specific cochaperone p50(cdc37). *Cell* **116**, 87–98.
75. Altieri, D. C. (2003) Validating Survivin as a Cancer Therapeutic Target. *Nat. Rev. Cancer* **3**, 46–54.
76. Isaacs, J. S.; Xu, W.; Neckers, L. (2003) Heat shock protein 90 as a molecular target for cancer therapeutics. *Cancer Cell* **3**, 213–217.
77. Neckers, L.; Ivy, S. P. (2003) Heat Shock Protein 90. *Curr. Op. Oncol.* **15**, 419–424.
78. Chen, Y. N.; Sharma, S. K.; Ramsey, T. M.; Jiang, L.; Martin, M. S.; Baker, K.; P. D. Adams; Bair, K. W. (1999) Selective killing of transformed cells by cyclin/cyclin-dependent kinase 2 antagonists. *Proc. Natl. Acad. Sci. USA* **96**, 4325–4329.
79. Sawada, M.; Hayes, P.; Matsuyama, S. (2003) Cytoprotective membrane-permeable peptides designed from the Bax-

binding domain of Ku70. *Nat. Cell Biol.* **5**, 352–357.

80. Swanton, C. (2004) Cell-cycle targeted therapies. *Lancet Oncol.* **5**, 27–36.
81. Vogelstein, B.; Lane, D.; Levine, A. J. (2000) Surfing the p53 network. *Nature* **408**, 307–310.
82. Johnstone, R. W.; Ruefli, A. A.; Lowe, S. W. (2002) Apoptosis: A link between cancer genetics and chemotherapy. *Cell* **108**, 153–164.
83. Vassilev, L. T.; Vu, B. T.; Graves, B.; Carvajal, D.; Podlaski, F.; Filipovic, Z.; Kong, N.; Kammlott, U.; Lukacs, C.; Klein, C. (2004) In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**, 844–848.
84. Sausville, E. A.; Tomaszewski, J. E.; Ivy, P. (2003) Clinical development of 17-allylamino, 17-demethoxygeldanamycin. *Curr. Cancer Drug Targets* **3**, 377–383.
85. Beere, H. M. (2004) "The stress of dying": The role of heat shock proteins in the regulation of apoptosis. *J. Cell Sci.* **117**, 2641–2651.
86. Kamal, A.; Boehm, M. F.; Borrows, F. J. (2004) Therapeutic and diagnostic implications of Hsp90 activation. *Trends Mol. Med.* **10**, 283–290
87. Guillemard, V.; Saragovi, H. U. (2004) Novel approaches for targeted cancer therapy. *Curr. Cancer Drug Targets* **4**, 313–326.
88. Muchmore, S. W.; Chen, J.; Jakob, C.; Zakula, D.; Matayoshi, E. D.; Wu, W.; Zhang, H.; Li, F.; Ng, S. C.; Altieri, D. C. (2000) Crystal structure and mutagenic analysis of the inhibitor-of-apoptosis protein survivin. *Mol. Cell* **6**, 173–182.
89. Li, A.; Daggett, V. (1998) Molecular Dynamics Simulation of the Unfolding of Barnase: Characterization of the Major Intermediate. *J. Mol. Biol.* **275**, 677–694.
90. Vogelstein, B.; Kinzler, K. W. (2004) Cancer genes and the pathways they control. *Nat. Med.* **10**, 789–799.
91. O'Dwyer, M. E.; Druker, B. J. (2000) Status of Bcr-Abl tyrosine kinase inhibitors in chronic myelogenous leukemia. *Curr. Op. Oncol.* **12**, 594–597.
92. O'Dwyer, M. E.; Druker, B. J. (2000) STI571: An inhibitor of the BCR-ABL tyrosine kinase for the treatment of chronic myelogenous leukemia. *Lancet Oncol.* **1**, 207–211.
93. Paez, J. G.; Janne, P. A.; Lee, J. C.; Tracy, S.; Greulich, H.; Gabriel, S.; Herman, P.; Kaye, F. J.; Lindeman, N.; Boggon, T. J.; Naoki, K.; Sasaki, H.; Fujii, Y.; Eck, M. J.; Sellers, W. R.; Johnson, B. E.; Meyerson, M. (2004) EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy *Science* **304**, 1497–1500.
94. Workman, P. (2003) Overview: Translating Hsp90 Biology into Hsp90 Drugs. *Curr. Cancer. Drug. Targets* **3**, 297–300.
95. Mehler, E. L.; Solmajer, T. (1991) Electrostatic effects in proteins. Comparison of dielectric and charge models. *Prot. Eng.* **4**, 903–910.
96. Daura, X.; Jaun, B.; Seebach, D.; van Gunsteren, W. F. v.; Mark, A. E. (1998) Reversible Peptide Folding in Solution by Molecular Dynamics Simulation. *J. Mol. Biol.* **280**, 925–932.
97. Catalyst (2005) Accelrys, San Diego CA, USA.
98. Ferris, J. P.; Devadas, B.; Huang, C. H.; Ren, W. Y. (1985) Nucleosides from carbohydrate adducts of diaminomaleonitrile. A novel synchapter of 5-amino-1-(beta-D-ribofuranosyl)imidazole-4-carboxamide and 5-amino-1-(beta-D-ribopyranosyl)imidazole-4-carboxamide. *J. Org. Chem.* **50**, 747–754.
99. Folini, M.; Brambilla, C.; Villa, R.; Gandellini, P.; Vignati, S.; Paduano, F.; Daidone, M. G.; Zaffaroni, N. (2005) Antisense oligonucleotide-mediated inhibition of hTERT, but not hTERC, induces rapid cell growth decline and apoptosis in the absence of telomere shortening in human prostate cancer cells *Eur. J. Cancer* **41**, 624–634
100. Vilenchik, M.; Solit, D.; Basso, A.; Huezo, H.; Lucas, B.; He, H.; Rosen, N.; Spampinato, C.; Modrich, P.; Chiosis, G. (2004) Targeting wide-range oncogenic transformation via PU24FCL, a specific inhibitor of Tumor Hsp90. *Chem. Biol.* **11**, 787–797.
101. Chiosis, G.; Lucas, B.; Shtil, A.; Huezo, H.; Rosen, N. (2002) Development of a purine-scaffold novel class of Hsp90 Binders that inhibit the proliferation of cancer cells and induce the degradation of Her2 tyrosine kinase. *Bioorg. Med. Chem.* **10**, 3555–3564.
102. Dymock, B. W.; Barril, X.; Brough, P. A.; Cansfield, J. E.; Massey, A.; McDonald, E.; Hubbard, R. E.; Surgenor, A.; Roughley, S. D.; Webb, P.; Workman, P.; Wright, L.; Drysdale, M. J. (2005) Novel, potent small-molecule inhibitors of the molecular chaperone Hsp90 discovered through structure-based design. *J. Med. Chem.* **48**, 4212–4215.
103. Wright, L.; Barril, X.; Dymock, B. W.; Sheridan, L.; Surgenor, A.; Beswick, M.;

Drysdale, M. J.; Collier, A.; Massey, A.; Davies, N.; Fink, A.; Fromont, C.; Aherne, W.; Boxall, K.; Sharp, S.; Workman, P.; Hubbard, R. E. (2004) Structure–Activity relationships in purine-based inhibitor binding to Hsp90 Isoforms. *Chem. Biol.* **11**, 775–785.

104. Holt, S. E.; Aisner, D. L.; Baur, J.; Tesmer, V. M.; Dy, M.; Ouellette, M.; Trager, J. B.; Morin, G. B.; Toft, D. O.; Shay, J. W.; Wright, W. E.; White, M. A. (1999) Functional requirement of p23 and Hsp90 in telomerase complexes *Genes & Devel.* **13**, 817–826.
105. Janin, Y. L. (2005) Heat Shock Protein 90 inhibitors. A text book example of medicinal chemistry? *J. Med. Chem.* **48**, 7503–7512.
106. Sausville, E. A. **(2004)** Geldanamycin Analogs. *J. Chemother.* **16**(S4), 68–69.
107. Giri, S.; Nath, N.; Smith, B.; Viollet, B.; Singh, A. K.; Singh, I. (2004) 5-aminoimidazole-4-carboxamide-1-beta-4-ribofuranoside inhibits proinflammatory response in glial cells: A possible role of AMP-activated protein kinase. *J. Neurosci.* **24**, 479–487.
108. Nath, N.; Giri, S.; Prasad, R.; Salem, M. L.; Singh, A. K.; Singh, I., (2005) 5-Aminoimidazole-4-carboxamide ribonucleoside: A novel immunomodulator with therapeutic efficacy in experimental autoimmune encephalomyelitis *J. Immunol.* **175**, 566–574.
109. Rattan, R.; Giri, S.; Singh, A. K.; Singh, I. (2005) 5-Aminoimidazole-4-carboxamide-1-beta-D-ribofuranoside inhibits cancer cell proliferation in vitro and in vivo via AMP-activated Protein Kinase. *J. Biol. Chem.* **280**, 39582–39593.
110. Paduano, F.; Villa, R.; Pennati, M.; Folini, M.; Binda, M.; Daidone, M. G.; Zaffaroni, N. (2006) Silencing of survivin gene by small interfering RNAs produces supra-additive growth suppression in combination with 17-allylamino-17-demethoxygeldanamycin in human prostate cancer cells. *Mol. Cancer Ther.* **5**, 179–186.
111. Rutledge, G.; Behnke, E.; Liang, Y. X.; You, S. W.; Tay, D. K. C.; Zhang, S.; So, K. F.; Schneider, G. E. (2006) Nano neuro knitting: Peptide nanofiber scaffold for brain repair and axon regeneration with functional return of vision. *PNAS* **103**, 5054–5059.
112. Daidone, I.; Amadei, A.; Di Nola, A. (2005) Thermodynamic and kinetic characterization of a ß-hairpin peptide in solution: an extended phase space sampling by molecular dynamics simulations in explicit water. *Proteins* **59**, 510–518.
113. Ma, B.; Nussinov, R. (2002) Molecular dynamics simulation of alanine beta-sheet oligomers: insight into amyloid formation. *Protein Sci.* 11, 2335–2350.
114. Urbanc, B.; Ding, C. L.; Sammond, D.; Khare, S. (2004) Molecular dynamics simulation of amyloid beta dimer formation. *Biophys. J.* **87**, 2310.
115. Dokholyan, N. V. (2006) Studies of folding and misfolding using simplified models. *Curr. Opin. Struct. Biol.***16**, 79–85.
116. Flock, D.; Daidone, I.; Di Nola, A. (2004) A Molecular Dynamics study of acylphosphatase in aggregation promoting conditions: the influence of TFE/water solvent. *Biopolymers* **75**, 491–496.
117. Wei, G.; Shea, J. E. (2006) Effects of solvent on the structure of the Alzheimer amyloid-beta(25-35) peptide. **91**, 1638–1647.
118. Cuniasse, P.; Raynal, I.; Yiotakis, A.; Dive, V. (1997) Accounting for Conformational Variability in NMR Structure of Cyclopeptides: Ensemble Averaging of Interproton Distance and Coupling Constant Restraints. *J. Am. Chem. Soc.* **19**, 5239–5248.
119. Favrin, G.; Irbäck, A.; Mohanty, S. **(2004)** Oligomerization of Amyloid Aß16–22 Peptides Using Hydrogen Bonds and Hydrophobicity Forces. *Biophys. J.***87**, 3657–3664.
120. Burkoth, T. S.; Benzinger, T. L. S.; Jones, D. N. M.; Hallenga, K.; Meredith, S. C.; Lynn, D. G. (1998) C-Terminal PEG Blocks the Irreversible Step in –Amyloid(10-35) Fibrillogenesis. *J. Am. Chem. Soc.* **120**, 7655–7656.
121. Burkoth, T. S.; Benzinger, T. L. S.; Urban, V.; Lynn, D. G.; Meredith, S. C.; Thiyagarajan, P. (1999) Self-Assembly of A(10-35)-PEG Block Copolymer Fibrils. *J. Am. Chem. Soc.* **121**, 7429–7430.
122. Zheng, J.; Tsai, M. B.; Nussinov, R. (2006) Structural stability and dynamics of an amyloid-forming peptide GNNQQNY from the yeast prion sup-35. *Biophys. J.* **91**, 824–833.
123. Walsh, D. M.; Klyubin, I.; Fadeeva, J. V.; Cullen, W. K.; Anwyl, R.; Wolfe, M. S.; Rowan, R. J.; Selkoe, D. J. (2002) Naturally secreted oligomers of amyloid beta protein potently inhibit hippocampal long-term potentiation in vivo. *Nature* **416**, 535–539.
124. Bucciantini, M.; Giannoni, E.; Chiti, F.; Baroni, F.; Formigli, L.; Zurdo, J.; Taddei, N.; Ramponi, G.; Dobson, C. M.;

Stefani, M. (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416**, 507–511.

125. Kayed, R.; Head, E.; Thompson, J. L.; McIntire, T. M.; Milton, S. C.; Cotman, C. W.; Glabe, C. G. (2003) Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* **300**, 486–489.
126. Mousseau, N.; Derreumaux, P. (2005) Exploring the Early Steps of Amyloid Peptide Aggregation by Computers. *Acc. Chem. Res.* **38**, 885–891.
127. Nilsson, M. R. (2004) Techniques to study amyloid fibril formation in vitro. *Methods* 34, 151–160.
128. Bitan, G.; Teplow, D. B. (2004) Rapid photochemical cross-linking-a new tool for studies of metastable, amyloidogenic protein assemblies. *Acc. Chem. Res.* **37**, 357–364.
129. Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. O.; Riekel, C.; Grothe, R.; Eisenberg, D. (2005) Structure of the cross-β spine of amyloid-like fibrils. *Nature* **435**, 773–778.
130. Fersht, A. R. (2002) On the simulation of protein folding by short time scale molecular dynamics and distributed computing. *PNAS* **99**, 14122–14125.
131. Derrick, J. P.; Wigley, D. B. (1992) Crystal-structure of a streptococal protein-G domain bound to an fab fragment. *Nature* **359**, 752–754.
132. Gabius, H.-J. (1998) The How and Why of Protein–carbohydrate Interaction: a Primer to Theoretical Concept and a Guide to application in drug Design. *Pharmaceutical Res.* **15**, 23–30.
133. Ban, T.; Yamaguchi, K.; Goto, Y. (2006) Direct Observation of Amyloid Fibril Growth, Propagation, and Adaptation. *Acc. Chem. Res.* **39**, 663–670.
134. Patino, M. M.; Liu, J. J.; Glover, J. R.; Lindquist, S. (1996) Support for the prion hypochapter for inheritance of a phenotypic trait in yeast. *Science* **273**, 622–626.
135. Serio, T. R. (2000) Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science* **289**, 1317–1321.

Section II

Applications

Chapter 5

Synthesis of Peptide Arrays Using SPOT-Technology and the CelluSpots-Method

Dirk F.H. Winkler, Kai Hilpert, Ole Brandt, and Robert E.W. Hancock

Abstract

Peptide synthesis on cellulose using the SPOT technology follows the standard Fmoc-chemistry and can be performed manually or automated. This method allows the synthesis of low-cost peptide arrays containing around 900 large spots of addressable peptides on a cellulose sheet of 19 cm × 29 cm. These peptides can be cleaved from the cellulose support by ammonia gas and afterward spotted on glass microchips. Alternatively, the peptides can be synthesized on modified cellulose discs and CelluSpot microarrays can be produced.

Key words: Spot synthesis, peptide array, screening.

1. Introduction

Solid-phase peptide synthesis has been automated for tBoc- as well as for Fmoc-chemistry. Established methods can be used to synthesize large numbers of peptides for screening procedures, including tea bag synthesis (1), digital photolithography (2), pin synthesis (3), and SPOT^{TM} synthesis on cellulose (4). Additionally, biological techniques such as phage (5), bacterial (6), or ribosome display (7) can be applied to synthesize and screen large numbers of peptides.

The success of DNA and protein microchips has inspired researchers to develop microchips for peptides arrays (8) (*see* **Table 5.1**). The small amounts of peptides used to fabricate microchips, together with existing technologies to analyze binding events on chips, has opened up new directions for peptide libraries. Several hundreds of different peptides can be placed on a microchip. The effort of synthesizing all these different peptides is high, but is balanced by the large number of chips that can be synthesized with the same peptide set. In particular for medical screens, this type of screening strategy is required since the purity of the peptides needs to be

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_5,

Table 5.1
Recommended literature

Description	References
Handbook about spot synthesis and applications	(45)
Detailed protocol for spot synthesis	(30)
Characterization of some useful types of filter paper and description of some linkers	(44)
Investigations on the quality of spot syntheses	(14)
Detailed review about synthesis and application using spot technique	(12)
Review about peptide arrays	(8)

very high. However, synthesis of peptide sets on resin is often very expensive. One strategy to overcome this issue is to synthesize the required peptide sets on cellulose membranes. The SPOT synthesis method described here permits the synthesis of multiple peptides for less than 1% of the cost compared to common synthesis on resin. In addition, we describe the synthesis of peptides on cellulose discs using a conventional peptide synthesizer. The synthesized peptides can then either be released as free peptides to produce common peptide microarrays on glass chips (**Fig. 5.1**), or remain bound to cellulose for direct screening on that membrane as well as for the production of CelluSpots microarrays (9, 10).

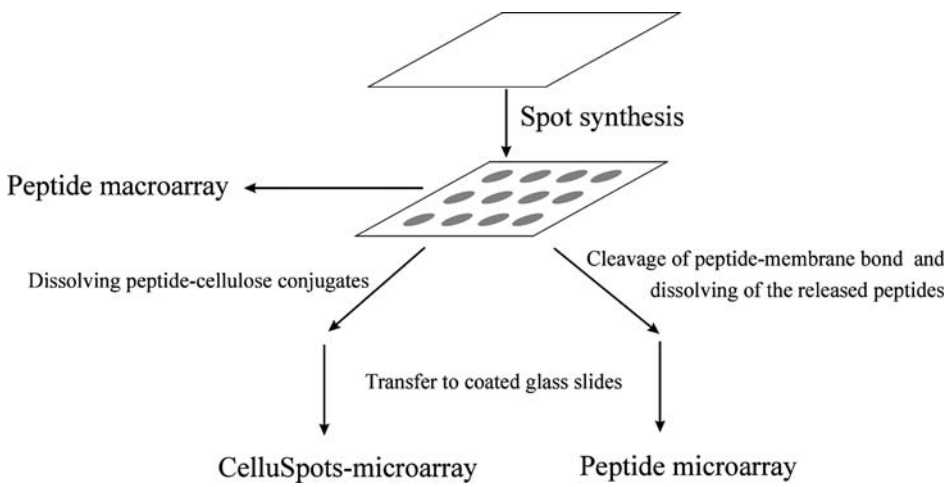


Fig. 5.1. Outline of the preparation of peptide arrays using spot synthesis.

SPOT synthesis was developed by Frank and co-workers in 1990 (11). This special type of solid-phase peptide synthesis follows standard Fmoc-chemistry using cellulose as the solid support (12).

Initially the method was performed manually. Although it is easy to perform and no special equipment is necessary, it is quite laborious to perform by hand and, therefore, is only feasible for a small number of peptides. Automation of the method permitted the synthesis of up to 8000 different peptides on a single sheet (19 cm × 29 cm) in a highly parallel and addressable manner (13), and this improvement has led to a great variety of applications of SPOT synthesis (14) (**Fig. 5.2**). The large number of different peptides synthesized in a parallel manner results in differences in coupling yields, and, therefore, in the quality and quantity of the different peptides. Nevertheless, the SPOT technology has been frequently reported to be very reliable (14–16). For transferring peptides synthesized on cellulose onto a microchip, a larger peptide amount is required than for direct screening on cellulose membranes. A sufficient amount of peptide can be obtained by synthesis on cellulose membranes with larger peptide spots or by multiple syntheses on cellulose discs. Using large cellulose sheets (19 cm × 29 cm) around 900 large peptide spots can be synthesized.

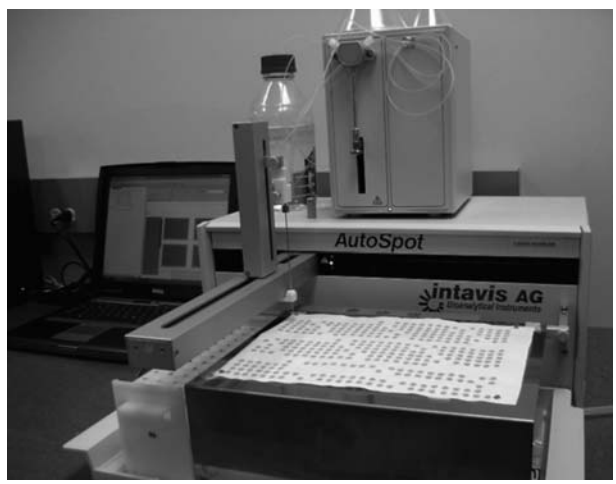


Fig. 5.2. Semi-automated synthesis of large spots using AutoSpot robot (INTAVIS).

Peptide synthesis on cellulose has several advantages: cellulose is inexpensive and withstands the organic solvents and acids used during peptide synthesis. In addition, cellulose is stable in aqueous solutions and, because it is non-toxic, it is appropriate for screening biological samples. Furthermore, peptides up to a length of about 40 amino acids can be synthesized using the SPOT method (e.g., 17, 18), although the optimal range is between 6 and 15 amino acids. With shorter peptides, the affinity for binding reactions might be too low. For the synthesis of long peptides, the following must be considered: with the increase of peptide length the likelihood of difficulties during coupling increases, such that the number of possible side reactions might increase and purity of the desired peptide will correspondingly decrease.

In contrast to biological peptide synthesis, SPOT technology is not restricted to gene-encoded amino acids, and more than 600 commercially available building blocks can be used (e.g., 19, 20). Another advantage of using SPOT synthesis on cellulose is the possibility of modifying the peptide. For detection purposes peptides can be labeled with, e.g., biotin (21, 22), dyes (23), or fluorophores (24). The biotinylation of peptides is also useful for the directed immobilization of peptides to streptavidin-coated glass slides to produce peptide microarrays (25). Maleimide-functionalized peptides can be attached to thiol-modified array surfaces (26). Various types of cyclization strategies can also be performed using this technology (e.g., 27, 28). This allows the synthesis and screening of more rigid molecules that might be more resistant to proteolytic cleavage and heat denaturation.

Peptides synthesized using the SPOT technology can be released from the cellulose support by treatment with ammonia gas or other strong bases such as tertiary amines, and transferred to glass slides. Some peptide arrays have the disadvantage that the peptide molecules can interact with the glass surface in a fashion that might interfere with the screening process. Alternatively, a TFA-soluble cellulose membrane can be used to obtain soluble peptide–cellulose conjugates (29). The resulting solutions of individual peptides covalently linked to macromolecular cellulose can then be spotted onto planar surfaces such as coated glass slides, to produce CelluSpots microarrays. After evaporation of the solvent, a three-dimensional layer is formed that is not soluble in the aqueous reagents used for standard assays (**Fig. 5.3**). It is possible to use detection methods such as autoradiography, chemiluminescence, or enzymatic color development which can be performed without expensive instrumentation.

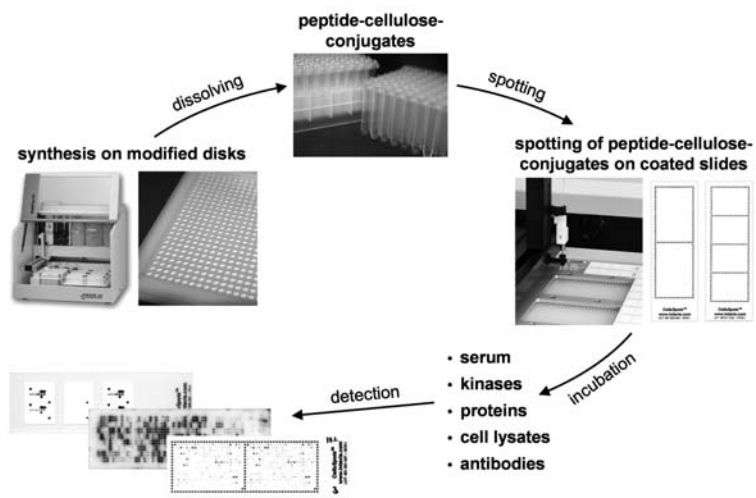


Fig. 5.3. Scheme of preparation cycle for CelluSpots peptide microarrays.

2. Materials

2.1. Solvents

1. *N,N′*-dimethylformamide (DMF; VWR).
2. Methanol or ethanol (MeOH or EtOH; VWR).
3. *N*-methylpyrrolidone (NMP; Fluka).
4. Diethyl ether (DEE; VWR) or *t*-butyl methyl ether (MTBE, VWR).
5. Dichloromethane (methylene chloride, DCM; VWR) (*see* **Note 1**).

2.2. Spot Synthesis of Macroarrays

2.2.1. Preparation of the Membrane

1. Cellulose membranes for releasing peptides from membrane for the preparation of peptide arrays are prepared from filter paper Whatman 50, Whatman 540, or Chr1 (Whatman, Maidstone, UK) (4, 30). Ready-to-use modified membranes for synthesis of peptides attached onto soluble cellulose for the preparation of CelluSpots microarrays are available from INTAVIS (Köln, Germany).
2. Diisopropylcarbodiimide (DIPC, DIC; Fluka, *see* **Note 2**).
3. *N*-methylimidazole (NMI; Sigma).
4. Fmoc-β-alanine (GL Biochem or EMD Biosciences, *see* **Note 4**).

2.2.2. Preparation of Activated Amino Acid Solutions

1. Coupling reagents: diisopropylcarbodiimide (DIPC, DIC; Fluka) and *N*-hydroxybenzotriazole (HOBt; EMD Biosciences). Coupling reagents are necessary only if non-pre-activated amino acid derivatives are used (*see* **Note 2**).
2. For in situ activation of amino acids and other building blocks, utilize derivatives compatible with the Fmoc-protection strategy (31, 32) (GL Biochem or EMD Biosciences). Pre-activated amino acid derivatives must also be derivatives with protecting groups according to the Fmoc-protection strategy. These are usually pre-activated as pentafluorophenyl esters (OPfp ester; EMD Biosciences or Bachem) (33) (*see* **Note 2**).

2.2.3. Spot Synthesis on Planar Cellulose Membranes

1. Solution for Fmoc-deprotection: 20% (v/v) piperidine (Sigma) in DMF.
2. Capping solution: 2% (v/v) acetic anhydride (Sigma) in DMF. To support the reaction, 2% (v/v) ethyl-diisopropylamine (DIPEA, DIEA; VWR or Sigma) should be added.
3. Staining solution: 0.002% bromophenol blue (BPB; Sigma) in MeOH (20 mg in 1 L).
4. Cleavage solution I: 90% trifluoroacetic acid (v/v) (TFA; VWR) + 5% (v/v) dist. water + 3% (v/v) triisopropylsilane or triisobutylsilane (TIPS or TIBS; Aldrich) + 1% (w/v) phenol (Sigma) + 1% (v/v) DCM.

5. Cleavage solution II: 60% (v/v) TFA + 3% (v/v) TIPS or TIBS + 2% (v/v) dist. water + 1% (w/v) phenol + 44% (v/v) DCM (*see* **Note 3**).
6. Ammonia gas (for releasing peptides from cellulose membrane).

2.3. Preparation of the Peptide Microarrays (CelluSpots Arrays)

2.3.1. Synthesis of Peptide–Cellulose Conjugates for CelluSpots Microarrays

1. Modified membranes (INTAVIS) or multi-well plates holding 384 modified membrane discs (INTAVIS) for the synthesis of peptide–cellulose conjugates.
2. Solution for Fmoc-deprotection: 20% (v/v) piperidine (Sigma) in DMF.
3. Cleavage solution (CelluSpots): side-chain cleavage solution CA: 80% (v/v) trifluoroacetic acid (TFA; VWR) + 5% (v/v) dist. water + 3% (v/v) triisopropylsilane (TIPS; Aldrich) + 12% (v/v) DCM.
4. Cleavage solution CB: 88.5% (v/v) trifluoroacetic acid (TFA; VWR) + 4% (v/v) trifluoromethanesulfonic acid (TFMSA; Fluka) + 5% (v/v) dist. water + 2.5% (v/v) triisopropylsilane (TIPS; Aldrich) (*see* **Note 3**).
5. Dimethyl sulfoxide (DMSO; Fluka).
6. SSC buffer (20x): dissolve 175 g sodium chloride (Sigma) + 88 g tribasic sodium citrate dihydrate (Fluka) in 1 L dist. water, adjust to pH 7.0 with 1 M HCl.

3. Methods

3.1. Spot Synthesis of Macroarrays

Spot synthesis of peptides can be carried out manually, semi-automated, or fully automated. Manual spot synthesis is useful for the synthesis and screening of a limited number of peptides (up to 100) with a pipetting volume per spot of about 1 μL or more. For all other situations, a semi-automated or fully automated synthesis is recommended. Spot synthesis with pipetting volumes per spot of 0.5 μL and more can be performed by most of commercially available *x*–*y*–*z* programmable pipetting robots. For synthesis of smaller spots with a pipetting volume less than 0.5 μl, spot synthesizers from INTAVIS AG (Köln, Germany) are recommended (34). In principle, for all types of spot synthesis, the preparation and synthesis procedure and final treatment of the membranes are similar and can be carried out according to the protocol below.

When not noted, all methods correspond to the preparation of an approximately 19 cm × 29 cm cellulose membrane, which can accommodate about 400 large spots with a diameter of about 7 mm. The synthesis on smaller membranes follows the same procedures. When the phrase "use of non-pre-activated or in situ

activated amino acids" appears, it includes the use of other organic building blocks (e.g., unnatural amino acids, PNA monomers, peptoidic elements) (35, 36), which can be used under conditions where spot synthesis is performed. Here we describe only the basic procedures for spot synthesis of linear conventional peptides to yield free peptides for the preparation of peptide arrays. The synthesis of modified peptides including cyclizations or side-chain modifications was described previously (30).

3.1.1. Preparation of a Cellulose Membrane for Spot Synthesis

1. Cut a piece of filter paper to the size required to accommodate all peptide spots including controls (here 19 cm × 29 cm).
2. For amine functionalization of filter paper, dissolve 0.64 g Fmoc-β-alanine in 10 mL amine-free DMF (*see* **Notes 1** and **4**). Add 374 μL DIC and 317 μL NMI, mix well and transfer the reaction solution into a chemically resistant box with a lid. To achieve homogeneous functionalizations of the cellulose membrane, avoid the presence of air bubbles under the paper during placement of the filter paper in the box and ensure that the surface of the membrane is slightly covered by the solution. Close the box and leave the membrane in the reaction mixture for at least 2 h, or overnight.
3. Wash the membrane three times with DMF for at least 30 s each. If the membrane needs to be stored for a limited time, wash the modified membrane at least twice with methanol or ethanol and dry it in an air stream in a fume hood, or by using a hair dryer without heat (*see* **Note 5**). The membrane can be stored at −20°C (if the membrane is stored, warm up the membrane to room temperature and wash the membrane once with DMF for at least 20 min prior to step 4).
4. For Fmoc-deprotection treat the membrane twice with 20% piperidine in DMF for at least 5 min each.
5. Staining (optional) (37): wash the membrane four times with DMF and then at least twice with methanol or ethanol for at least 30 s each. Treat the membrane with staining solution until the filter paper shows a homogeneous blue color. After staining wash the membrane at least twice with methanol or ethanol, until the wash solution remains colorless. Dry the membrane in air.

3.1.2. Preparation of Activated Amino Acid Solutions

Preparation of the coupling solutions can be performed according to two different protocols. One method involves synthesis using preactivated Fmoc-protected amino acids (e.g., pentafluorophenyl ester). The advantage of this method is the use of only one reagent for each solution, making the preparation of the amino acid solutions very simple and the likelihood of mistakes low. A disadvantage is the higher price of the amino acid derivatives; however, due to the small amounts of activated amino acids used, the absolute difference is only

a few dollars for the synthesis of an entire peptide membrane array. Another disadvantage is the fact that activated esters are only commercially available for the standard amino acids.

The second procedure is based on the use of in situ activated amino acids. Activation of the amino acids is carried out by adding an activator and coupling reagent to the non-activated Fmoc-protected amino acid derivative. Although the preparation according to this method is more laborious than the first described method, it has the advantage that it can be applied for all derivatives of amino-acid-like building blocks that can be protected according to the Fmoc-protecting group strategy.

Both methods can also be used for automated synthesis of peptide–cellulose conjugates on modified cellulose discs for the preparation of CelluSpots arrays (**Section 3.2**). For that purpose one must use commercially available CelluSpots membranes (INTAVIS).

Method 1: preparation of coupling solutions using pre-activated amino acid derivatives

Except for the arginine derivative, all solutions of pre-activated amino acids can be prepared and used for a week. Dissolve all amino acid derivatives, except serine pentafluorophenyl ester, in amine-free NMP at a concentration of 0.3 M (*see* **Table 5.2**). Due to poor solubility, serine pentafluorophenyl ester must be dissolved in amine-free DMF. Solutions can be stored in sealed tubes at –20°C for at least one week. At room temperature, the solutions are stable for about 1 day. Therefore, before starting the first synthesis cycle of a day, always discard the solutions from the previous day and replace

Table 5.2
Weights (mg) of amino acid pentafluorophenyl esters for 0.3 M solutions for spot synthesis using pre-activated derivatives

Amino acid		MW (g/mol)	0.5 mL	0.75 mL	1.0 mL
	Fmoc-β-Ala-OPfp	477.39	71.6	107.4	143.2
A	Fmoc-Ala-OPfp	477.39	71.6	107.4	143.2
C	Fmoc-Cys(Trt)-OPfp	751.79	112.8	169.2	225.5
D	Fmoc-Asp(OtBu)-OPfp	577.5	86.6	129.9	173.3
E	Fmoc-Glu(OtBu)-OPfp	591.50	88.7	133.1	177.5
F	Fmoc-Phe-OPfp	553.49	83.0	124.5	166.0
G	Fmoc-Gly-OPfp	463.36	69.5	104.3	139.0
H	Fmoc-His(Trt)-OPfp	785.78	117.9	176.8	235.7

(continued)

Table 5.2 (continued)

Amino acid		MW (g/mol)	0.5 mL	0.75 mL	1.0 mL
I	Fmoc-Ile-OPfp	519.47	77.9	116.9	155.8
K	Fmoc-Lys(Boc)-OPfp	634.60	95.2	142.8	190.4
L	Fmoc-Leu-OPfp	519.47	77.9	116.9	155.8
M	Fmoc-Met-OPfp	537.51	80.6	120.9	161.3
N	Fmoc-Asn(Trt)-OPfp	762.74	114.4	171.6	228.8
P	Fmoc-Pro-OPfp	503.43	75.5	113.3	151.0
Q	Fmoc-Gln(Trt)-OPfp	776.77	116.5	174.8	233.0
R	Fmoc-Arg(Pbf)-OPfp	814.84	122.2	183.3	244.5
S	Fmoc-Ser(tBu)-OPfp	549.5	82.4	123.6	164.9
T	Fmoc-Thr(tBu)-OPfp	563.52	84.5	126.8	169.1
V	Fmoc-Val-OPfp	505.44	75.8	113.7	151.6
W	Fmoc-Trp(Boc)-OPfp	692.70	103.9	155.9	207.8
Y	Fmoc-Tyr(tBu)-OPfp	625.60	93.8	140.8	187.7

them by transferring the appropriate amount for that day from the stock solutions. Due to the instability of dissolved pre-activated arginine derivatives, this solution must be prepared fresh every day.

Method 2: Preparation of coupling solutions using in situ activated derivatives of amino acids or other building blocks

Prepare a 0.9 M solution of HOBt in amine-free NMP. Dissolve the Fmoc-amino acids or protected building blocks to a concentration of 0.45 M using the HOBt-solution (*see* **Table 5.3**). Except for the arginine derivatives, these solutions can be stored at –20°C for at least a week. Each day prepare a 20% mixture of DIC in amine-free NMP (*see* **Notes 1** and **2**). Replace the coupling solutions of the previous day by preparation of a fresh one. Then take the desired amount of the amino acid/HOBt stock solution and add to those 20% DIC/NMP at a ratio of 3:1 (e.g., 75 µL amino acid solution and 25 µL of fresh prepared DIC).

3.1.3. Spot Synthesis on Planar Cellulose Membranes

1. Definition of the spot pattern: place the dry membrane on a planar surface. Deliver the required volume of activated Fmoc-β-alanine/DMSO solution to all spot positions (*see* **Note 4**). Repeat the delivery after 20 min, and allow the reaction to proceed once more for at least another 20 min.

Table 5.3
Weights (mg), prior to adding DIC solution, of amino acid derivative solutions used in spot synthesis with non-pre-activated derivatives. If after dissolving of amino acids in the corresponding amount of HOBt solution the volume does not achieve the value described in the table, make up the volume with NMP

Amino acid		MW (g/mol)	0.5 mL	0.75 mL	1.0 mL
	Fmoc-β-Ala-OH	311.3	70.0	105.1	140.1
A	Fmoc-Ala-OH	311.3	70.0	105.1	140.1
C	Fmoc-Cys(Trt)-OH	585.7	131.8	197.7	263.6
D	Fmoc-Asp(OtBu)-OH	411.5	92.6	138.9	185.2
E	Fmoc-Glu(OtBu)-OH	425.5	95.7	143.6	191.5
F	Fmoc-Phe-OH	387.4	87.2	130.7	174.3
G	Fmoc-Gly-OH	297.3	66.9	100.3	133.8
H	Fmoc-His(Trt)-OH	619.7	139.4	209.1	278.9
I	Fmoc-Ile-OH	353.4	79.5	119.3	159.0
K	Fmoc-Lys(Boc)-OH	468.5	105.4	158.1	210.8
L	Fmoc-Leu-OH	353.4	79.5	119.3	159.0
M	Fmoc-Met-OH	371.5	83.6	125.4	167.2
N	Fmoc-Asn(Trt)-OH	596.7	134.3	201.4	268.5
P	Fmoc-Pro-OH	337.4	75.9	113.9	151.8
Q	Fmoc-Gln(Trt)-OH	610.7	137.4	206.1	274.8
R	Fmoc-Arg(Pbf)-OH	648.8	146.0	219.0	292.0
S	Fmoc-Ser(tBu)-OH	383.4	86.3	129.4	172.5
T	Fmoc-Thr(tBu)-OH	397.5	89.4	134.2	178.9
V	Fmoc-Val-OH	339.4	76.4	114.5	152.7
W	Fmoc-Trp(Boc)-OH	526.6	118.5	177.7	237.0
Y	Fmoc-Tyr(tBu)-OH	459.6	103.4	155.1	206.8
Added 0.9 M HOBt-solution			0.38 mL	0.56 mL	0.75 mL

2. Blocking free areas: add an appropriate amount of capping solution to the box and place the membrane face down into the solution. Avoid air bubbles under the surface. Do not shake! After a minimum of 5 min of reaction time, take out

the membrane and pour out the used capping solution. Add an appropriate amount of capping solution with 2% DIPEA and place the membrane face up in this solution for 20 min.

3. Fmoc-deprotection: wash four times with DMF for at least 30 s each. Treat the membrane twice with 20% piperidine/DMF for 5 min each. Wash three times with DMF for at least 30 s each. Wash at least twice with MeOH or EtOH.
4. Staining (optional) (37): shake the membrane with staining solution in a box. If the staining solution changes to blue very quickly, renew the solution. Let the reaction proceed for at least 15 s until spots are visible (*see* **Note 6**). Wash at least twice with MeOH or EtOH until the wash solution remains colorless.
5. Dry the membrane in an air stream. If faster drying is necessary, additionally wash the membrane once or twice with DEE before drying. The membrane is now ready for a coupling cycle.
6. Coupling step: Deliver the prepared coupling solutions of desired activated amino acids to the corresponding positions on the membrane with the required volumes (*see* **Note 7**). Repeat the spotting after 20 min.
7. Blocking unreacted free amino groups (capping): place the membrane face down for at least 5 min in an appropriate amount of capping solution. Do not shake! Take out the membrane and pour out the used capping solution. Add an appropriate amount of capping solution with 2% DIPEA and place the membrane face up in this solution for at least another 5 min.
8. Building up the peptide chain: peptides are built up step by step (one amino acid at a time), starting from the C-terminus to the N-terminus. Except for the last coupling cycle, repeat steps 3–7. At the last coupling cycle carry out the steps above without capping and staining!
9. Final Fmoc-deprotection: Wash four times with DMF for at least 30 s each. Treat the membrane twice with 20% piperidine/DMF for 5 min. Wash again four times with DMF followed by three times washing with DCM for at least 30 s each.
10. Final side-chain deprotection (*see* **Note 3**): treat home-made membranes with cleavage solution I for 30 min. For each small membrane (96-well plate size) use at least 25 mL cleavage solution and for a large membrane (19 cm $\times$ 29 cm) at least 80 ml. The surface of the membrane must be well covered by the cleavage solution. Keep the box closed. Do not shake! Pour off the solution very carefully. Wash the membrane five times with DCM for at least 1 min each. Treat home-made membranes with cleavage solution II for 3 h. Use at least the same volume as for cleavage solution I. The surface of the membrane must be well covered by the cleavage solution. To avoid evaporation keep the

box tightly closed. Do not shake! Pour off the solution very carefully. Commercially available membranes are usually more stable than the homemade, which is why most of them can be treated with cleavage solution I for 3.5 h. Wash the membrane five times for at least 1 min each with PBS or TBS until the pH value is around 7 (*see* **Note 8**). Wash the membrane three times with water and then three times MeOH or EtOH for at least 1 min each. Dry the membrane in the air stream of a fume hood or with a hair dryer without heat.

11. For the preparation of CelluSpots arrays punch out the spots, transfer all single discs into deep-well plates or 96-well tube racks with a minimum holding capacity of 1 mL and proceed with the preparation according to the instructions of **Section 3.2.2**.

12. For the preparation of common peptide arrays, it is necessary to release the peptides from the membrane. One method is to treat the dry membrane or punched-out spots overnight in a glass desiccator containing ammonia gas. The strong basic environment leads to a cleavage of the ester bond to the cellulose by forming a C-terminal amide (*see* **Note 9**). If not done previously, on the next day punch out the spots and place the discs into wells of a microtiter plates or into vials in which one can dissolve the released peptides (*see* **Note 10**). If the peptides need to have a free carboxyl group at the C-terminus, another treatment is necessary. In this case punch out the spots from the membrane and transfer the discs into wells of microtiter plates or into vials. Then treat them with an aqueous basic solution, for example ammonium hydroxide, aqueous solution of triethylamine (38), or sodium hydroxide (39).

Another method is to release the peptide by forming C-terminal diketopiperazines. Therefore, modify the membrane with Boc-Lys(Fmoc)-OH instead of the Fmoc-β-Ala-OH, followed by coupling of Fmoc-proline as first coupled amino acid. After TFA treatment, punch out the spots and transfer them into microtiter plates or vials (*see* **Note 10**). The peptides are then released by treatment with basic aqueous buffers (pH $\geq$ 7.5) overnight (1).

3.2. Preparation of the Peptide Microarrays (CelluSpots Arrays)

Peptide–cellulose conjugates for the preparation of CelluSpots microarrays on coated microscope slides can be synthesized on modified membranes or in multi-well plates holding up to 384 modified discs in a semi-automated or fully automated manner. For semi-automated synthesis on CelluSpots membranes using the SPOT-technique follow the instructions under **Section 3.1**.

3.2.1. Synthesis of Peptide – Cellulose Conjugates for CelluSpots Microarrays

1. 384-Well filter-plates containing modified cellulose discs are placed into the synthesizer.
2. Fmoc-deprotection: to each well with a disc add 2X 4 μL of a 20% piperidine/DMF solution for 5 and for 10 min, respectively. Wash seven times with DMF (25 μL per well/disc) for at least 10 s each. Wash six times with EtOH or MeOH (25 μL per well/disc).
3. Dry the discs in a steady air stream for at least 12 min. The synthesis plate with the discs is now ready for the first coupling cycle.
4. In situ activation of amino acids and coupling: for each cycle all necessary amino acids are to be in situ activated with HOBt and DIC. To each disc add 1.2 μL of a coupling mix of 0.33 mole eq. DIC, 0.45 mole eq. HOBt, and 0.3 mole eq. Fmoc-amino acid (*see* **Note 11**). Repeat the distribution after a reaction time of 20–40 min.
5. Blocking un-reacted free amino groups on peptide fragments: add to each disc, 4 μL of capping solution. Allow the reaction to occur for at least 5 min. Remove the liquid and wash seven times with DMF (25 μL per well/disc) for at least 10 s each. Wash six times with EtOH or MeOH (25 μL per well/disc) and dry the discs in a steady air stream for at least 12 min.
6. Building up the peptides on modified cellulose discs: repeat steps 2–5.
7. Final Fmoc-deprotection: treat each disc with 4 μL of a 20% piperidine/DMF for 5 min. Repeat the treatment twice for 10 min and then once for 15 min. Wash seven times with DMF (25 μL per well/disc) for at least 10 s each. If N-terminal acetylated peptides are needed add three times 4 μL capping solution and let it react for at least 5 min. Wash another seven times with DMF, followed by six washing steps with EtOH or MeOH (25 μL per well/disc). After drying in a steady air stream for at least 12 min the discs are ready for the side-chain deprotection and dissolving (*see* **Note 12**).
8. Transfer all single discs from the synthesis plate into deep-well plates or 96-well tube racks with a minimum holding capacity of 1 mL.
9. Side-chain deprotection: add 150 μL of the side-chain cleavage solution CA to each well and incubate for 1–2 h in a fume hood (*see* **Note 13**). Remove the cleavage solution by suction, e.g., with a multi-channel device connected to a water jet pump or by a multi-channel pipette (*see* **Note 3**).

3.2.2. Preparation of Peptide–Cellulose Conjugate Solutions

1. Cleavage and dissolving of discs: add 250 μL of the cleavage solution CB to each well and close the plate with a lid or with stopper stripes (*see* **Note 13**). Incubate overnight in a hood,

until the cellulose discs are completely dissolved (*see* **Note 3**). Add 750 µL of ice-cold ether to each well and close the plate with a lid or with stopper stripes. Shake the plate gently for 10 s and let it cool down at –20 or –70°C for 1 h. Leave the plate in the freezer over night or spin down the precipitated peptide–cellulose conjugates (*see* **Note 14**).

2. Remove the supernatant and wash the pellet one or two times with fresh ether. Remove the remaining ether by evaporation under a fume hood until the residue remains as a gel (*see* **Note 15**).
3. To dissolve the conjugates, add 500 µL DMSO to each well. Close the plate with caps and store the stock solutions at –20°C.

3.2.3. Preparation of CelluSpots Microarrays

1. Spotting of peptide–cellulose conjugates onto coated microscope slides: take a new plate (e.g., 384-well plate for spotting) and dilute the peptide–cellulose conjugates from stock solutions in DMSO or in a mixture of DMSO and SSC buffer (2 parts DMSO and one part 1X SSC buffer). Dilutions of 1:2–1:6 can be tested (*see* **Note 16**). Deliver about 80 nL of that solution to coated microscope slides. Different types of slides can be used. The spotting of identical CelluSpots slides can be done by using a conventional contact printer (e.g., Slide Spotting Robot, INTAVIS). For detection via chromogenic substrate reactions, white-coated glass slides (e.g., INTAVIS) are favored.
2. Let the spots dry in air and put the slides into an oven at 75°C for 1 h. Until use, CelluSpots peptide arrays should be kept in a dry and dark place at 4°C. Under these conditions the arrays are stable for several months.

4. Notes

1. Solvents for dissolving reagents must be amine and water free. To check whether the solvent is sufficiently amine free, add to 1 mL of the solvent few microliters of the methanolic BPB solution. If it remains yellow or turns only to yellow green, the solvent can be used (34). All other solvents should be at least ACS or synthesis grade. Organic solvents should be stored in the dark.

2. All synthesis reagents must be protected from moisture. Reagent bottles and vials with solutions must be sealed with parafilm before storage. To avoid condensation, reagent bottles stored in the fridge or freezer should not be opened until they are warmed up to room temperature for about 30 min.
3. Safety: for all handling with TFA lab coat, gloves, safety glasses and fume hood are mandatory!
4. Due to the flexibility and linear structure of the molecule, β-alanine is mostly used for modification of the cellulose. However, the use of other amino acids is also possible for the functionalization of the membrane (40). For example, if all peptides have the same C-terminal amino acid, the corresponding derivative of that amino acid can be used for the amine functionalization of the membrane. The necessary amount of that derivative must be calculated according to its molecular weight. After cleavage from the cellulose, the amino acid used for functionalization remains on the peptide.
5. The membrane can be stored at –80°C for several months until use. For longer storage a loss of peptide is possible. Storage of the membranes at –80°C is recommended. The membranes can also be stored at –20°C or +4°C; but the risk of losing activity over a long storage time increases with the storage temperature and the length of the storage.
6. Do not stain the spots too much. If the absorbed amount of bromophenol blue is too high, some of the dye could become incorporated in the peptide. The removal of that incorporated dye is almost impossible and might affect the detection. Since bromophenol blue only turns blue in a basic environment, e.g., in the presence of primary amino functions, due to different acidity of the coupled amino acids and the built-up peptide chain, a difference in the intensity and type of the color of the spots is normal.
7. To ensure that the whole spot area is covered by the coupling solution, it is recommended to use at least 20% more amino acid solution volume than in step 1 of **Section 3.1.3**.
8. After TFA treatment, the home-made membranes might become very soft and can be easily damaged due to strong agitation. Be careful and shake very gently! Do not try to lift the membrane out until it becomes harder and less likely to break apart during the washing steps (at around the time of the MeOH washing step)!
9. Release of peptides with ammonia might work only on membranes prepared according to the instructions of **Section 3.1.1**. Commercially available amino-functionalized

membranes often have bonds other than ester bonds between the peptide and cellulose. For such membranes their modification with additional linkers is necessary (e.g., thioester (41), HMB linker (42), and Rink linker (43)).

10. Due to chromatographic effects during the transfer of the coupling solution to the spots, the concentration of side products increases toward the edge of the spots. To reduce the amount of side products in the released peptides, the discs should be punched out in the center of the spot and their diameter should be smaller than the spot diameter.
11. It is also possible to use solutions of pre-activated amino acids. For preparation of those, please follow instructions in method 1 in **Section 3.1.2**.
12. The discs must be completely dry. Otherwise, remaining traces of EtOH or MeOH can lead to partial esterification of the peptides during TFA treatment.
13. If the cellulose discs or solutions turn from yellow to red, add a few more microliters of TIPS to the cleavage mix.
14. It is prohibited to spin open tubes with ether solutions in a non-explosion-proof centrifuge!
15. Take care that the pellet is not completely dry before adding the DMSO at a subsequent step – its consistency should be like a gel. If the pellet is dried completely, it will not dissolve in DMSO!
16. Especially if aqueous solutions like SSC buffers are used, some amount of conjugates may precipitate and can be centrifuged down. There will still be sufficient conjugate in solution.

References

1. Houghten, R. A. (1985) General method for the rapid solid-phase synthesis of large numbers of peptides: Specificity of antigen–antibody interaction at the level of individual amino acids. *Proceedings of the National Academy of Sciences (USA)* **82**, 5131–5135.
2. Pellois, J. P., Zhou, X.; Srivannavit, O., Zhou, T., Gulari, E. and Gao, X. (2002) Individually addressable parallel peptide synthesis on microchips. *Nature Biotechnology* **20**, 922–926.
3. Geysen, H. M., Meloen, R. H. and Barteling, S. J. (1984) Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. *Proceedings of the National Academy of Sciences (USA)* **82**, 3998–4002.
4. Frank, R. (1992) Spot-synthesis: An easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. *Tetrahedron* **48**, 9217–9232.
5. Paschke, M. (2006) Phage display systems and their applications. *Applied Microbiology and Biotechnology* **70**, 2–11.
6. Westerlund-Wikstrom, B. (2000) Peptide display on bacterial flagella: principles and applications. *International Journal of Medical Microbiology* **290**, 223–230.
7. Yan, X. and Xu, Z. (2006) Ribosome-display technology: applications for directed evolution of functional proteins. *Drug Discovery Today* **11**, 911–916.
8. Reineke, U., Schneider-Mergener, J. and Schutkowski, M. (2006) Peptide arrays in proteomics and drug discovery, in *BioMEMS and Biomedical Nanotechnology, Vol. II: Micro/Nano technologies for genomics*

and proteomics (Ferrari, M., Ozkan, M. and Heller, M. I., eds.), Springer US, New York/Boston, USA, pp. 161–282.

9. Zander, N., Beutling, U., Dikmans, A., Thiele, S. and Frank, R. (2005) A special cellulose membrane support for the combinatorial and parallel synthesis of peptide libraries suitable for the SC^2-type manufacturing of high density multi-purpose chemical micro-arrays, in *Peptides 2004: Proceedings of the 3rd International and 28th European Peptide Symposium* (Flegel, M., Fridkin, M., Gilon, C. and Slaninova, J., eds.), Kenes International, Geneva, Switzerland, pp. 405–406.
10. Dikmans, A., Beutling, U., Schmeisser, E., Thiele, S. and Frank, R. (2006) SC^2: A novel process for manufacturing multipurpose high-density chemical microarrays. *QSAR & Combinatorial Science* **25**, 1069–1080.
11. Frank, R., Güler, S., Krause, S. and Lindenmaier, W. (1991) Facile and rapid "spot-synthesis" of large numbers of peptides on membrane sheets, in *Peptides 1990: Proceedings of the 21st European Peptide Symposium* (Giralt, E. and Andreu, D., eds.), ESCOM, Leiden, NL, pp. 151–152.
12. Hilpert, K., Winkler, D. F. H. and Hancock, R. E. W. (2007) Cellulose-bound peptide arrays: Chemistry and Applications. *Biotechnology and Genetic Engineering Reviews* **24**, 31–106.
13. Kramer, A. and Schneider-Mergener, J. (1998) Synthesis and screening of peptide libraries on continuous cellulose membrane supports. *Methods in Molecular Biology* **87**, 25–39.
14. Kramer, A., Reineke, U., Dong, L., Hoffmann, B., Hoffmüller, U., Winkler, D., Volkmer-Engert, R. and Schneider-Mergener, J. (1999) Spot synthesis: observations and optimizations. *Journal of Peptide Research* **54**, 319–327.
15. Weiser, A.A., Or-Guil, M., Tapia, V., Leichsenring, A., Schuchardt, J., Frömmel, C. and Volkmer-Engert, R. (2005) SPOT synthesis: Reliability of array-based measurement of peptide binding affinity. *Analytical Biochemistry* **342**, 300–311.
16. Molina, F., Laune, D., Gougat, C., Pau, B. and Granier, C. (1996) Improved performances of spot multiple peptide synthesis. *Peptide Research* **9**, 151–155.
17. Toepert, F., Knaute, T., Guffler, S., Pires, J. R., Matzdorf, T., Oschkinat, H. and Schneider-Mergener, J. (2003) Combining SPOT synthesis and native peptide ligation to create large arrays of WW protein domains. *Angewandte Chemie International Edition* **42**, 1136–1140.
18. Toepert, F., Pires, J.R., Landgraf, C., Oschkinat, H. and Schneider-Mergener, J. (2001) Synthesis of an array comprising 837 variants of the hYAP WW protein domain. *Angewandte Chemie International Edition* **40**, 897–900.
19. Heine, N., Ast, T., Schneider-Mergener, J., Reineke, U., Germeroth, L. and Wenschuh, H. (2003) Synthesis and screening of peptoid arrays on cellulose membranes. *Tetrahedron* **59**, 9919–9930.
20. Blackwell, H. E. (2006) Hitting the SPOT: small-molecule macroarrays advance combinatorial synthesis. *Current Opinion in Chemistry and Biology* **10**, 203–212.
21. Winkler, D. F. H. and McGeer, P. L. (2008) Protein labeling and biotinylation of peptides during spot synthesis using biotin p-nitrophenyl ester. *Proteomics* **8**, 961–967.
22. Kim, D.-H., Shin, D.-S. and Lee, Y.-S. (2007) Spot arrays on modified glass surfaces for efficient SPOT synthesis and on-chip bioassay of peptides. *Journal of Peptide Science* **13**, 625–633.
23. Licha, K., Bhargava, S., Rheinländer, C., Becker, A., Schneider-Mergener, J. and Volkmer-Engert, R. (2000) Highly parallel nano-synthesis of cleavable peptide–dye conjugates on cellulose membranes. *Tetrahedron Letters* **41**, 1711–1715.
24. Reineke, U., Bhargava, S., Schutkowski, M., Landgraf, C., Germeroth, L., Fischer,G. and Schneider-Mergener, J. (1999) Spatial addressable fluorescence-quenched peptide libraries for the identification and characterization of protease substrates, in *Peptides 1998: Proceedings of the 25th European Peptide Symposium* (Bajusz, S. and Hudecz, F., eds.) Akademiai Kiado, Budapest, Hungary, pp. 562–563.
25. Andresen, H., Grötzinger, C., Zarse, K., Kreuzer, O. J., Ehrentreich-Förster, E. and Bier, F. F. (2006) Functional peptide microarrays for specific and sensitive antibody diagnostics. *Proteomics* **6**, 1376–1384.
26. Maier, S., Frank, M., Rau, H., Lewandrowski, P., Uhrig, R., Keil, O., Deppe, H., Müller, N., Vanier, C., Mannsperger, H., Zepter, S. and Junker, H.-D. (2006) Synthesis and quality control of thiol compound libraries for chemical microarrays. *QSAR & Combinatorial Science* **25**, 1047–1054.
27. Hahn, M., Winkler, D., Welfle, K., Misselwitz, R., Welfle, H., Wessner, H., Zahn, G., Scholz, C., Seifert, M., Harkins, R., Schneider-Mergener, J. and Höhne, W. (2001) Cross-reactive binding of cyclic peptides to an anti-TGFα

antibody Fab fragment. An X-ray structural and thermodynamic analysis. *The Journal of Molecular Biology* **314**, 293–309.

28. Hilpert, K., Hansen, G., Wessner, H., Volkmer-Engert, R. and Höhne, W. (2005) Complete substitutional analysis of a sunflower trypsin inhibitor with different serine proteases. *Journal of Biochemistry (Tokyo)* **138**, 383–390.
29. Beutling, U., Dikmans, A., Thiele, S. and Frank, R. (2005) A novel process for manufacturing high density multi-purpose chemical microarrays, in *Peptides 2004: Proceedings of the 3rd International and 28th European Peptide Symposium* (Flegel, M., Fridkin, M., Gilon, C. and Slaninova, J., eds.), Kenes International, Geneva, Switzerland, pp. 152–153.
30. Hilpert, K., Winkler, D. F. H., Hancock, R. E. W. (2007) Peptide arrays on cellulose support: SPOT synthesis – a time and cost efficient method for synthesis of large numbers of peptides in a parallel and addressable fashion. *Nature Protocols* **2**, 1333–1349.
31. Fields, G. B. and Noble, R. L. (1990) Solid phase synthesis utilizing 9-fluorenylmethoxycarbonyl amino acids. *International Journal of Peptide and Protein Research* **35**, 161–214.
32. Zander, N. and Gausepohl, H. (2002) Chemistry of Fmoc peptide synthesis on membranes, in *Peptide Arrays on Membrane Support*. (Koch, J. and Mahler, M., eds.) Springer-Verlag, Berlin Heidelberg, Germany, pp. 23–39.
33. Atherton, E. and Sheppard, R.C. (eds.) (1989) 7.2. Activated esters of Fmoc-amino acids, in *Solid phase peptide synthesis – a practical approach*. IRL press at Oxford University Press, Oxford, UK, pp. 76–78.
34. Gausepohl, H. and Behn, C. (2002) Automated synthesis of solid-phase bound peptides, in *Peptide Arrays on Membrane Support* (Koch, J. and Mahler, M., eds.), Springer-Verlag, Berlin Heidelberg, Germany, pp. 55–68.
35. Ast, T., Heine, N., Germeroth, L., Schneider-Mergener, J. and Wenschuh, H. (1999) Efficient assembly of peptomers on continuous surfaces. *Tetrahedron Letters* **40**, 4317–4318.
36. Weiler, J., Gausepohl, H., Hauser, N., Jensen, O.N. and Hoheisel, J. D. (1997) Hybridisation based DNA screening on peptide nucleic acid (PNA) oligomer arrays. *Nucleic Acids Research* **25**, 2792–2799.
37. Krchnak, V., Wehland, J., Plessmann, U., Dodemont, H., Gerke, V. and Weber, W. (1988) Noninvasive continuous monitoring of solid phase peptide synthesis by acid–base indicator. *Collection of Czechoslovak Chemical Communications* **53**, 2542–2548.
38. Lizcano, J. M., Deak, M., Morrice, N., Kieloch, A., Hastie, C.J., Dong, L., Schutkowski, M., Reimer, U. and Alessi, D. R. (2002) Molecular basis for the substrate specificity of NIMA-related kinase-6 (NEK-6). *The Journal of Biological Chemistry* **277**, 27839–27849.
39. Bhargava, S., Licha, K., Knaute, T., Ebert, B., Becker, A., Grötzinger, C., Hessenius, C., Wiedemann, B., Schneider-Mergener, J. and Volkmer-Engert, R. (2002) A complete substitutional analysis of VIP for better tumor imaging properties. *Journal of Molecular Recognition* **15**, 145–153.
40. Kamradt, T. and Volkmer-Engert, R. (2004) Cross-reactivity of T lymphocyctes in infection and autoimmunity. *Molecular Diversity* **8**, 271–280.
41. Boisguerin, P., Leben, R., Ay, B., Radziwill, G., Moelling, K., Dong, L. and Volkmer-Engert, R. (2004) An improved method for the synthesis of cellulose membrane-bound peptides with free C termini is useful for the PDZ domain binding studies. *Chemistry and Biology* **11**, 449–459.
42. Volkmer-Engert, R., Hoffmann, B. and Schneider-Mergener, J. (1997) Stable attachment of the HMB-linker to continuous cellulose membranes for parallel solid phase spot synthesis. *Tetrahedron Letters* **38**, 1029–1032.
43. Scharn, D., Wenschuh, H., Reineke, U., Schneider-Mergener, J. and Germeroth, L. (2000) Spatially addressed synthesis of amino- and amino-oxy-substituted 1,3,5-triazine arrays on polymeric membranes. *Journal of Combinatorial Chemistry* **2**, 361–369.
44. Frank, R., Hoffmann, S., Overwin, H., Behn, C. and Gausepohl, H. (1996) Easy preparation of synthetic peptide repertoires for immunological studies utilizing SPOT synthesis, in: *Peptides in Immunology* (Schneider, C. H., ed.), John Wiley & Sons, Ltd., New York, USA, pp. 197–204.
45. Koch, J. and Mahler, M. (eds.) (2002) *Peptide Arrays on Membrane Support*, Springer-Verlag, Berlin Heidelberg, Germany.

Chapter 6

Rapid Identification of Linear Protein Domain Binding Motifs Using Peptide SPOT Arrays

Douglas J. Briant, James M. Murphy, Genie C. Leung, and Frank Sicheri

Abstract

Understanding protein–protein interactions is a key step in unravelling the roles proteins play in cellular function. The ability to analyse protein–protein interactions rapidly and economically is a powerful research tool. Using peptide SPOT arrays, peptides of known sequence can be synthesized directly in discrete spots on a cellulose membrane and assayed for an interaction with a protein of interest. Several hundred peptides can be synthesized on each cellulose membrane; therefore, this method is amenable to designing high-throughput peptide binding studies. SPOT arrays are particularly well suited for deducing peptidic binding motifs within proteins that are difficult to purify in sufficient quantities for traditional biochemical analyses, as well as for determining binding specificities and targets for proteins of undefined function. Peptide SPOT arrays have been used extensively to define protein–protein interaction surfaces. In this chapter, we will outline the steps involved in designing and probing a peptide SPOT array to identify peptide binding motifs for a protein of interest.

Key words: Protein–protein interaction, peptide synthesis, peptide SPOT array, identifying protein targets.

1. Introduction

Protein–protein interactions underpin all functional aspects of cellular biology, including signal transduction, cell adhesion, growth factor recognition, enzyme–substrate interactions and antibody–antigen interactions. Typically, these interactions are mediated by modular protein domains, which have the capacity to recognize short peptide motifs with great specificity (1, 2).

The identification of linear peptide motifs that mediate protein–protein interactions has historically been a time-consuming and laborious task. To address this, high-throughput technologies,

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_6,

such as synthetic peptide arrays, have been utilized to allow rapid identification of linear protein binding motifs within the amino acid sequence of an interacting protein. This is especially useful for identifying targets of modular domains of unknown function or for detection of interaction sites on proteins that are difficult to purify in sufficient quantities for other binding studies. Peptide arrays hold many advantages over other methodologies, such as phage display, for several key reasons: (a) the capacity to methodically vary the amino acid at a prescribed position to deduce binding specificities/consensus motifs; (b) the potential to conveniently incorporate unnatural and modified amino acids, such as phosphotyrosine or acetyl-lysine, into a peptide; (c) the peptides are synthesized in high density on a cellulose membrane using standard 9-fluoroenyl-methoxycarbonyl (Fmoc) chemistry (3), and their preparation is rapid, convenient, automated and cost-effective. Peptide arrays synthesized on cellulose membranes are very versatile, and have been used for the characterization of a broad range of protein–protein interactions to date, including mapping antibody recognition epitopes (4), mapping the epitopes recognized by interacting proteins (5), defining the residues that are critical for recognition by protein domains (5–7) and the characterization of enzyme–substrate recognition determinants (*see* Leung et al., Chapter 7, this issue (8)). The integrity of the identified linear peptide motifs as protein binding motifs has been validated in vivo and in vitro (5–7), and suggests that this methodology is not only convenient and rapid, but also robust and reliable for deducing physiologically relevant protein–protein interactions.

Herein, we describe the methodology for the design and synthesis of peptide SPOT arrays and the application of these arrays for rapid identification of linear protein binding motifs.

2. Materials

2.1. SPOT Blot Synthesis

1. The materials utilized in this technique are hazardous and should be handled in a fume hood.
2. Peptide synthesis grade 1-methyl-2-pyrrolidone (NMP) dried by adding molecular sieves.
3. Derivatized cellulose membrane (Intavis AG, catalog number 32.100).
4. 0.75 M Hydroxybenzotriazole (HOBt) dissolved in NMP.
5. Fmoc amino acids.
6. Peptide synthesis grade dimethylformamide (DMF).

7. 20% v/v Piperidine (redistilled, 99.5% grade) in DMF for deprotection of Fmoc amino acids.
8. 2% v/v Acetic anhydride (99% grade) in DMF to serve as a capping reagent for amino acids.
9. Diisopropyl carbodiimide (DIC; peptide synthesis grade) in NMP for activation of amino acids.
10. Dichloromethane (DCM; HPLC grade) as a solvent for amino acid side chain deprotection.
11. 3% v/v Triisopropylsilane (99% grade), 2% v/v dH_2O, 47.5% v/v trifluoroacetic acid (TFA; HPLC grade) in DCM for deprotection of amino acid side chains.
12. 0.01% w/v Aqueous bromophenol blue for visualizing spots.
13. 95% EtOH.

2.2. SPOT Blot Analysis

1. Tris-buffered saline with 0.1% Tween 20 (TBST): 20 mM Tris-HCl, pH 7.5, 135 mM NaCl, 0.1% w/v Tween 20.
2. Blocking agent: TBST + 10% w/v milk powder or 3% bovine serum albumin (BSA) (*see* **Note 1**).
3. Primary and secondary antibody in TBST with 5% w/v milk powder or 3% BSA.
4. Chemiluminescent substrate, X-ray film and film cassette.
5. Stripping buffer A: 8 M urea, 1% w/v sodium dodecyl sulphate (SDS), 0.5% β-mercaptoethanol.
6. Stripping buffer B: 50% v/v ethanol, 10% v/v acetic acid (*see* **Note 2**).

3. Methods

SPOT blot techniques are a powerful tool for rapidly identifying peptide binding motifs. This method is particularly useful for identifying binding sites on proteins that are difficult to purify or that are not amenable to other methods such as proteolytic digests or expression of stable domains, or for elucidating binding targets of proteins or protein domains of unknown function. It is important, however, to ensure that peptides identified are specific for binding the protein of interest, and not artefacts arising from non-specific interactions between the ligand and peptide spot (e.g. charge–charge interactions). As a control, duplicate spot arrays are constructed and one is incubated with the protein of interest while the other is probed with an equivalent amount of an unrelated protein. Interactions that are specific for the protein of

interest should be verified by an independent method such as site-directed mutagenesis of the predicted binding site in the full-length protein followed by in vitro and/or in vivo tests to assess binding, if possible.

3.1. Designing the Peptide Array

1. SPOT arrays are carefully designed to maximize yield and simplify use. Generally, several experiments are included in a single synthesis as the amount of consumables used is the same regardless of how many peptide spots are synthesized. As a general rule, we try not to divide an experiment over two blots, and the synthesis is designed to avoid this.
2. For initial binding experiments, a "peptide walk" is performed. Generally, 12 amino acid peptides offset by two amino acids spanning the region of interest is an effective approach (*see* **Note 3** and **Fig. 6.1**).
3. To confirm peptide–protein binding interactions from a peptide walk, an alanine scan is performed (*see* **Fig. 6.1**). Peptides of interest from the peptide walk are identified as spots that interact strongly with the protein of interest, but not with the negative control protein. If the structure of the protein is known, this peptide should be mapped to see if the peptide is indeed a good candidate for binding (*see* **Note 4**). Blots are then constructed that comprise the entire peptide with each amino acid position replaced, in turn, with alanine. Positions that are required for binding the ligand protein will result in a loss of protein binding when they are substituted with alanine (*see* **Fig. 6.1**).
4. Once the peptide that binds to the ligand protein is identified, it may be desirable to further refine and/or optimize the binding sequence. For this, a total amino acid scan is performed where each amino acid in the peptide is substituted in turn with every amino acid (*see* Figure 2, Leung et al., Chapter 7, this issue). The optimized peptide binding sequence can subsequently be used to mine for other potential binding targets of the protein of interest through protein sequence database searches.

3.2. SPOT Blot Synthesis

1. The following instructions refer to an Intavis MultiPEP SPOT synthesizer, but may be adapted for other systems. The method describes the synthesis of two blots, each 30 spots by 20 spots (600 each, 1200 total spots). Wash volumes should be adjusted for different sized membranes.
2. The synthesis sequence is programmed into the SPOT synthesizer (*see* **Note 5**). A typical synthesis requires 3–4 days for completion.
3. 0.5 mmol Fmoc amino acids are weighed into 2 ml tubes, or alternately, preweighed amino acids can be purchased (Intavis, catalog number 30.700). Sufficient amino acid for the

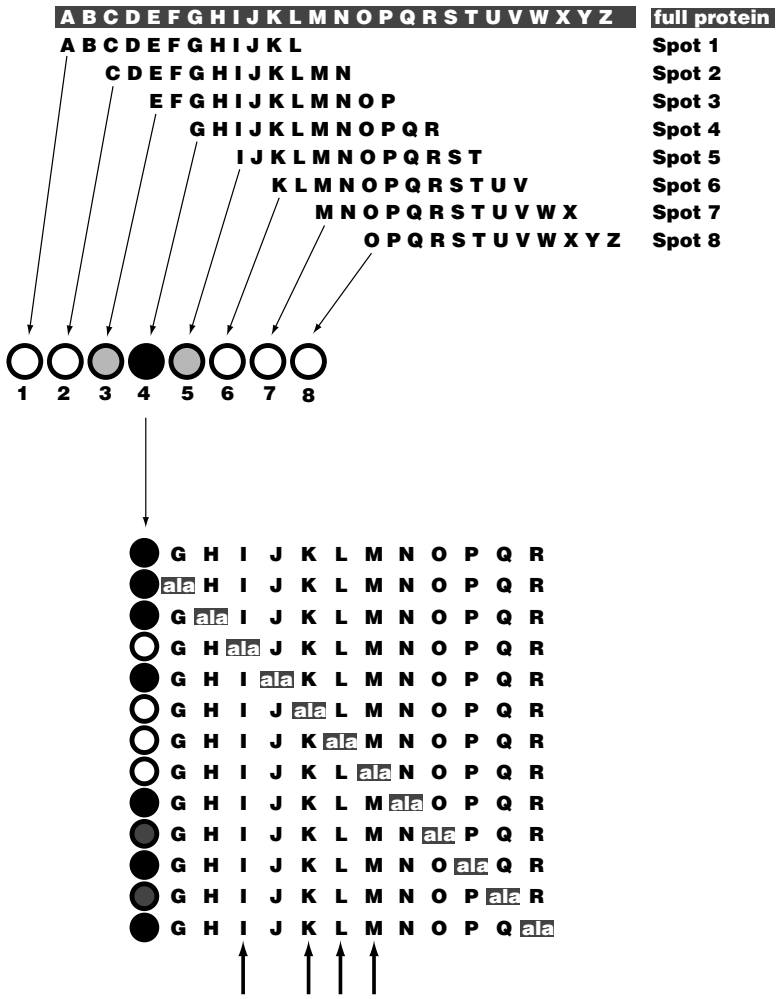


Fig. 6.1. Designing SPOT arrays for the identification of peptide binding motifs. The *top* of the panel depicts a full protein or protein binding domain of interest, with the primary amino acid sequence sequentially represented as letters a through z. Overlapping 12 amino acid peptides, offset by 2 amino acids (Spot 1–8) are synthesized on the membrane and probed with the protein of interest. In this example, Spot 4 represents a positive hit, determined by the procedure outlined in **Section 3.3**. To identify residues required for binding, each amino acid in the Spot 4 peptide is replaced with an alanine (ala) residue. In this example, residue J at position 3, K at position 5, L at position 6 and M at position 7 (all indicated with vertical arrows) are required for binding the protein of interest.

entire synthesis needs to be prepared. This may require splitting the reaction into multiple 2 ml tubes if large amounts of an amino acid are required (*see* **Note 6**).

4. 1 ml fresh 0.75 M HOBt solution is added to each amino acid, and tubes are vortexed or shaken for 30–60 min at room temperature. Once the amino acids are dissolved, a further 400 µl of NMP is added to each tube and mixed thoroughly.
5. While the amino acids dissolve, the membranes for synthesis are prepared (this section, steps 6–10).

6. Wash the membranes thoroughly with dH_2O (approximately 50–100 ml).
7. Each membrane should be aligned on its corresponding platform on the SPOT synthesizer. It is important to ensure that each blot is centred and that there are no wrinkles preventing contact between the membrane and rubber sheet atop the platform. The metal frame that holds the membranes is then carefully placed on the blots and secured in place by tightening the screws (*see* **Note 7**).
8. The vacuum waste bottle is emptied, with care given to properly dispose of the waste solvents.
9. The vacuum is turned on and the blots are each washed three times with copious amounts of 95% EtOH (10–20 ml each from a squeeze bottle).
10. The blots are dried under vacuum for at least 15 min during which time solutions for synthesis are prepared.
11. Solution bottles for EtOH and dimethylformamide (DMF) are filled. During the course of synthesis, approximately 4 l each of EtOH and DMF will be consumed.
12. A 20% v/v solution of piperidine in DMF is made on the first day of synthesis. An adequate volume for the entire synthesis should be made.
13. Capping reagent is made on the first day of synthesis, and consists of 2% v/v acetic anyhdride in DMF. Sufficient quantity should be made for the entire synthesis process.
14. Fresh aliquots of activated amino acids should be prepared just prior to starting each day of the SPOT synthesis. The activator solution is 9.5% diisopropyl carbodiimide (DIC) in NMP that has been dried on molecular sieves. Activator is then added in a 3:2 ratio with dissolved amino acids and mixed thoroughly, ensuring that there is sufficient volume of amino acids for all the cycles of the round of synthesis for a single day (usually four or five cycles per day). It is imperative to ensure that the solution is completely mixed by repeated pipetting. Activated amino acids are then placed in their corresponding position in the synthesizer. Double check that all tubes and synthesis solutions are properly positioned and pushed tightly into place to prevent contact with the synthesizer needle as this may result in damage.
15. The synthesis is initiated and allowed to proceed overnight. The reaction should be checked during the first cycle to ensure that the spotting is proceeding properly.
16. The following morning, the remaining activated amino acids are disposed according to waste solvent guidelines (*see* **Note 8**), and the EtOH and DMF solvent bottles are

topped up. New batches of amino acids are activated, and the synthesis is resumed where the previous synthesis terminated. This is repeated until the complete peptides are synthesized (usually 3 days for a 12- or 13-mer).

17. Following the final synthesis, the membranes are removed and placed in a single solvent-resistant dish that is minimally larger than the membrane. Before removing the membranes from the synthesizer, ensure they are marked with pencil in the corner to (a) distinguish between the two blots and (b) mark the orientation of the blots.
18. The blots are washed three times for 5 min with 20 ml dH_2O on an orbital shaker.
19. The membranes are mixed with an aqueous solution of 0.01% w/v bromophenol blue (50 ml) on an orbital shaker at room temperature for approximately 1 h.
20. Membranes are washed with excess dH_2O until most of the background staining is removed from the membrane. The peptide spots should be visible on the blot (*see* **Note 9**).
21. The membranes are washed with EtOH and air-dried on the bench top.
22. Once the membranes are completely dry, a grid is marked in pencil on the membrane to surround the location of each peptide spot. This is very important because the bromophenol blue staining will be removed by the subsequent deprotection steps (*see* **Note 10**).
23. At this point, the spots are still visible within the grid, and the membrane should be either photocopied or scanned for reference.
24. The blot is now ready for deprotection. These steps should be carried out at room temperature on an orbital shaker that is placed in a fume hood. A solvent-resistant container close to the size of the blot with a sealing lid should be used for deprotection, and two blots can be placed in a single container.
25. *Very important note*: TFA waste must go in a separate container as it reacts violently with DMF.
26. First, 30 ml of a 1:1 mix of TFA and dichloromethane (DCM) is prepared. To this TFA:DCM solution, 3% v/v triisopropyl silane and 2% v/v dH_2O are added (*see* **Note 11**) and this mixture is incubated with the blots on an orbital shaker for 60–90 min.
27. The blots are subjected to four washes, each 2 min, with 20 ml DCM. These washes are discarded in the TFA waste.
28. At this point, the TFA waste should be sealed to prevent accidental addition of DMF wash waste.
29. Blots are washed four times for 2 min with 20 ml DMF.

30. Two washes, each 2 min with 20 ml 95% EtOH, are performed and the blots are now ready for either analysis or storage. For storage, the blots should be air-dried on the bench top, heat-sealed in a polyester bag and stored at –20°C (*see* **Note 12**).

3.3. Analysis of Peptide Binding on SPOT Blot

1. Pre-wet the membrane by rinsing with EtOH, followed by thorough washing with TBST (*see* **Note 1**). This is performed at room temperature in a dish that is minimally larger than the size of the membrane.
2. The membrane is blocked overnight at 4°C on a rocker with an appropriate blocking reagent in TBST (typically, for a 20 by 30 membrane, about 10–15 ml will be required).
3. Blots are washed for 5 min, three times in TBST.
4. Binding assays are performed in TBST. Assays are performed in a minimum volume in heat-sealed polyester bags on a rocking platform at 4°C. The time of incubation and concentration of the binding protein are dependent upon binding affinity between the protein of interest and its corresponding peptide spot (*see* **Note 13**).
5. The membrane is washed three times, 5 min each with TBST on a rocking platform.
6. The blot is incubated in heat-sealed polyester bags with the primary antibody in TBST with blocking agent for 1 h at 4°C (for a 20 by 30 blot, approximately 5–10 ml of antibody solution will be required).
7. Three, 5 min washes in TBST are performed in a dish minimally larger than the membrane.
8. The membrane is mixed with HRP-conjugated secondary antibody (for a 20 by 30 blot, approximately 5–10 ml of antibody solution will be required) at the manufacturer's recommended dilution in TBST in a heat-sealed polyester bag at 4°C for 30–60 min.
9. Three washes, each 5 min with TBST are performed on the blot at 4°C.
10. The blot is now ready to be developed. To increase sensitivity, SuperSignal West Femto Maximum Sensitivity substrate (Pierce) is used. The peroxide buffer and luminol/enhancer solutions are mixed in a dark room and added immediately to the blot which is then sandwiched between acetate sheets. Add a luminescent marker, such as a "glow-in-the-dark" sticker to a corner of the blot so the orientation and location of the spots on the membrane can be deciphered. The blot is then exposed to X-ray film for several exposure lengths which may take up to 30 min (*see* **Note 14**).

3.4. Stripping the SPOT Blot

1. To reprobe the blots with other potential peptide spot-binding partners, the membranes must be stripped. All stripping steps are performed at room temperature in a dish as close as possible to the size of the membrane on a rocker.
2. The membrane is washed three times for 10 min with deionized water.
3. Three washes, each 10 min with about 20 ml stripping buffer A are performed (*see* **Note 15**).
4. Three further 10 min washes with 20 ml stripping buffer B are performed.
5. Finally, the membranes are washed three times with EtOH and bench-dried overnight, or, alternately, washed three times for 5 min with TBST and immediately reprobed.
6. Membranes can now be stored in heat-sealed polyester bags at –20°C.
7. For reprobing dried blots, the membranes must first be wetted with EtOH followed by three, 5 min TBST washes (*see* **Note 16**).

4. Notes

1. The correct choice of blocking agent can be critical for sensitivity of the antibodies. If the primary antibody is commercial, check the manufacturer's instructions, and if the antibody is not commercial, test Western blots can be performed to determine the conditions for maximum antibody sensitivity. This blocking agent should be used for all subsequent steps where a blocking agent is required.
2. Stripping buffer A without β-mercaptoethanol can be made in advance and stored at room temperature. β-mercaptoethanol should be added just prior to use. Stripping buffer B should be made fresh on the day of use.
3. The region of interest for the peptide walk should take any known binding information into account. If, for example, the ligand is known to bind to Domain A of the protein, then the initial peptide walk will encompass Domain A. This is especially important for large proteins as it may not be practical to walk their entire length.
4. Generally, a good candidate peptide for binding should be surface exposed on the structure.
5. This should be performed to maximize the number of peptides constructed on the membranes. Since the first and last spots of the synthesis tend to incur a larger degree of pipetting

error, we always lead and finish the synthesis with a row of peptides composed of a homogeneous amino acid. We also limit our daily synthesis cycles to four or five amino acids to ensure freshness of the activated amino acids.

6. For automated SPOT synthesizers such as the Intavis Multi-PEP SPOT, an amino acid usage calculator may be included in the operating software and this information can be used to ensure sufficient quantities of each amino acid are prepared.
7. Care should be taken to prevent scratching or denting the metal frame, as this will reduce the effectiveness of the seal. When removing the frame, we always set it on a soft surface, such as a foam computer mouse pad to protect the surface. When attaching the frame to the synthesizer, we tighten the screws in a cross-wise pattern to ensure an even seal.
8. Guidelines for disposal will vary, but care should be taken to properly dispose of all chemicals used during construction of the blots.
9. If the staining is very weak at this point, it may indicate a problem with the synthesis. Inclusion of a positive control on the blot may determine if the synthesis was unsuccessful.
10. Pencil is used to mark the blots, as pen will fade during the deprotection steps. We also mark the individual rows, to prevent any confusion in orientation or spot identity.
11. TFA is very volatile and corrosive. Proper protective equipment should be worn and all steps should be carried out in a fume hood. Due to its volatility, TFA is difficult to pipette without some spillage. To minimize spillage, the mouth of the TFA bottle and the deprotection mix tube should be in close proximity to minimize the amount of time TFA is in the pipette. It is also advisable to use a manual pipette bulb, since the corrosive, volatile TFA may damage a mechanical pipetting aid.
12. Blots stored at –20°C are stable for several months.
13. For binding assays, it is best to use protein that is purified as closely to homogeneity as possible. If specific antibodies against the protein of interest are available, we utilize these and remove any affinity tags from the protein as they may interfere with binding and/or give a false-positive signal. If the protein used for probing the blot is affinity-tagged, we use a sample of purified affinity tag, or an alternate affinity-tagged protein as our negative control. Care should be taken when using glutathione-*S*-transferase (GST) fusion proteins as a probe, since GST is a constitutive homodimer and avidity effects may influence binding to a peptide motif.

14. The SPOT blot can also be visualized on an imager, although we find that film gives a clearer signal, and that identification of the peptides by realigning with the membrane is simpler.
15. Washes with buffer A can be extended for up to 60 min each, if the interaction between the ligand protein and peptide spot is very strong.
16. If the binding interaction between a peptide spot and a ligand is very strong, the stripping procedure may not remove all of the ligand. To test the efficiency of stripping, the blot can be blocked (**Section 3.3**, step 2) probed with primary antibody and secondary antibody followed by developing the blot (**Section 3.2**, steps 6–10). If no signal shows up on the blot, the stripping was successful. The blot can then be washed four times for 5 min with TBST at room temperature and reprobed with another protein of interest (**Section 3.2**, steps 2–10).

Acknowledgments

This work was supported by grants from the National Cancer Institute of Canada and the Canadian Institutes of Health Research to F.S., and from the Terry Fox Foundation through an award from National Cancer Institute of Canada to GCL. JMM is a recipient of a CJ Martin (Biomedical) Fellowship from the National Health and Medical Research Council of Australia.

References

1. Pawson, T. (2007) Dynamic control of signaling by modular adaptor proteins *Curr Opin Cell Biol* **19**, 112–6
2. Kanduc, D., Serpico, R., Lucchese, A., and Shoenfeld, Y. (2008) Correlating low-similarity peptide sequences and HIV B-cell epitopes *Autoimmun Rev* **7**, 291–6
3. Frank, R. (1992) Spot-synthesis: An easy technique for the positionally addressable, parallel chemical synthesis on a membrane support *Tetrahedron* **48**, 9217–32
4. Reineke, U., Ivascu, C., Schlief, M., Landgraf, C., Gereicke, S., Zahn, G., Herzel, H., Volkmer-Enger, R., and Schneider-Mergener, J. (2002) Identification of distinct antibody epitopes and mimitopes from a peptide array of 5520 randomly generated sequences *J Immunol Methods* **267**, 37–51
5. Ogunjimi, A. A., Briant, D. J., Pece-Barbara, N., LeRoy, C., DiGuglielmo, G. M., Kavsak, P., Rasmussen, R. K., Seet, B. T., Sicheri, F., and Wrana, J. L. (2005) Regulation of Smurf2 ubiquitin ligase activity by anchoring the E2 to the HECT domain *Mol Cell* **19**, 297–308
6. Lim, C. S., Seet, B. T., Ingham, R. J., Gish, G., Matskova, L., Winberg, G., Ernberg, I., and Pawson, T. (2007) The K15 protein of Kaposi's sarcoma-associated herpesvirus recruits the endocytic regulator Intersectin 2 through a selective SH3 domain interaction. *Biochemistry* **46**, 9874–85
7. Smith, M. J., Hardy, W. R., Murphy, J. M., Jones, N., and Pawson, T. (2006) Screening for PTB domain binding partners and ligand specificity using proteome-derived NPXY peptide arrays *Mol Cell Biol* **26**, 8461–74
8. Leung, G. C., Murphy, J. M., Briant, D. J., and Sicheri, F. (2009) Characterization of kinase target phosphorylation consensus motifs using peptide SPOT arrays. *Methods Mol Biol*

Chapter 7

Characterization of Kinase Target Phosphorylation Consensus Motifs Using Peptide SPOT Arrays

Genie C. Leung, James M. Murphy, Doug Briant, and Frank Sicheri

Abstract

The human proteome is known to contain >500 protein kinases, which regulate almost all facets of cellular biology by the post-translational attachment of a phosphate moiety to serine, threonine, or tyrosine residues within a substrate protein. Most protein kinases remain poorly characterized and, as a result, current studies are directed toward defining their target substrates experimentally to gain a comprehensive view of the signaling proteins and pathways modulated by these kinases. Herein, we describe a rapid and convenient method for elucidating the consensus substrate motif for phosphorylation by a protein kinase using peptide SPOT arrays that are custom-synthesized on a cellulose membrane support. The definition of the target consensus motif provides an important starting point for the identification of physiologically relevant kinase substrates.

Key words: Kinase specificity, phosphorylation consensus motifs, peptide SPOT arrays.

1. Introduction

Eukaryotic protein kinases, numbering over 500 predicted kinases in the human proteome, play critical roles in multiple signaling networks to regulate essential processes including metabolism, cell growth, differentiation, motility, polarity, division, and death (1). Protein kinases share the general ability to catalyze the transfer of a phosphate moiety to serine, threonine, or tyrosine residues within a target sequence of a protein substrate. Protein kinases exhibit preferences for specific protein targets that, in effect, tailor them to specific signaling networks. These substrate preferences are partly dictated by a complementarity of charge, hydrophobicity, and shape between the kinase active site and the substrate peptide motif. In this way, the phospho-acceptor residue and its flanking

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_7,

residues serve as determinants for kinase recognition and phosphorylation (2–4). The characterization of the phosphorylation consensus motif of a kinase is key to understanding overall specificity of the protein kinase system, which is augmented by secondary recognition mechanisms, and also to be able to predict potential downstream targets.

The ability to directly synthesize peptides as SPOT arrays on a membrane has emerged as a powerful analytical tool to characterize kinase phosphorylation consensus motifs (5). Peptide SPOT arrays can be designed to vary the amino acid residues at each position flanking a phospho-acceptor residue. The array is subjected to phosphorylation by a kinase of interest in a single reaction mixture and the phosphorylation consensus motif determined by comparing the intensities of ^{32}P incorporation into different spots of known peptide sequence. This method has been used to characterize the phosphorylation consensus motifs of the NIMA-related kinase-6 (NEK6) and the polo-like kinase 4 (Plk4/Sak) (6, 7).

This chapter will describe a method to characterize the target phosphorylation consensus motifs for kinases using peptide SPOT arrays. While we will specifically describe this method using the murine serine/threonine kinase Plk4/Sak as a representative kinase, this protocol can be adapted to determine the target phosphorylation consensus of any protein kinase with detectable in vitro catalytic activity.

2. Materials

2.1. Peptide SPOT Array Synthesis

All materials required for peptide SPOT synthesis are listed in **Chapter 6** (p. 176) of this edition of Methods in Molecular Biology.

2.2. Kinase Assay

For one 13 × 20 peptide spot array membrane (*see* **Note 1**):

1. Peptide SPOT array membranes (for SPOT array synthesis, *see* **Chapter 6**, p. 178).
2. Tray (large enough to allow the entire surface of the membrane to be covered when rinsing).
3. 95% Ethanol.
4. Tris-buffered saline with Tween 20 (TBS-T): 20 mM Tris-HCl, pH 7.5, 135 mM NaCl, 0.1% w/v Tween 20 (*see* **Note 2**).
5. 50 mL Polypropylene tube with conical base.

6. Kinase reaction buffer: 20 mM Hepes, pH 7.5, 250 mM NaCl, 2 mM dithiothreitol (DTT), 20 mM $MgCl_2$, 20 mM $MnCl_2$ (*see* **Notes 3** and **4**).
7. Purified protein kinase of interest (here, we use Plk4/Sak) (*see* **Note 5**).
8. 10 mM (cold) unlabeled adenosine triphosphate (ATP) stock (Sigma Aldrich).
9. ^{32}P-γ-ATP (hot ATP), 10 μCi/μL (Perkin Elmer).
10. Temperature-controlled "rotisserie" hybridization oven, set to 30°C.
11. Stripping buffer A to remove non-covalently associated material after the phosphorylation reaction: 1% sodium dodecyl sulfate (SDS), 8 M urea, 0.5% β-mercaptoethanol (*see* **Note 6**).
12. Stripping buffer B: 50% ethanol, 10% glacial acetic acid (*see* **Note 7**).
13. Phosphor-screen and cassette.
14. Phosphor-imaging system and data analysis program (*see* **Note 8**).

3. Methods

To determine the consensus phosphorylation motif for a kinase using a peptide SPOT array, a peptide sequence that is detectably phosphorylated by the kinase of interest must initially be identified (**Section 3.2**). Once this initial peptide hit is determined, the preferences for specific residues at positions flanking the phospho-acceptor site can be determined by systematic replacement of these residues with each of the remaining 19 naturally occurring amino acids. By arraying the peptides onto a single membrane and incubating the membrane with active kinase in the presence of ^{32}P-γ-ATP, relative phosphorylation of spots can be easily detected and the preference for particular amino acid residues at each position flanking the phospho-acceptor site can be determined (the SPOT array phosphorylation assay will be described in **Sections 3.3** and **3.4**).

3.1. SPOT Array Synthesis

The protocol for peptide SPOT synthesis is shown in **Chapter 6** (p. 176) of this edition of Methods in Molecular Biology.

3.2. Design of the Peptide Array

1. If a peptide sequence that is detectably phosphorylated by the kinase of interest is known, proceed to step 5.

2. If a specific protein or protein fragment is known to be a substrate for the kinase of interest, each serine/threonine or tyrosine residue within the protein sequence is inferred to be a potential target site for phosphorylation (*see* **Fig. 7.1a**).
3. A series of 13 residue peptides (13 mers), each peptide sequence encompassing one of the potential phospho-acceptor residues (positioned in the seventh amino acid position) are synthesized as peptides in discrete spots on a membrane to determine which residue is preferentially phosphorylated by the kinase of interest. For each peptide, an adjacent peptide with the potential phospho-acceptor site altered to an alanine is synthesized (*see* **Fig. 7.1b** and **Note 9**).
4. This series of peptide spots is tested for phosphorylation by the kinase of interest (*see* **Sections 3.3** and **3.4**). ^{32}P-γ-ATP incorporation into a particular peptide spot, but not its

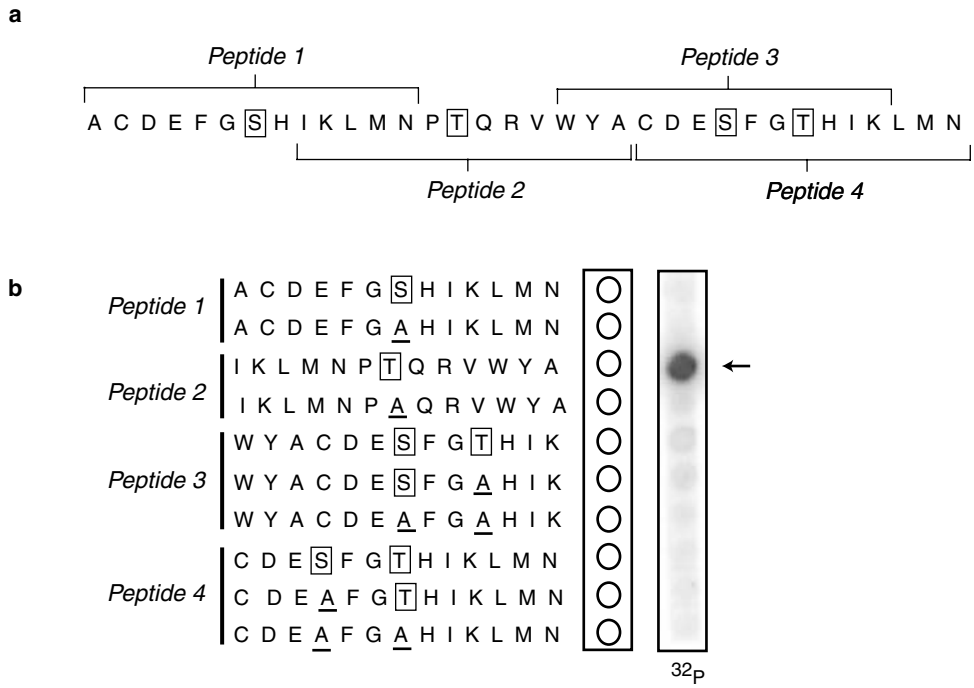


Fig. 7.1. Identification of a peptide phosphorylated by a kinase of interest. **a.** Sequence of a fictional protein fragment. All the potential phospho-acceptor residues for a Ser/Thr kinase are shown in boxes. Each 13mer peptide to be synthesized (framed in brackets) encompasses one phospho-acceptor site. **b**. Illustration of peptides 1–4 synthesized onto a SPOT membrane. The sequence of each peptide is shown on the left. Each peptide is synthesized twice: one peptide with the original sequence, and an additional peptide where the potential phospho-acceptor site is altered to an Ala (underlined). For peptides 3 and 4, which have more than one potential phospho-acceptor site, one of the sites is changed to an Ala (underlined). Following phosphorylation of the SPOT membrane by a kinase of interest, results (shown on the right) would indicate that the kinase of interest specifically phosphorylates peptide 2 (indicated by an arrow).

adjacent spot with alanine in place of the phospho-acceptor site, identifies a peptide sequence as a bona fide in vitro phosphorylation target of the kinase of interest (*see* **Fig. 7.1b**).

5. To determine the target phosphorylation consensus motif of the kinase of interest (using either a known peptide sequence or a peptide sequence identified in the previous step), a peptide SPOT array is synthesized by systematically altering each of the peptide residues to each of the 20 amino acids (*see* **Fig. 7.2a** and **Note 10**).
6. The peptide SPOT array is tested for phosphorylation by the kinase of interest (**Sections 3.3** and **3.4**).

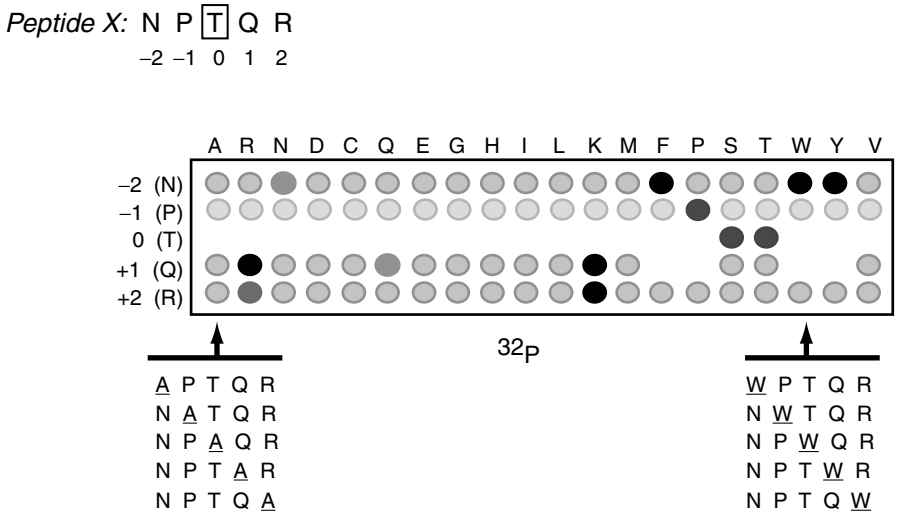


Fig. 7.2. Design of a peptide SPOT array. A 5 × 20 peptide SPOT array of a fictitous Peptide X with a 5 amino acid sequence as shown. The sequence of the peptide is labeled in brackets on the vertical axis. Numbers indicate the residue position relative to the phospho-acceptor site (labeled 0). Each residue in the vertical axis is altered to every other residue in the horizontal axis. The sequences to the bottom left are the sequences of peptides in the "A" column; the sequences to the bottom right are the sequences of peptides in the "W" column. Spots illustrate ^{32}P incorporation following phosphorylation of a peptide SPOT array by a kinase of interest. Preferred amino acids at each position can be determined by examining the relative intensities of the spots. Note that replacement of the phospho-acceptor residue (T) with a residue other than Ser or Thr results in loss of phosphorylation. Interpretation of the phosphorylation determinants would be as follows: In the –2 position, the residues Asn, Phe, Trp and Tyr are favoured, where Phe, Trp and Tyr are strongly preferred. In the –1 position, proline is favoured. In the +1 position, Arg, Gln, and Lys are favoured, where the positively charged Arg and Lys are strongly preferred, but Phe, Pro, Trp and Tyr are not tolerated. In the +2 position, Lys and Arg are favoured with a stronger preference for Lys.

3.3. Preparation of SPOT Membrane and Reaction Mixture

For one 13 × 20 spot peptide SPOT array (*see* **Note 1**):

1. The SPOT array membrane is placed in a plastic tray and rinsed briefly in 95% ethanol. When handling with filter forceps, the membrane is grasped along the edges taking care not to scrape the peptide-coated surface.

2. The membrane is washed two to three times with a generous amount of TBS-T, approximately 1 min per wash with gentle shaking.
3. 4 mL of kinase reaction buffer is mixed with purified kinase (*see* **Note 11**).
4. The membrane is transferred to a 50 mL polypropylene tube containing reaction mixture described in the previous step. The surface of the membrane on which the peptide spots were synthesised should face inward to allow for complete exposure to the reaction mixture as the tube rotates.

3.4. Kinase Assay

All radioactive work is done according to local guidelines.

1. A hot ATP mixture is prepared by mixing 5 μL of 10 mM unlabeled ATP and 100 μCi ^{32}P-γ-ATP in a 1.5 mL microcentrifuge tube.
2. Phosphorylation reactions are started by adding the hot ATP mixture to the reaction mixture in the 50 mL polypropylene tube (final concentration of unlabeled ATP = 12.5 μM, ^{32}P-γ-ATP: 25 μCi/mL, *see* **Note 12**).
3. The reaction is incubated in a temperature-controlled hybridization oven for 30 min at 30°C. The 50 mL polypropylene tube is positioned such that the entire surface of the membrane is exposed to the reaction mixture as the tube is rotated (*see* **Note 13**).
4. The reaction mixture is discarded into a radioactive liquid waste container and the membrane is washed four times with ~10 mL of stripping buffer A for approximately 5 min (*see* **Note 14**). Each wash is discarded into the radioactive liquid waste.
5. The membrane is washed three times with approximately 10 mL each of stripping buffer B for approximately 5 min.
6. The membrane is washed three times with approximately 10 mL each of 95% ethanol for approximately 2 min.
7. Filter forceps are used to carefully remove the membrane from the 50 mL polypropylene tube. The membrane is placed on laboratory tissues to dry for 10 min at room temperature in a fume hood.
9. The membrane is sealed in plastic wrap and exposed to a blanked phosphor-screen for 1–2 h. The phosphor-screen is scanned using an appropriate imaging system such as Storm (Molecular Dynamics). For examples of actual experimental results, see (6, 7) (*see* **Note 15**).

4. Notes

1. The dimensions of the 13 × 20 spots arise from systematically altering each residue of a 13 amino acid peptide with each of the 20 amino acids (*see* **Fig. 7.2a**).
2. All solutions should be made in filtered deionized water.
3. While this kinase reaction buffer has supported the catalytic activity of several kinases in our laboratory, it may be necessary to optimize these conditions for your kinase of interest. Primary parameters to optimize include pH, $MgCl_2$, and $MnCl_2$ concentrations. It may be helpful to add the phosphatase inhibitors, sodium orthovanadate or sodium fluoride, to the reaction should any contaminating phosphatases be present in the protein kinase sample. Kinase reaction buffer without DTT can be made in advance and stored at 4°C. Warm to room temperature and add fresh DTT before use.
4. It is foreseeable that some kinases may bind the cellulose membrane non-specifically. In such cases, the inclusion of a detergent, such as Tween 20 (0.02% v/v or higher) in the kinase reaction buffer may be warranted. The use of proteins such as casein (milk powder) or bovine serum albumin is not recommended as blocking reagents as these proteins can be phosphorylated causing non-specific background. An alternative is to synthesize the peptide array on a cellulose membrane with a cleavable linker, cleave the peptide, and covalently attach to a glass chip via aldehyde chemistry as described previously (6).
5. It is preferable to obtain soluble kinase protein that is untagged and purified to homogeneity. The Plk4 protein used for our assays was >90% pure.
6. Stripping buffer A can be made ahead of time and stored at room temperature with β-mercaptoethanol added immediately before use. β-mercaptoethanol should be handled with care. Wear protective clothing and gloves and avoid breathing fumes by opening the stock bottle in a fume hood and by keeping stripping buffer A capped, opening only when needed.
7. Stripping buffer B is made fresh on the day of use.
8. For phosphor-imaging, we use the Storm Imaging system with ImageQuant analysis software (Molecular Dynamics).
9. If more than one serine/threonine or tyrosine residue is present in a 13 amino acid peptide sequence, synthesize two versions of this peptide to determine which site is

being phosphorylated; one peptide with a sequence identical to the naturally occurring sequence found in the protein substrate and the other peptide with the second potential phospho-acceptor residue changed to an alanine (*see* **Fig. 7.1b**).

10. If more than one peptide is phosphorylated by the kinase (**Section 3.2**, step 3), peptide SPOT arrays can be synthesized for each peptide sequence. Multiple arrays would allow you to validate a target phosphorylation consensus motif and/or determine if your kinase phosphorylates several distinct target phosphorylation consensus motifs.
11. The Plk4 kinase domain protein described in this assay encompasses residues 1–390 of murine Plk4. We use 300 μg of purified Plk4 kinase protein for one 13 × 20 array, an amount that was predetermined using standard in vitro phosphorylation assays (7). The ratio of reaction volume to protein used in these phosphorylation assays was maintained in the phosphorylation of the peptide SPOT arrays.
12. It is important to maintain the ratio of unlabeled ATP: ^{32}P-γ-ATP when adjusting the reaction mixture volume.
13. After 30 min, the kinase reaction is within the linear range for phosphorylation by Plk4. A time course should be performed to determine the linear range of each particular kinase. Multiple peptide spots of identical sequence are synthesized on a single membrane which is then cut into segments, each with an equal number of peptide spots. Each segment is then incubated with kinase reaction mixture for a single time point within a time course (between 0 and 60 min) and quantitation of ^{32}P-γ-ATP incorporation at each time point would determine the linearity of the phosphorylation reaction.
14. The membrane should adhere to the side of the conical as the reaction mixture or individual washes are discarded into the waste. Each wash is done by hand, gently rocking, and simultaneously rotating the tube to ensure that the entire surface of the membrane is exposed.
15. Following identification of a consensus phosphorylation motif using peptide SPOT arrays, we confirm that the peptides are phosphorylated in solution using in vitro kinase assays with free peptides (in solution and not synthesized on a solid membrane support). Proteins in the reaction mixture are resolved by gel electrophoresis, and ^{32}P-γ-ATP incorporation is detected by phosphorimaging.

Acknowledgments

We thank C.S. Ho and I.M. Blasutig for helpful comments on the manuscript, and T. Pawson for use of his Multipep SPOT synthesizer. This work was supported by grants from the National Cancer Institute of Canada and the Canadian Institutes of Health Research to F.S., and from the Terry Fox Foundation through an award from National Cancer Institute of Canada to GCL. JMM is a recipient of a CJ Martin (Biomedical) Fellowship from the National Health and Medical Research Council of Australia.

References

1. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002) The protein kinase complement of the human genome. *Science*, **298**, 1912–34.
2. Kennelly, P. J., and Krebs, E. G. (1991) Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. *J Biol Chem*, **266**, 15555–8.
3. Songyang, Z., and Cantley, L. C. (1995) Recognition and specificity in protein tyrosine kinase-mediated signalling. *Trends Biochem Sci*, **20**, 470–5.
4. Songyang, Z., Lu, K. P., Kwon, Y. T., Tsai, L. H., Filhol, O., Cochet, C., Brickey, D. A., Soderling, T. R., Bartleson, C., Graves, D. J., DeMaggio, A. J., Hoekstra, M. F., Blenis, J., Hunter, T., and Cantley, L. C. (1996) A structural basis for substrate specificities of protein Ser/Thr kinases: primary sequence preference of casein kinases I and II, NIMA, phosphorylase kinase, calmodulin-dependent kinase II, CDK5, and Erk1. *Mol Cell Biol*, **16**, 6486–93.
5. Frank, R. (2002) The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports – principles and applications. *J Immunol Methods*, **267**, 13–26.
6. Lizcano, J. M., Deak, M., Morrice, N., Kieloch, A., Hastie, C. J., Dong, L., Schutkowski, M., Reimer, U., and Alessi, D. R. (2002) Molecular basis for the substrate specificity of NIMA-related kinase-6 (NEK6). Evidence that NEK6 does not phosphorylate the hydrophobic motif of ribosomal S6 protein kinase and serum- and glucocorticoid-induced protein kinase in vivo. *J Biol Chem*, **277**, 27839–49.
7. Leung, G. C., Ho, C. S., Blasutig, I. M., Murphy, J. M., and Sicheri, F. (2007) Determination of the Plk4/Sak consensus phosphorylation motif using peptide spots arrays. *FEBS Lett*, **581**, 77–83.

Chapter 8

CelluSpots™: A Reproducible Means of Making Peptide Arrays for the Determination of SH2 Domain Binding Specificity

Chenggang Wu and Shawn S.-C. Li

Abstract

Peptide arrays differ from conventional peptide synthesis in that hundreds upon thousands of peptides are synthesized and presented on a planar surface at a time. While direct synthesis of peptide arrays on a functionalized surface is feasible, reprinting of pre-made peptides offers flexibility and reproducibility and drastically reduces cost when multiple copies of the same or related peptide arrays are needed. CelluSpot™, a method developed by Intavis, opens a new route in peptide array synthesis and printing and overcomes certain limitations of the SPOT membrane. This technique was used to produce hundreds of phosphotyrosine-oriented peptide array libraries for determining the specificity of the Src homology 2 (SH2) domain.

Key words: Peptide array, peptide conjugate, CelluSpot, SPOT, SH2.

1. Introduction

CelluSpots™ are arrays of peptide–cellulose conjugates spotted on a planar surface. In this method, peptides are first synthesized on a modified cellulose support and are subsequently dissolved in a trifluoroacetic acid (TFA) cocktail after synthesis. Solutions of individual peptides covalently linked to the cellulose polymer are then spotted onto a surface of choice in a fully addressable manner. Upon evaporation of the solvent, peptide–cellulose spots are formed that are water insoluble and can withstand repeated washes required for the most stringent biochemical assays. The main

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_8,

advantages of CelluSpots™ arrays are as follows: (a) multiple identical copies of the same peptide array can be prepared by one batch of SPOT synthesis followed by a one-step printing (or re-spotting); (b) modification of neither the peptide nor the surface of the support is necessary prior to printing; (c) the cellulose matrix has a low capacity for non-specific protein binding, thus reducing background signals; and (d) an optimal peptide concentration can be obtained by serial dilutions before printing. These advantages made CelluSpots™ an efficient method for producing OPAL arrays for the determination of Src homology 2 (SH2) domains binding specificity as described below.

2. Materials

2.1. Synthesis of Peptide Arrays by SPOT

1. Auto-Spot ASP 222 Robot (Abimed, CA).
2. Derivatized cellulose membrane (Intavis AG).
3. Fmoc amino acids with orthogonal sidechain protection (NovoBiochem).
4. 1-Methyl-2-pyrrolidone (NMP) (OmniSolv).
5. Hydroxybenzotriazole (HOBT) (NovoBiochem).
6. Diisopropyl carbodiimide (DIC) (Sigma).
7. *N*,*N*-Dimethylformamide (DMF) (OmniSolv).
8. Acetic anhydride (Sigma).
9. Piperidine (Sigma).
10. Anhydrous ethyl alcohol (various suppliers).
11. Dichloromethane (DCM) (OmniSolv).
12. TFA cocktail: 94% trifluoroacetic Acid (TFA) (Sigma), 4% triisopropylsilane (TIPS) (Sigma), 2% deionized water.

2.2. Producing (pTyr) OPALs by the CelluSpots™ Technology

1. Auto-Spot ASP 222 Robot (Abimed, CA) or Multipep Robot (Intavis AG).
2. Membranes for Celluspots (Intavis AG).
3. Fmoc amino acids with orthogonal sidechain protection (NovoBiochem).
4. 1-Methyl-2-pyrrolidone (NMP) (OmniSolv).
5. Hydroxybenzotriazole (HOBT) (NovoBiochem).
6. Diisopropyl carbodiimide (DIC) (Sigma).
7. *N*,*N*-Dimethylformamide (DMF) (OmniSolv).
8. Acetic anhydride (Sigma).

9. Piperidine (Sigma).
10. Anhydrous ethyl alcohol (Commercial Alcohols).
11. TFA cocktail: 88% trifluoroacetic acid (TFA) (Sigma), 4% trifluoromethanesulphonic acid (TFMSA) (Fluka), 4% triisopropylsilane (TIPS) (Sigma), 4% deionized water.
12. *tert*-Butyl-methyl ether (Sigma).
13. Dimethyl sulphoxide (DMSO) (Sigma).
14. SSC buffer: 15 mM sodium citrate (pH 7.0) and 150 mM sodium chloride.
15. Brady label with polyester surface (Brady).

2.3. Screening of a pTyr OPAL by an SH2 Domain

1. TBS buffer: 0.1 M Tris-HCl (pH 7.4), 150 mM NaCl.
2. TBST buffer: 0.1 M Tris-HCl (pH 7.4), 150 mM NaCl, and 0.1% Tween 20.
3. Glutathione Sepharose 4 Fast Flow bead (GE Healthcare).
4. L-Glutathione reduced (Sigma).
5. Blocking buffer: 5% bovine serum albumin (BSA) (Roche) in TBST buffer.
6. Glutathione-horseradish peroxidase (GSH-HRP) (Santa Cruz Biotech).
7. Detection reagents: Western Lightning Chemiluminescence Reagent Plus (Perkin Elmer) prepared according to the manufacturer's protocol immediate prior to use.
8. Fluor-S MultiImager (BIORAD) or any other fluorescein imaging equipment.

2.4. Screening of a Targeted Peptide Array by an SH3 Domain

1. PBS buffer: 2.68 mM KCl, 1.47 mM KH_2PO_4, 136.89 mM NaCl, 8.1 mM Na_2HPO_4 (pH 7.4).
2. HIS-Select Nickel Affinity Gel (Sigma).
3. Imidazole (Sigma).
4. Fluorescein-5-maleimide (Pierce).
5. 2-Mercapto-ethanol (Sigma).
6. TBS buffer: 0.1 M Tris-HCl (pH 7.4), 150 mM NaCl.
7. TBST buffer: 0.1 M Tris-HCl (pH 7.4), 150 mM NaCl, and 0.1% Tween 20.
8. Blocking buffer: 5% bovine serum albumin (BSA) (Roche) in TBST buffer.
9. Fluor-S MultiImager (BIORAD) or another fluorescein imaging apparatus.

3. Methods

If unspecified, all the experiment steps are carried out at room temperature.

3.1. Synthesis of Peptide Arrays (e.g. SH3 Ligand Array) by SPOT

1. Search the SwissProt protein database for candidate peptide sequences that match the Class I (XXX[R/K]XΦPXXPXXX) or Class II (XXXXPXΦPX[R/K]XXXX) motifs, where Φ denotes a hydrophobic residue and X denotes any amino acid.
2. Synthesize the peptide array on the derivatized cellulose membrane with a capacity of 1536 by Auto-Spot ASP 222 Robot according to the manufacturer's protocol.

3.2. Producing pTyr OPALs by the CelluSpots™ Technology

1. Generate grids on the membranes for CelluSpots to help localize/address peptides on the array.
2. Dissolve individual Fmoc amino acid in organic solvent and prepare the mixture "X" by mixing the 19 Fmoc amino acid solutions (except Cys) in the following ratio (A:R:N:D:Q:E:G:H:I:L:K:M:F:P:S:T:W:Y:V = 3:6:5:5:5:5:3:7:6:3:5:3:4:5:5:5:6:4:6) (*see* **Note 1**).
3. Synthesize pTyr OPAL on the membrane for Celluspots with a capacity of 384 on an Auto-Spot ASP 222 Robot or Multipep Robot according to the manufacturer's manual while leaving the sidechain protected (*see* **Note 2**).
4. Punch each peptide spot and transfer the peptide-membrane disc to a numbered TFA-resistant tube.
5. Add 250 μl of the TFA cocktail to each tube and close the cap. Incubate overnight in a fume hood until the cellulose discs are completely dissolved.
6. Add 750 μl of ice-cold ether to each tube when the peptide–cellulose conjugates were completely dissolved and tighten the tube cap.
7. Rock the tube gently to generate the peptide–cellulose conjugate suspension, which should appear cloudy to the eye.
8. Spin down the peptide–cellulose conjugate suspension in an explosion-proof centrifuge at 3000 $\times g$ and at 0°C.
9. Remove the supernatant and wash the pellet with fresh ether three times.
10. Evaporate the remaining ether in a fume hood after the last ether wash.
11. Add DMSO to each well to dissolve the conjugates (*see* **Note 3**).
12. Close the tubes with caps and store these stock solutions in a 4°C freezer.

13. Dilute the peptide–cellulose conjugates from stock solutions in SSC buffer (*see* **Note 4**).
14. Re-spot the diluted peptide–cellulose conjugates on Brady labels (*see* **Note 5**).
15. Let the spots air-dry and place the labels in an oven set at 70°C for 1∼2 h.
16. Seal and store the peptide array labels in a 20°C freezer.

3.3. Screening of a pTyr OPAL Using an SH2 Domain

1. Purify the SH2-GST fusion protein using the Glutathione Sepharose 4 Fast Flow beads according to the manufacturer's protocol. Pass the elutant through a gel filtration column on a fast protein liquid chromatography (FPLC) system run at 4°C.
2. Block the OPAL membrane with blocking buffer for 1 h.
3. Label the SH2-GST protein with GSH-HRP for 30 min to form a stable SH2-GST-GSH-HRP conjugate in a ratio specified in the product manual.
4. Add the SH2-GST-GSH-HRP conjugate to the blocking buffer at a final concentration of 1 μg/ml.
5. Incubate the OPAL membrane with a solution of the SH2-GST-GSH-HRP conjugate for 1 h.
6. Wash the OPAL membrane three times with TBST buffer for 15 min each.
7. Incubate the OPAL membrane with the detection reagent for approximately 1 min.
8. Develop the film for a suitable time in a dark room.
9. Scan the film on a Fluor-S MultiImager.

3.4. Peptide Array Target Screening (PATS) Assay for SH3 Domains

1. Each SH3 domain was subcoloned in the pET28a vector to contain a C-terminal Gly-Gly-Cys triad sequence for attachment of fluorescein and an N-terminal $(His)_6$ tag for purification.
2. Purify the $(His)_6$-tagged SH3 protein by metal ion affinity chromatography using HIS-Select Nickel Affinity Gel according to the manufacturer's protocol followed by FPLC at 4°C in PBS buffer.
3. Label the purified SH3 protein with fluorescein-5-maleimide according to the manufacturer's protocol.
4. Block the SH3 target array membrane in the blocking buffer for 1 h.
5. Incubate the SH3 target array membrane with 1 μM fluorescein-labelled SH3 protein for 1 h.
6. Wash the membrane three times in TBST for 5 min each.
7. Scan the membrane using a fluorescein imager.
8. Identify positive binding peptides.

4. Notes

1. Since the coupling efficiency of each amino acid varies, it is important that the mixture take into account of this difference. However, the mixing ratios provided here should be used as a reference only. Depending on the particular amino acid derivatives used in the synthesis, the user should determine the ratios empirically.
2. Sidechain protection groups are removed in the TFA cocktail used to dissolve the spot disc.
3. The volume of DMSO should be adjusted according to the amount of peptides synthesized and to the conditions of the assay. We typically use 100 µl DMSO for each peptide spot synthesized in a density of 384 peptides per membrane.
4. The ratio of dilution should be adjusted in different applications. For the SH2-binding assay, we diluted the peptide stock in a 1:5 ratio. A higher ratio of the SSC buffer will accelerate drying of the spots while resulting in a lower concentration of the peptides.
5. Volume of peptide–cellulose conjugates used for re-spotting should be adjusted according to the size of the microarray and sensitivity of the assay.

References

The relevant references are listed in accompanying **Chapter 3.**

Chapter 9

High-Density Peptide Microarrays for Reliable Identification of Phosphorylation Sites and Upstream Kinases

Alexandra Thiele, Johannes Zerweck, Matthias Weiwad, Gunter Fischer, and Mike Schutkowski

Abstract

The human genome encodes about 25,000 genes. This number seems to be very small compared to the multitude of different protein functions in highly regulated pathways that are responsible for complex biochemical mechanisms like growth, metabolism, signal transduction and reproduction. Obviously, there are mechanisms creating additional protein diversity. The most important mechanism is post-translational modification (PTM) changing protein surfaces by phosphorylation, sulfation, acetylation, methylation and sumoylation resulting in an about 100-fold higher complexity (1, 2). This chapter presents a very efficient way to detect potential phosphorylation sites in proteins using overlapping peptide scans immobilized on glass slides.

Results from 35 different human kinases using peptide microarrays displaying overlapping peptide scans through either all human cyclophilins or all human FK506-binding proteins are shown. Additionally, detection of phosphorylation sites in a proteome-wide manner is demonstrated using peptide microarrays displaying cytomegalovirus proteome in the form of more than 17,000 overlapping peptides.

Key words: Peptide microarray, kinase, phosphorylation site, overlapping peptides, peptidyl-prolyl-*cis/trans*-isomerase, cyclophilin, FK506-binding protein, Pin1.

1. Introduction

There are more than 200 different post-translational modifications (PTMs) known, of which phosphorylation of serines, threonines and tyrosines is the most important one. There are estimations that about one-third of all human proteins is modified by kinases resulting in more than 100, 000 human phosphorylation sites (3).

The functions of phosphorylation are diverse and can result in either regulation of enzymatic activity by factor of 1000 (4, 5) or incorporated phosphate moiety can create binding sites for interacting proteins that have phospho-specific recognition domains,

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_9,

like 14-3-3, SH2 or subsets of WW-domains. It can influence localization of the modified protein and can even be involved in the regulation of protein degradation.

Efficient methods are needed enabling reliable detection of phosphorylation sites in a proteome-wide scale. Commonly used methods are based on the enrichment of phosphorylated peptides resulting from proteolytic digests of complex mixtures using, e.g. immobilized metal affinity chromatography (IMAC) or phosphoramidate chemistry (PAC) followed by the identification of the phosphopeptides using mass spectrometry (6). Low abundant proteins or short lifetimes of the phosphorylation events can be limiting factors for phosphorylation site determination of proteins using above-mentioned enrichment methods. Furthermore, results using phosphopeptide enrichment methods strongly depend on the used method (6).

Here we provide a robust protocol to determine potential phosphorylation sites of complete protein family of interest. We searched systematically for phosphorylation sites in human peptidyl-prolyl-*cis/trans*-isomerases (PPIases) in a proteome-wide manner as this modification is expected to be one mechanism to regulate PPIase enzymatic activity (*see* **Figs. 9.1** and **9.2**). The

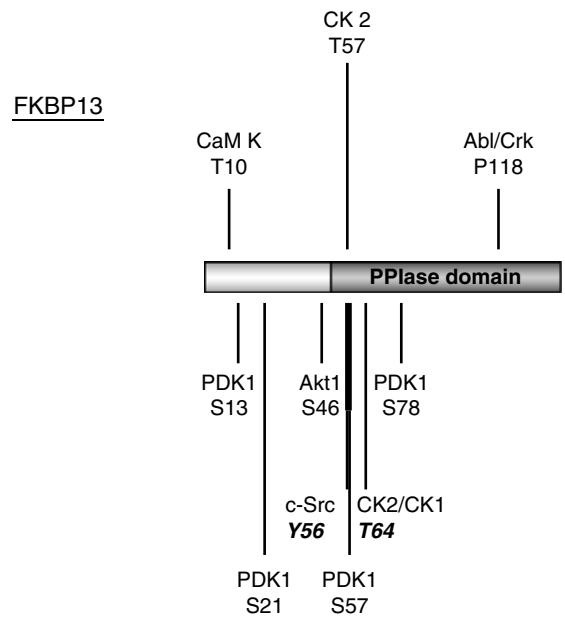


Fig. 9.1. Schematic presentation of human FKBP13 is shown together with predicted (*top*) and detected (*bottom*) phosphorylation sites. Prediction was made using www.scansite.mit.edu at medium stringency settings. Detected phosphorylation sites marked in italic represent residues found to be important for the active site of the PPIase (Abl = Abelson tyrosine kinase, Akt = protein kinase B, CaM K = calmodulin-dependent protein kinase, CK1 = casein kinase 1, CK2 = casein kinase 2, Crk = Cdc2-related kinase, c-Src = cellular homolog to Rous sarcoma tyrosine kinase, PDK1 = 3-phosphoinositide-dependent protein kinase 1).

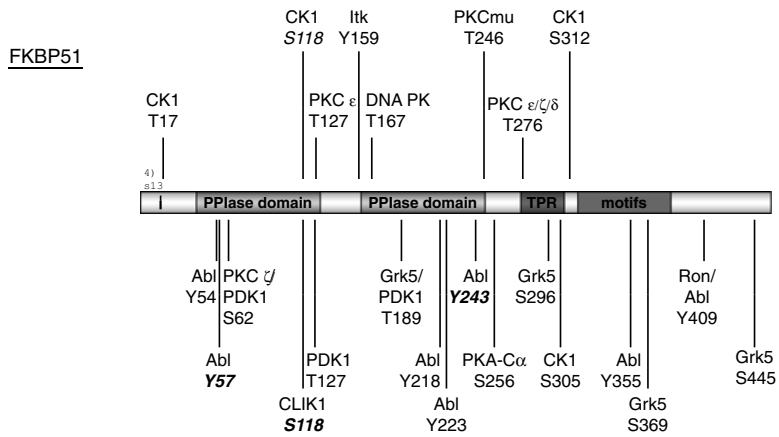


Fig. 9.2. Schematic presentation of human FKBP51 is shown together with predicted (*top*), published (*black bar* marked with a *star*) and detected (*bottom*) phosphorylation sites. Prediction was made using www.scansite.mit.edu at medium stringency settings. Detected phosphorylation sites marked in italic represent residues found to be important for the active site of the PPIase (Abl = Abelson tyrosine kinase, CaM K = calmodulin-dependent protein kinase, CK1 = casein kinase 1, CK2 = casein kinase 2, CLIK1 = CLP-36 interacting kinase 1, DNA PK = DNA dependent protein kinase, Grk5 = G protein-coupled receptor kinase 5, Itk = Interleukin-2 inducible T-cell kinase, PDK1 = 3-phosphoinositide-dependent protein kinase 1, PKA–Cα = protein kinase A isoform Cα, PKC = (protein kinase C), Ron = recepteur d'origine nantais) phosphorylation site at serine 13 was published (15).

PPIases are a family of enzymes catalysing isomerization of peptidyl-prolyl bonds. This isomerization seems to be important for transport through membranes, cell cycle progression, channel gating and virus replication (7, 8). Furthermore, peptidyl-prolyl-*cis/trans*-isomerization is a slow process and can, therefore, be the rate-limiting step in protein folding (9).

Depending on their primary structure and drug-binding ability PPIases are subdivided into three distinct families: cyclophilins, FK506-binding proteins (FKBPs) and parvulins resulting in a total of 35 human members all together.

The used proteome-wide PPIase-scan peptide microarray contains the complete primary structure of all human PPIases in the form of overlapping peptide scans. The overlapping peptides have been synthesized by SPOT technology (10, 11) as aminooxyacetyl derivatives allowing chemoselective and covalent immobilization onto epoxy-modified glass slides. These include the following FKBPs, cyclophilins and parvulins resulting in 3280 peptides: (listed in brackets are the corresponding SwissProt identification numbers followed by the length of the PPIase derived peptides and the number of overlapping amino acids) FKBP12 (P20071, 15/9), FKBP12.6 (Q16645, 15/9), FKBP13 (P26885, 15/9),

FKBP19 (Q9NYL4, 15/9), FKBP22 (Q9NWM8, 15/9), FKBP23 (Q9Y680, 15/9), FKBP25 (Q00688, 15/9), FKBP36 (O75344, 15/8), FKBP37 (O00170, 15/7), FKBP38 (Q14318, 15/7), FKBP44 (Q9NZN9, 15/7), FKBP51 (Q13451, 15/7), FKBP52 (Q02790, 15/7), FKBP63 (O95302, 15/7), FKBP65 (Q96AY3, 15/7); FKBP135 (Q9Y4DO, 15/3), Cyp18 (P62937, 13/7), Cyp18.1 (Q9H2H8, 13/9), Cyp18.2 (Q9Y536, 13/7), Cyp18.2a (Q9Y3C6, 15/9), Cyp19.2 (O43447, 13/7), Cyp22 (P30405, 13/7), Cyp23 (P23284, 13/7), Cyp23a (P45877, 15/9), Cyp33 (Q9UNP9, 15/9), Cyp35 (Q81XY8, 15/9), Cyp40 (Q08752, 15/9), Cyp57 (Q8WUA2, 15/9), Cyp59 (Q13356, 15/9), Cyp73 (Q96BP3, 15/9), Cyp89 (Q13427, 15/5), Cyp165 (P30414, 15/1), Pin1 (Q13526, 13/7), Par14 (Q9Y237, 13/7), Par17 (Q52M21, 15/1).

All peptides were synthesized with amino-oxy-acetylated *N*-(3-{2-[2-(3-amino-propoxy)-ethoxy]-ethoxy}-propyl)-succinyl moiety at the N-terminus allowing both optimal presentation of the peptide to the used enzymes and purification of the immobilized peptides by removal of truncated by-products during microarray printing (1). All FKBP-derived peptides were printed 6-fold to one standard industry glass slide resulting in $6 \times 1450 = 8700$ data points on the so-called FKBP-Chip. Peptides derived from human cyclophilins and parvulins were immobilized in quadruplicates on one standard industry glass slide resulting in $4 \times 1830 = 7320$ spots on the so-called Cyp-Chip. Peptide microarrays were designed and produced by JPT Peptide Technologies GmbH (Berlin, Germany).

Subsequent to incubation of these peptide microarrays with different kinases incorporated ^{33}P-phosphate was visualized by phosphorimaging (*see* **Figs. 9.3** and **9.4**). The signals found in the phosphorimages can be correlated with peptides from the corresponding PPIases using spot recognition software packages like Genepix 6.0. Phosphorylation sites found on microarrays were validated in vitro using full-length proteins together with appropriate kinase and radioisotopically labelled ATP (*see* **Fig. 9.5**).

Technologically there is no need for restriction to just one family of enzymes. In principle production of peptide microarrays displaying overlapping peptides covering complete proteomes of viruses or bacteria resulting in about 50,000 or 250,000 peptides, respectively, should be feasible. We synthesized cytomegalovirus (CMV) derived peptides representing the primary structure of all proteins of complete CMV proteome as 17,181 overlapping peptides. All CMV peptides were printed onto one standard industry glass slide in duplicates (*see* **Fig. 9.6**) resulting in $2 \times 17{,}181 = 34{,}362$ individual data points on one peptide microarray. To fit all of these peptide spots on one standard industry glass slide we reduced spot size down to 85–120 μm. Resolution of most

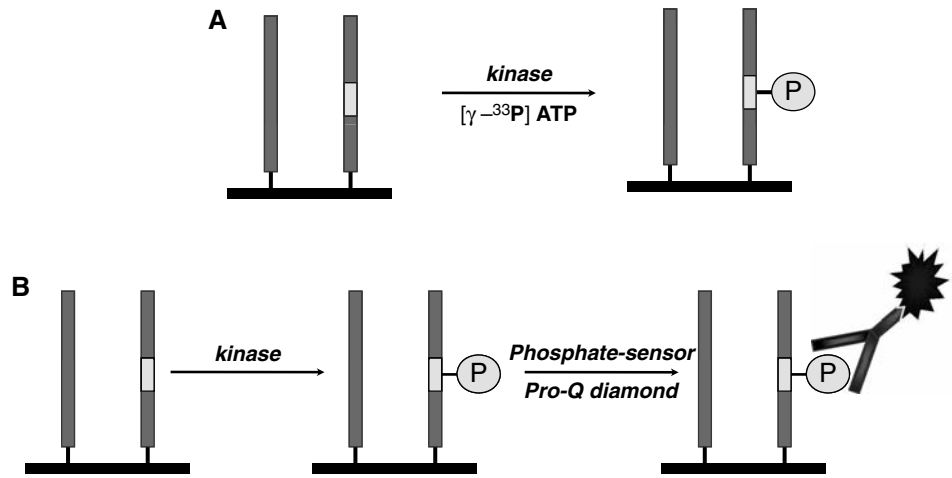


Fig. 9.3. Kinase assay principles. (**A**) Kinase-mediated transfer of radioisotopically labelled phosphate residues from γ^{33}P-ATP to glass surface bound peptides results in radioisotopically labelled phosphopeptides which could be detected by phosphorimaging. (**B**) Kinase-mediated phosphate transfer yields bound phosphopeptides which will capture phospho-specific fluorescence dyes like ProQ-Diamond. Bound dye molecules could be detected by fluorescence scanning using standard microarray imagers equipped with *green* lasers.

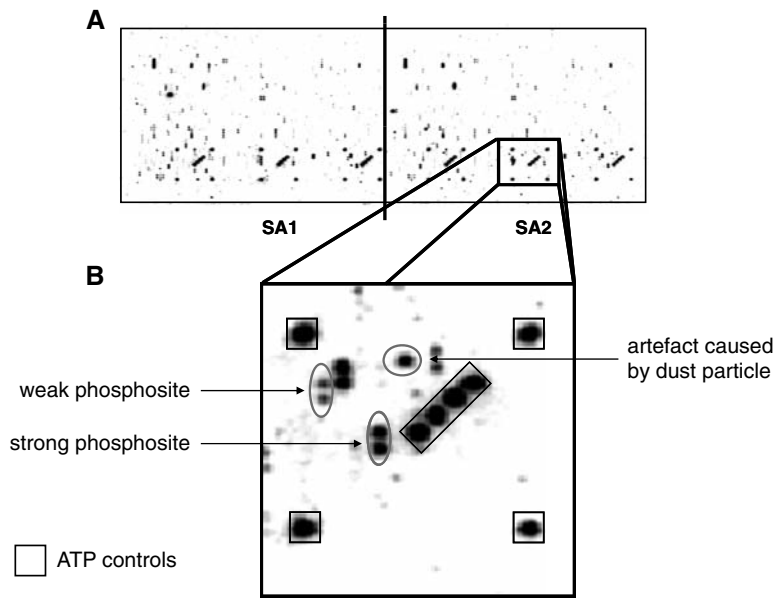


Fig. 9.4. (**A**) Phosphorimage of Cyp-Chip subsequent to incubation with PKCα kinase in the presence of γ–^{33}P-ATP is shown. Black spots represent peptide spots which are radioisotopically labelled by incorporation of ^{33}P-phosphate moieties. The vertical, solid line separates the two identical subarrays. (**B**) Enlarged region within one subarray which was printed by one needle. ATP binding control spots (eight identical peptide spots which are known to capture ATP from the assay solution) are visible at the edges of the *square* and as *diagonal line* (marked with *squares*). Additionally, strong end weak phosphorylation sites are visible as pairs of spots which are arranged side by side. This arrangement enables easy identification of artefacts which would result in single spots only.

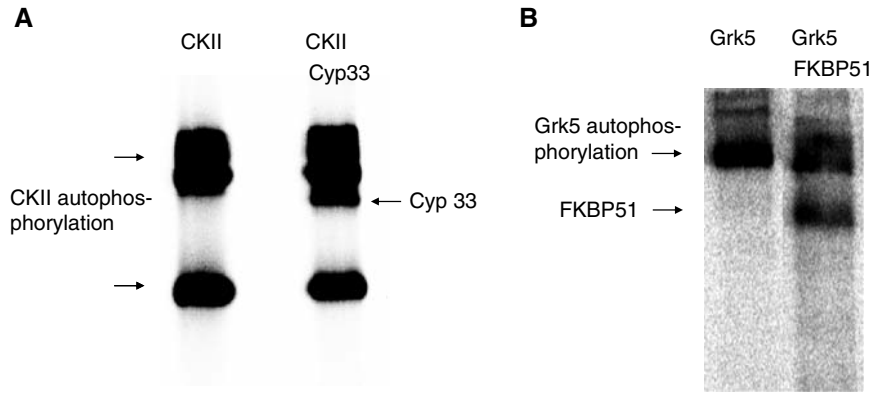


Fig. 9.5. (**A**) Analysis of in vitro phosphorylation of full-length cyclophilin 33 (Cyp33) by recombinant casein kinase 2 (CK2). 30 μ*M* Cyp33 were treated with 0.1 U of CK2 for 4 h in the presence of γ–^{33}P ATP in 10 μL total assay volume, boiled with 10 μL 2X sample buffer and used for separation in 15% polyacryl amide gel electrophoresis. Radioisotopically labelled proteins were visualized by phosphorimaging using Fuji FLA 3000 reader. Assay solution with casein kinase 2 but without Cyp33 was treated in similar way to correct for signals caused by autophosphorylation of casein kinase subunits. (**B**) Analysis of in vitro phosphorylation of full-length FKBP51 by recombinant G-protein-coupled receptor kinase 5 (Grk5). 30 μ*M* of FKBP51 were treated with 0.1 U of Grk5 for 4 h in the presence of γ^{33}P-ATP in 10 μL total assay volume, boiled with 10 μL 2X sample buffer and used for separation in 15% polyacryl amide gel electrophoresis. Radioisotopically labelled proteins were visualized by phosphorimaging using Fuji FLA 3000 reader. Assay solution with Grk5 but without FKBP51 was treated in similar way to correct for signals caused by autophosphorylation of Grk5.

phosphorimagers (pixel size in the range of 25–50 μm) is not sufficient for readout of kinase-mediated phosphate transfer to peptides within such small spots. Therefore, we used commercially available, fluorescently labelled, phospho-specific stain ProQ Diamond (Invitrogen) which was claimed to bind specifically to peptides containing phosphorylated amino acids residues like phosphotyrosine, phosphothreonine and phosphoserine (12, 13). We validated this stain extensively using phosphopeptide microarrays either displaying more than 15,000 phoshopeptides derived from human phosphorylation sites or 18,432 human peptides in the non-phosphorylated form (14). Phosphorylation could be measured by reading fluorescence using standard DNA microarray scanners at green channel settings. Spot recognition software identified all pixels corresponding to each spot and calculates the mean of signal intensities for all of these pixels (signal intensity). Additionally, mean intensity for all pixel around the identified spot is calculated (background signal intensity) and finally background signal intensity is substracted from signal intensity resulting in background corrected signal intensities (signal minus background). We were able to demonstrate for kinases Grk5, CK2 and Abl that this approach allows determination of phosphorylation sites in a proteome-wide manner (*see* **Fig. 9.6**). We identified phosphorylation sites for CK2 which were known from the literature plus additional unknown phosphorylation sites (*see* **Tables 9.1** and **9.2**).

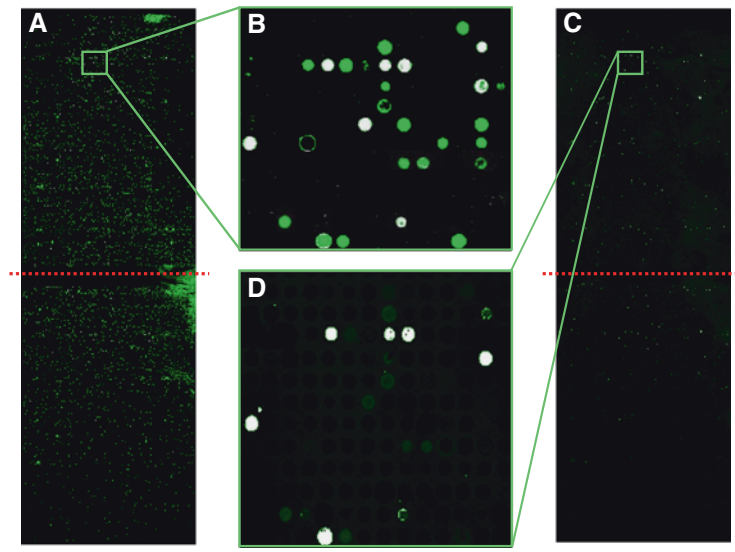


Fig. 9.6. Fluorescence images of high-content peptide microarrays displaying complete human cytomegalovirus proteome as overlapping peptides are shown. (**A**) Image of CMV-Chip displaying 2 × 17,181 = 34,362 individual peptides is shown. CMV-Chip was treated with casein kinase 2 followed by incubation with ProQ-Diamond stain which is specific for phosphorylated peptides and proteins. Two identical subarrays each displaying 17,181 peptides are separated by *dotted red line*. (**B**) Area of fluorescence image is zoomed out. White spots are related to fluorescence signals. No signal is indicated by *black* colour. (**C**) Image of CMV-Chip displaying 2 × 17,181 = 34,362 individual peptides is shown. CMV-Chip was treated with G protein coupled receptor kinase 5 followed by incubation with ProQ-Diamond stain which is specific for phosphorylated peptides and proteins. Two identical subarrays each displaying 17,181 peptides are separated by *dotted red line*. (**B**) Area of fluorescence image is zoomed out. White spots are related to fluorescence signals. No signal is indicated by *black* colour.

2. Materials

2.1. Equipments

1. Micropipettes adjustable from 0.5 to 10 µL and from 100 to 1000 µL (Eppendorf or Gilson) with corresponding plastic tips.
2. Metal trays with cover.
3. 1 mL Eppendorf tubes for the preparation of the phosphorylation assay solution.
4. Single-use syringe and syringe filters for filtration of all assay solutions (Sartorius, Göttingen, Germany).
5. A plastic sheet cut in 1 mm × 3 cm pieces. We used sheets with a thickness of 0.8 mm resulting in a final assay volume of 400 µL.
6. A centrifuge for standard industry glass slides.
7. Spot-recognition software like GenePix Pro 6.0.

Table 9.1
Peptide sequences for top substrates identified for casein kinase 2 using peptide microarray displaying complete human cytomegalovirus proteome as overlapping peptides are given. Kinase-mediated phosphate transfer to microarray-bound peptides was detected by fluorescently labelled phospho-specific stain ProQ Diamond. Potential phosphorylation sites are marked in bold. If more than one phosphorylation site is possible in one sequence several residues are marked. Right column lists corresponding data base identifiers for the proteins from which peptides are derived followed by position of first, N-terminal amino acid of the peptide in the primary structure of the protein

Top 40 detected substrates for CK2 on CMV-Chip	Corresponding CMV proteins
WLLND**S**DGEEEEMS	CAA35312_045
TLMPYVLFRRD**T**DTE	CAA35406_217
AAFESCCYDI**T**EAES	CAA35438_041
QSKCAFLKGYL**S**EGC	CAA35385_129
FYRAFRSGRFDLC**T**D	CAA35391_169
KKLVEMMEQHDRG**S**D	CAA35341_329
CYVIEFKTTY**S**DADD	CAA35391_097
ERLFRDPL**T**TYEYLD	CAA35426_029
MELD**S**VEEEDDFGAS	CAA35404_137
SDFDADCWCMWGRFG	CAA35449_129
GNARLDALMSA**S**EWW	CAA35404_645
KDNQCM**T**DYDYLEVS	P09704_169
WVA**S**EDELDVSRGDA	CAA35418_084
RTMAFLRFERYD**T**DY	CAA35408_153
PSQ**S**EVDCASLMETL	CAA35418_061
SGLLP**S**CEEDERELC	CAA35447_229
KYAE**S**DYIFLQDMCP	CAA35396_353
PQT**S**DLECAKQYWQE	CAA35293_421
HLLGTE**S**DDEETTVW	CAA74073_053
MLCYY**T**EKLEEIDSK	CAA35442_069
P**S**TLETFPDLFCLPL	CAA35390_561
SEPEDDDEDP**T**YDEL	CAA35311_721
LP**S**LREDYAQL**S**DVI	CAA35424_453

(continued)

Table 9.1 (continued)

Top 40 detected substrates for CK2 on CMV-Chip	Corresponding CMV proteins
E**S**CEDVPSGKLFMH	CAA35357_229
EFEDIFGSAAV**T**DGP	CAA35293_061
KNSVRHMSSFVS**S**DI	CAA35406_261
IRYYVSVYDEL**T**ASE	CAA35295_225
DEF**S**FCD**S**DIEDFER	CAA35424_173
ATAP**S**FDEAFL**T**DRL	CAA35407_389
MYPGEKNTQQLLD**S**D	CAA35262_173
WDSMHCTPFWS**T**DLE	AAA85887.1_061
KVYLE**S**FCEDVPSGK	CAA35357_225
MDWKAHVEYAHPA**S**E	CAA35312_093
RLFFPRED**S**EPLM**S**D	CAA35416_081
DFKKWLDGGFS**T**AVE	CAA35363_133
GYMPIHCPSE**T**D**T**DS	CAA35327_013
PVVPEECYDQRFT**T**E	CAA35311_697
DEF**S**FCDSDIEDFER	CAA35424_173
TVYPTYDCVL**S**DLEA	CAA35392_265
RTPLCYA**S**ELCDESV	CAA35404_253

Table 9.2
Peptide sequences for top substrates identified for G-protein coupled receptor kinase 5 using peptide microarray displaying complete human cytomegalovirus proteome as overlapping peptides are given. Kinase-mediated phosphate transfer to microarray-bound peptides was detected by fluorescently labelled phospho-specific stain ProQ Diamond. Potential phosphorylation sites are marked in bold. If more than one phosphorylation site is possible in one sequence several residues are marked. Right column lists corresponding data base identifiers for the proteins from which peptides are derived followed by position of first, N-terminal amino acid of the peptide in the primary structure of the protein

Top 40 detected substrates for GRK5 on CMV-Chip	Corresponding CMV proteins
NVVSYVCEEHLH**S**FT	CAA35363_0361
AAFESCCYDI**T**EAES	CAA35438_0041
GGLCSSMAVYDEE**T**M	CAA35430_0341

(continued)

Table 9.2 (continued)

Top 40 detected substrates for GRK5 on CMV-Chip	Corresponding CMV proteins
RTPLCYASELCDE**S**V	CAA35404_0253
DIEDFERECYRV**S**VA	CAA35424_0181
TPLSEAMFAGFEEA**S**	CAA35356_0437
KYAE**S**DYIFLQDMCP	CAA35396_0353
SSMAVYDEE**T**MRQSQ	CAA35430_0345
KVYLESFCEDVP**S**GK	CAA35357_0225
DEFSFCD**S**DIEDFER	CAA35424_0173
ELLIEDFDIYVD**S**FP	CAA35431_0269
LHDCAAFE**S**CCYDIT	CAA35438_0037
ESFCEDVP**S**GKLFMH	CAA35357_0229
KDNQCMTDYDYLEV**S**	P09704_0169
DWIRFLSLPDHD**T**VL	CAA35316_0033
LHPFFDFTHCQEN**S**E	CAA35360_0533
AAEGDEFSFCD**S**DIE	CAA35424_0169
DWCSMRNSLDEV**S**GT	CAA35293_0393
HHYDLCFTCDRNL**S**L	AAA85896.1_0245
ETFPDLFCLPLGE**S**F	CAA35390_0565
RFTEDTFVE**T**FCDFL	CAA35426_0465
MELDSVEEEDDFGA**S**	CAA35404_0137
LDDAFLD**T**LALLYNN	P16832_0101
VDAVVHPLALDFR**S**E	CAA35282_0193
MLCYYTEKLEEID**S**K	CAA35442_0069
LDMSSLYYNE**T**MFVE	CAA35389_0377
PAVY**T**CVDDLCRGYD	P16832_0149
QKLGWCLADDIHT**S**F	CAA35405_0249
ASRGTVFEEETVW**S**L	CAA35404_0789
GYEGFGWDGE**T**LMEL	CAA35417_0133
PVVPEECYDQRF**T**TE	CAA35311_0697
CYASELCDE**S**VRRFV	CAA35404_0257
RDCYERFVCPVYD**S**G	CAA35277_0121
VGLRLHDCAAFE**S**CC	CAA35438_0033
ITG**S**EFEGDFARYRS	CAA35407_1157

(continued)

Table 9.2 (continued)

Top 40 detected substrates for GRK5 on CMV-Chip	Corresponding CMV proteins
LSVEEICEEH**T**LNDL	CAA35424_0397
DL**S**RWFGENMDEY**S**G	CAA35454_0089
MAYYLFEGQY**S**TIST	CAA35386_0277
ATAPSFDEAFL**T**DRL	CAA35407_0389
TFNVSMD**T**AGMYECV	AAA85880.1_0217
FLGDMQLPADNFL**T**S	CAA35430_0677

2.2. Radioisotopic Readout

1. Phosphorimager with a pixel size of at least 50 μm pixel sizes smaller than 50 μm will result in more accurate data points. We used a Fuji FLA 3000 scanner at 50 μm resolution.
2. Double distilled water.
3. Phosphoric acid.
4. γ–^{33}P-ATP.
5. Methanol.
6. Blocking buffer: Tris-buffered saline (50 m*M* Tris-HCl, 150 m*M* NaCl, pH 8.0, 3% bovine serum albumin, w/vol).
7. Phosphorylation assay solution:
   - 1 Unit of respective kinase (unit = 1 nmol transferred phosphate per mg enzyme per min reaction time)
   - 4 μL Kinase assay buffer (according to the manufacturer protocol)
   - 2 μL 10 m*M* ATP, 10 m*M* $MgCl_2$
   - 18 μL γ–^{33}P-ATP (180 μCi)
   - Add 400 μL H_2O

   Before you add the γ^{33}P-ATP it is recommended to filter the solution through a 0.45 μm filter and pre-incubate the "cold" assay solution for 5 min at room temperature to allow "invisible" autophosphorylation of the kinase (*see* **Note 12**).

2.3. Fluorescence Readout

1. Fluorescence scanner with a pixel size smaller than 20 μm. Pixel sizes smaller than 20 μm will result in more accurate data points. We used an Axon4000 scanner at 10 μm resolution.
2. ProQ Diamond stain (Invitrogen, Karlsruhe, Germany).
3. ProQ Diamond destain solution or 20% acetonitrile, 50 *mM* sodium acetate pH 4.0.
4. Double distilled water.

5. Blocking buffer: Tris-buffered saline (50 m*M* Tris-HCl, 150 m*M* NaCl, pH 8.0, 3% bovine serum albumin, w/vol).
6. Washing buffer: 50 m*M* Tris-HCl, 150 m*M* NaCl, 0.2% Tween20, pH 8.0.
7. Phosphorylation assay solution for fluorescence readout :
   - 1 Unit of respective kinase (unit = 1 nmol transferred phosphate per mg enzyme per min reaction time)
   - 40 µL Kinase assay buffer (according to the manufacturer protocol)
   - 2 µL 10 m*M* ATP, 10 m*M* $MgCl_2$
   - Add 400 µL H_2O

3. Methods

3.1. Radioisotopic Readout

1. Syringe filters have to be pre-treated with blocking buffer for 1 h at room temperature (*see* **Notes 1–11**).
2. Prepare the phosphorylation assay (*see* **Notes 13–14**).
3. For the phosphorylation reaction two slides, one displaying the peptides and another slide without any peptides but coated with blocking buffer for 1 h at room temperature have to be assembled according to **Fig. 9.7** in a sandwich-like format. If two peptide microarrays should be screened the top slide could be another peptide-displaying chip. The two slides are separated by two spacers generated from a plastic sheet (*see* **Fig. 9.7**). The final assay volume will depend on the thickness of these plastic spacers (0.2 mm thickness will result in 100 µL assay volume; we recommend at least 0.8 mm thickness resulting in 400 µL). The sample has to be applied in between the two slides. Therefore, the top slide is shifted about 1 mm to one side (*see* **Fig. 9.7**). If the pipette tip is adjusted on the position directly over the uncovered bottom slide the capillary forces allow proper distribution of the sample solution without formation of air bubbles.
4. The phosphorylation reaction should be performed at 37°C for 3–4 h.
5. Remove the plastic spacers and rinse the peptide microarray with 2% phosphoric acid to stop the reaction and to remove excess of radioisotopically labelled ATP. Repeat this step two times.
6. Wash microarray with distilled water and methanol (three times each).

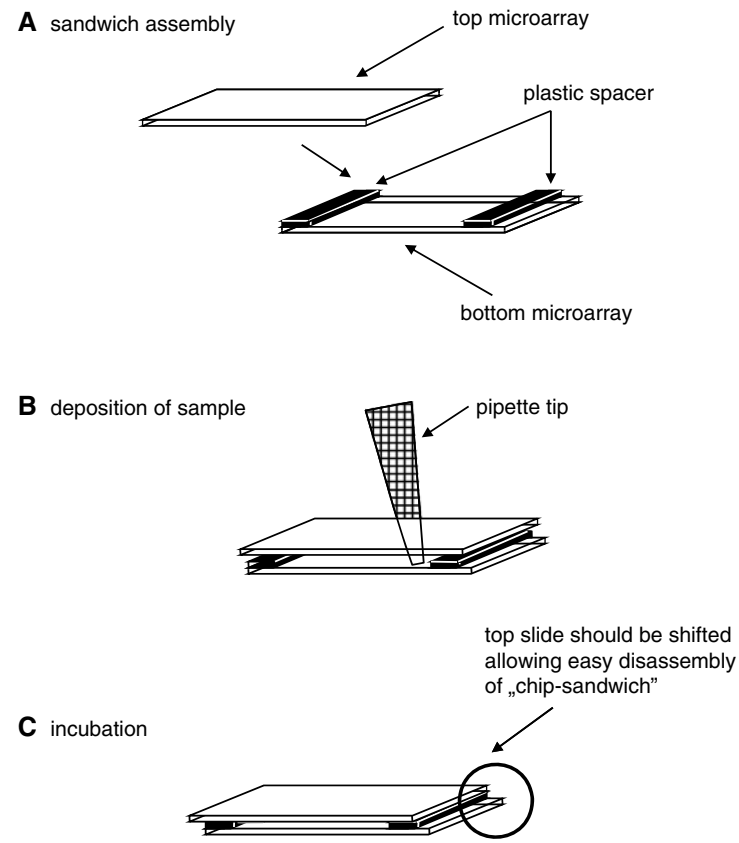


Fig. 9.7. (**A**) Assembly of "chip sandwich" is shown. Two plastic spacers are placed between the peptide displaying microarray (*bottom* microarray) and the dummy slide or second peptide displaying microarray (*top* microarray) resulting in a defined reaction chamber. (**B**) Assay solution is applied via pipette tip into the reaction chamber formed by the two slides. Capillary forces will soak-in the solution without formation of air bubbles. (**C**) *Top* microarray is shifted resulting in overlaying ends of the glass slides. This arrangement enables convenient disassembly after the incubation step.

7. Dry microarray using a microarray centrifuge. Alternatively, drying using a stream of oil-free nitrogen or argon can be applied.
8. Following incubation with target kinase in presence of radioisotopically labelled ATP, incorporated phosphate moiety can be detected by autoradiography or phosphorimaging for 1–24 h depending on the activity of the kinase.
9. Scan the screen using a phosphorimager.
10. Use spot recognition software to get signal intensities for each peptide spot.
11. Calculate the mean of background corrected signal intensities for identical peptides (replicates on the microarray).

3.2. Fluorescence Readout

1. Syringe filters have to be pre-treated with blocking buffer for 1 h at room temperature.
2. Prepare the phosphorylation assay solution (*see* **Note 15**).
3. For phosphorylation reaction two slides, one displaying the peptides and another slide without any peptides but coated with blocking buffer for 1 h at room temperature have to be assembled according to **Fig. 9.7** in a sandwich-like format. If two peptide microarrays should be screened the top slide could be another peptide-displaying chip. The two slides are separated by two spacers generated from a plastic sheet (*see* **Fig. 9.7**). The final assay volume will depend on the thickness of these plastic spacers (0.2 mm thickness will result in 100 µL assay volume; we recommend at least 0.8 mm thickness resulting in 400 µL). The sample has to be applied in between the two slides. Therefore, the top slide is shifted about 1 mm to one side (*see* **Fig. 9.7**). If the pipette tip is adjusted on the position directly over the uncovered bottom slide the capillary forces allow proper distribution of the sample solution without formation of air bubbles.
4. The phosphorylation reaction should be performed at 37°C for 3–4 h.
5. Remove the plastic spacers and rinse the peptide microarray with washing buffer three times for each side of the slide.
6. Incubate microarray with ProQ Diamond stain solution for 30 min.
7. Wash microarray with ProQ Diamond destain solution or 20% acetonitrile, 50 *mM* sodium acetate, pH 4.0 (three times, 30 min each time).
8. Wash two times 5 min with double distilled water.
9. Dry microarray using a microarray centrifuge. Alternatively, drying using a stream of oil-free nitrogen or argon can be applied.
10. Scan peptide microarray at appropriate laser/filter settings using microarray fluorescence scanner (*see* **Notes 16 and 17**).
11. Use spot recognition software to get mean of fluorescence intensities of all pixels for each peptide spot.
12. Calculate the mean of background corrected signal intensities for identical peptides (replicates on the microarray).

4. Notes

1. Always handle peptide microarray slides with care. They are made of glass and can damage your fingers.
2. Never touch the peptide microarray surface.

3. Always wear laboratory gloves when handling peptide microarray slides.
4. Hold peptide microarray slides at the end, which carries the engraved data label. This label provides unique identification of the array.
5. Please take care when dispensing solutions onto the slide surface. Make sure not to touch the surface with pipette tips or dispensers.
6. Apply the sample to the peptide displaying side of the glass slide. This is the side with the engraved label.
7. We strongly recommend arrangement of the two slides during incubation as shown in **Fig. 9.7** (circled region bottom). The slight shift of the top slide compared to the bottom slide allows easy disassembly of the two glass slides.
8. Never whisk the surface of the peptide microarray slide with a cloth.
9. Never use other chemicals as described. Inappropriate chemicals may destroy the chemical bond formed between the peptides and the modified glass surface.
10. Filter all solutions for the washing steps through 2 μm, preferably 0.45 μm particle filters before use.
11. Peptide replicates are printed side by side to allow differentiation between real signals and signals caused by artefacts during radioisotopic readout. Small dust particles or damages on the slide surface could cause round, spot-like looking signals but never two signals with similar shape side by side (*see* **Fig. 9.4**).
12. If your kinase tends to autophosphorylation, please preincubate your kinase with cold ATP only before adding hot ATP. In case your kinase shows unspecific binding to the hydrophobic microarray surface, the background will not be increased because it is not visible in the phosphorimage.
13. All plastic equipment which will contact your kinase solution (pipette tips, plastic tubes, etc.) should be coated for at least 1 h with bovine serum albumin solution (blocking buffer) at room temperature before use. It is important to coat for at least 1 h because that process is time dependent. If coating is not made or if coating time is too short, kinase from your assay solution will stick on un-coated parts of the hydrophobic plastic surfaces and is not available for phosphorylation of microarray-bound peptides. This could result in complete loss of kinase activity in the final assay especially if highly diluted kinase solutions are used.

14. Avoid dust or other particles during each step of the experiment. Dust particles and resulting scratches will cause artefacts during the final signal readout especially during radioisotopic readout.
15. For incubation with the fluorescently labelled phospho-specific stain it is important to use metal trays with a cover or plastic trays completely covered with aluminium foil as these reagents are sensitive towards light.
16. Make sure to scan the front face of the microarray slide displaying the peptides.
17. Fluorescence scanning could be very sensitive depending on the used scanner. Avoid any fluorescent impurities/contaminations inside your assay solution or washing solutions. You can easily check for such impurities by incubating and washing a dummy slide with the same solutions followed by fluorescence imaging at same laser settings.

References

1. Geiss-Friedlander, R., Melchior, F. (2007). Concepts in sumoylation: a decade on. *Nat Rev Mol Cell Biol.* **8**, 947–56.
2. Hunter, T. (2007). The age of crosstalk: phosphorylation, ubiquitination, and beyond. *Mol Cell.* **28**, 730–8.
3. Zhang, H., Zha, X., Tan, Y., Hornbeck, P. V., Mastrangelo, A. J., Alessi, D. R. Polakiewicz, R. D., Comb, M. J. (2002). Phosphoprotein analysis using antibodies broadly reactive against phosphorylated motifs. *J Biol Chem.* **277**, 39379–87.
4. Adams, J. A. (2003). Activation loop phosphorylation and catalysis in protein kinases: is there functional evidence for the autoinhibitor model? *Biochemistry.* **42**, 601–7.
5. Yu, R. K., Bieberich, E. (2001). Regulation of glycosyltransferases in ganglioside biosynthesis by phosphorylation and dephosphorylation. *Mol Cell Endocrinol.* **177**, 19–24.
6. Bodenmiller, B., Mueller, L. N., Mueller, M., Domon, B., Aebersold, R. (2007). Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nat Methods.* **4**, 231–237.
7. Kay, J. E. (1996). Structure–function relationships in the FK506-binding protein (FKBP) family of peptidylprolyl cis–trans isomerases. *Biochem J.* **314**, 361–85.
8. Fanghänel, J., Fischer, G. (2004). Insights into the catalytic mechanism of peptidyl prolyl cis/trans isomerases. *Front Biosci.* **9**, 3453–78.
9. Drakenberg, T., Forsen, S. (1971). The barrier to internal rotation in monosubstituted amides. *J Chem So. Chem Commun.* 1404–1405.
10. Frank, R. (2002). The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports – principles and applications. *J Immunol Methods.* **267**, 13–26.
11. Reineke, U., Volkmer-Engert, R., Schneider-Mergener, J. (2001). Applications of peptide arrays prepared by the SPOT-technology. *Curr Opin Biotechnol.* **12**, 59–64.
12. Martin, K., Steinberg, T. H., Cooley, L. A., Gee, K. R., Beechem, J. M., Patton, W. F. (2003). Quantitative analysis of protein phosphorylation status and protein kinase activity on microarrays using a novel fluorescent phosphorylation sensor dye. *Proteomics* **3**, 1244–55.
13. Steinberg, T. H., Agnew, B. J., Gee, K. R., Leung, W. Y., Goodman, T., Schulenberg, B., Hendrickson, J., Beechem, J. M., Haugland, R. P., Patton, W. F. (2003). Global quantitative phosphoprotein analysis using multiplexed proteomics technology. *Proteomics* **3**, 1128–44.

14. Zerweck, J., Masch, A., Schutkowski, M. (2009). Peptide microarrays for profiling of modification state specific antibodies, in *Epitope mapping protocols, methods in molecular biology* (Reineke, U. and Schutkowski, M., eds), Humana, Totowa, NJ, *Methods Mol. Biol.* **524**, 169–80.

15. Beausoleil. S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Vilén, J., Li, J., Cohn, M.A., Cantley, L. C., Gygi, S. P. (2004). Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci USA*. **101**, 12130–5.

Chapter 10

Epitope Mapping of Human Chromogranin A by Peptide Microarrays

Marina Cretich, Renato Longhi, Angelo Corti, Francesco Damin, Gabriele Di Carlo, Valentina Sedini, and Marcella Chiari

Abstract

In this chapter we report on the characterization of linear antigenic sites of human chromogranin A (CgA), a useful tissue and serum marker for neuroendocrine tumours and a precursor of many biologically active peptides. The epitope mapping of CgA has been carried out by peptide microarrays on glass slides coated by a copolymer of *N,N*-dimethylacrylamide (DMA), *N,N*-acryloyloxysuccinimide (NAS) and [3-(methacryloyl-oxy) propyl] trimethoxysilyl (MAPS). The microarray support provided sufficient accessibility of the ligand, with no need for a spacer, as the polymer chains prevent interaction of immobilized peptides with substrate. In addition, the polymeric surface constitutes an aqueous micro-environment in which, despite peptide random orientation, linear epitopes are freely exposed. The results reported are in accordance with those obtained in conventional ELISA assays using biotinylated and non-biotinylated peptides.

Key words: Epitope mapping, peptide microarray, human chromogranin A, polymer coating.

1. Introduction

Human chromogranin A (CgA) is an acidic glycoprotein of 439 aminoacids contained in secretory granules of many endocrine and neuroendocrine cells. CgA antigen is a useful tissue and serum marker for neuroendocrine tumours. It is believed that CgA plays a role in secretory granule biogenesis and as a precursor of many biologically active peptides with endocrine, paracrine or autocrine functions (1). For instance, the CgA fragment corresponding to amino acids 1–76 (vasostatin-1) inhibits vascular tension, exerts anti-fungal and antibacterial effects, promotes cell adhesion and induces negative inotropic effects (1, 2). Proteolitically processed products of CgA have also a role as new markers in endocrine

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_10,

neoplasias (3). Recently, a proteomic strategy based on chromatographic fractionation followed by MALDI-TOF mass spectrometry has been proposed to isolate and characterize a spectrum of CgA-derivated peptides (4).

The linear antigenic sites of human CgA were characterized by studying the cross-reaction of rabbit polyclonal and mouse monoclonal anti CgA antibodies with native and synthetic CgA peptides encompassing most of its sequence (5, 6). The study with rabbit anti human CgA identified several strong antigenic sites within residues 68–106, 222–230, 315–330, 376–394. Moreover, residues 53–57 and 68–70 were strongly recognized by the mouse monoclonal antibodies 5A8 and B4E11, respectively (5, 6). These antibodies (5A8 and B4E11) specifically neutralize the pro-adhesive activity of vasostatin-1 in fibroblast adhesion assays (6). In this report, a microarray approach is presented for the characterization of linear antigenic sites of human CgA using copoly (DMA-NAS-MAPS)-coated slides. The microarray substrates were obtained by physical adsorption on glass of a *N,N*-dimethylacrylamide (DMA), *N,N*-acryloyloxysuccinimide (NAS) and [3-(methacryloyl-oxy)-propyl] trimethoxysilyl (MAPS) copolymer (7), that provides a polymeric active surface on which peptides are immobilized. Microarraying peptides is a highly efficient method for epitope mapping, screening of enzyme specificity or ligand–receptor interactions. Since peptides have small molecular mass, they are not easily accessible when non-specifically adsorbed on solid supports (8). Moreover, since peptides often lack a well-defined three-dimensional structure, a correct orientation is essential to promote their interaction with the target (9). For these reasons, chemoselective N-terminal attachment of peptides on slides have been developed based on glyoxylic acid and N-terminally cysteine-containing peptides (10), on Diels Alder reactions (11), on native chemical ligation and on avidin–biotin interaction (10). A different approach is the SPOT-synthesis developed by Frank (12) based on the parallel synthesis of peptides directly on cellulose membranes.

In this protocol we show the suitability of a glass slide coated with a copolymer of *N,N*-dimethylacrylamide (DMA), *N,N*-acryloyloxysuccinimide (NAS) and [3-(methacryloyl-oxy) propyl] trimethoxysilyl (MAPS) as a peptide microarray substrate. The coating is formed by a fast, inexpensive and robust procedure consisting in simply immersing slides in a dilute aqueous polymer solution without time-consuming glass pre-treatments. Copoly (DMA-NAS-MAPS)-coated slides were already successfully used in protein–protein interaction experiments (13) where we demonstrated that both Fc and Fab domains of a capture antibody were freely accessible once bound to the surface. Copoly (DMA-NAS-MAPS)-coated slides immobilize probes in a random conformation; nevertheless, the polymer coating creates an aqueous

micro-environment in which linear epitopes are freely exposed. Herein the polymeric surface was used for peptide immobilization. The results demonstrated a good accessibility of the ligand without need for a spacer between the probe and the surface, which is generally necessary in non-polymeric slides (8).

The peptide array approach identified several linear antigenic sites of human CgA, in accordance with previous results obtained by ELISA assays using the same peptides and the same polyclonal and monoclonal antibodies. Peptide microarray experiments reported in this work provide a precious tool for structure–function studies of CgA and of its proteolytic peptides.

2. Materials

2.1. Synthesis of the Copoly(DMA-NAS-MAPS)

1. *N,N*-Dimethylacrylamide (DMA) from Sigma (St. Louis, MO, USA).
2. 3-(Trimethoxysilyl)propyl methacrylate (MAPS) from Sigma (St. Louis, MO, USA).
3. *N*-Acryloyloxysuccinimide (NAS) from Polysciences (Warrington, PA, USA).
4. Tetrahydrofuran (THF) from Sigma (St. Louis, MO, USA).
5. α,α'-Azoisobutyronitrile (AIBN) from Sigma (St. Louis, MO, USA).
6. Petroleum ether from Sigma (St. Louis, MO, USA).

2.2. Derivatization of Glass Slide with Copoly(DMA-NAS-MAPS)

1. Untreated glass microscope slides (25 mm × 75 mm) from Sigma (St. Louis, MO, USA).
2. Pre-treatment solutions: ethanol 96% v/v; sodium hydroxide (NaOH) 1 M; hydrogen chloride (HCl) 1 M.
3. Copoly (DMA-NAS-MAPS) synthesized as described in **Section 3.1**.
4. Coating solution: 1% w/v of copoly (DMA-NAS-MAPS) in a water solution of ammonium sulphate at a 20% saturation level.

2.3. Preparation and Spotting of Peptides

1. Lyophilized peptides were synthesized on pre-loaded NovaSyn TGA resin (Novabiochem).
   The peptide sequences are reported in **Table 10.1**.
2. Spotting buffer: phosphate buffered saline solution (PBS 1X) (pH 7.5): 0.137 M NaCl, 0.0027 M KCl, 0.012 M Na_2HPO_4, 1.76 mM KH_2PO_4 filtered by a 0.22 um filter.

Table 10.1
Sequences of the peptides used in this work

Probe	Peptide code	Chromogranin Aresidues	Sequence
1	T01	−2–20	TALPVNSPMNKGDTEVMKCIVEgy
2	T27	25–46	TLSKPSPMPVSQECEETLRGDEgy
3	CgA38-57	38–57	CEETLRGDERILSILRHQNL
4	G70	68–91	yGAKERAHQQKKHSGFEDELSEVLE
5	E93	91–113	yENQSSQAELKEAVEEPSSKDVME
6	S109	107–130	SSKDVMEKREDSKEAEKSGEATDGy
7	G132	130–153	GARPQALPEPMQESKAEGNNQAPGy
8	T165	163–187	TNTHPPASLPSQKYPGPQAEGDSEG
9	G189	187–210	GLSQGLVDREKGLSAEPGWQAKRE
10	G224	222–244	GEEAVPEEEGPTVVLNPHPSLGY
11	G233	231–255	GPTVVLNPHPSLGYKEIPKGESRSE
12	G277	275–297	GKGEQEHSQQKEEEEEMAVVPQGy
13	G299	297–319	GLFRGGKSGELEQEEERLSKEWE
14	S317	315–337	SKEWEDSKRWSKMDQLAKELTAE
15	A333	331–352	AKELTAEKRLEGQEEEEDNRDSy
16	R376	374–396	RGWRPSSREDSLEAGLPLQVRGY
17	G397	395–417	GYPEEKKEEEGSANRRPEDQELE
18	CgA#2025	416–439	yLESLSAIEAELEKVAHQLQALRRG
19	CgA47-68	47–68	RILSILRHQNLLKELQDLALQGy
20	gp41	–	LIEESQNQQEKNEQELLELDKWAS

3. SpotBot® 2 Microarrayer equipped with two split spotting pins (Telechem, Sunnyvale, CA, USA).

2.4. Blocking and Incubation of Microarrays

1. 50 mM Ethanolamine in 0.1 M Tris/HCl pH 9.0.
2. Mouse anti CgA monoclonal antibodies (B4E11 and 5A8) and one rabbit polyclonal antibody were obtained as reported in (5).
3. Cy3-labelled secondary antibodies produced in goat (Jackson ImmunoResearch, USA).

4. Incubation buffer: 0.05 M Tris-HCl pH 7.6, 0.15 M NaCl, 1% w/V BSA, 0.02% w/V Tween 20.
5. Washing buffer: 0.05 M Tris/HCl pH 9, 0.25 M NaCl, 0.05% Tween 20.

2.5. Fluorescent Scanning and Data Analysis

1. Fluorescence signals were detected by a ScanArray Lite (Perkin Elmer, MA, USA) analysed by using the software associated to the scanner. Data were exported to Microsoft Excel software for further processing and signal intensities were defined in fluorescence arbitrary units. Signal-to-noise ratios (S/N) for fluorescence signals were calculated as the ratio between the mean spot signal intensity and the standard deviation of background intensity fluorescence.

3. Methods

The methods described in this section outline: (a) the synthesis of the copoly (DMA-NAS-MAPS); (b) the coating of standard microscope slides with the copoly (DMA-NAS-MAPS); (c) the preparation of the synthetic peptides; (d) the spotting of the microarrays; (e) the subsequent blocking and incubation; (f) the fluorescent scanning; and (h) the data analysis.

3.1. Synthesis of the Copoly(DMA-NAS-MAPS)

The copolymer, composed of DMA (with 97% of moles in the monomer feed), NAS (2%) and 3-(trimethoxysilyl) propyl methacrylate (MAPS, 1%) is synthesized by free radical copolymerization and used as a functional coating for the fabrication of the peptide microarrays.

1. Dissolve 4 g of DMA, 140.6 mg of NAS and 13 mg of the initiator azoisobutyronitrile (AIBN) in 42 mL of dried tetrahydrofuran (THF) in a 100 mL round-bottomed flask equipped with condenser, magnetic stirring and nitrogen connection. The concentration of the monomer feed in the solvent is 0.1 g/mL (10% w/v) (*see* **Note 1**).
2. Degas the solution by alternating vacuum and nitrogen purging for 10 min.
3. Dissolve in the same solution 103.3 mg of MAPS.
4. Stir magnetically and warm the solution to 65°C and keep it at his temperature under nitrogen for about 18 h.
5. After the polymerization is completed, precipitate the polymer by pouring the reaction mixture into a large excess of petroleum ether (about 1:10 by volume).

6. Filter the white solid copolymer.
7. Dry the copolymer under vacuum for 1–2 h at room temperature and store it in a dry environment.

3.2. Derivatization of Glass Slides with Copoly(DMA-NAS-MAPS)

The coating of glass slides requires two steps: (a) surface cleaning and pretreatment and (b) adsorption of the copolymer (DMA-NAS-MAPS).

1. Clean the microscope glass slides by immersion in pure ethanol for 30 min in a glass chamber at room temperature (*see* **Note 2**).
2. Dry the slides for few minutes in air.
3. Incubate slides in a solution of 1 M NaOH for 30 min at room temperature in a glass chamber.
4. Wash slides vigorously two or three times by dipping them, one slide at a time, into a Becker plenty of ddH_2O to remove all traces of NaOH.
5. Dry the slides for few minutes at air.
6. Incubate slides in a solution of 1 N HCl for 1 h at room temperature in a glass chamber.
7. Repeat step 4.
8. Spin dry in a centrifuge at 780 rpm for 3 min.
9. Prepare a solution of copoly(DMA-NAS-MAPS), 1% w/v in a water solution of ammonium sulphate at a 20% saturation level (*see* **Note 3**).
10. Incubate the slides in the solution prepared in step 9 for 30 min at room temperature in a plastic chamber (*see* **Note 2**).
11. Put the slides in a rack and wash them vigorously with ddH_2O to remove the excess of the copolymer on the slides.
12. Dry the slides in a vacuum oven at 80°C for 20 min.
13. Store slides at room temperature in an dry environment, and use slides within 4 weeks after production (*see* **Note 4**).

3.3. Synthesis of Peptides and Preparation for Spotting

3.3.1. Synthesis of Peptides

Several synthetic peptides (*see* **Table 10.1**) spanning most of the CgA sequence and one unrelated peptide were synthesized on pre-loaded NovaSyn TGA resin (Novabiochem), using an apparatus for manual multiple peptide synthesis and a solid-phase method based on Fmoc chemistry as described in (5). Tyrosine or a dipeptide glycine–tyrosine (glycine as spacer) was added, respectively, at the amino or carboxy terminus of peptides lacking chromophoric groups to allow spectrophotometric quantitation.

3.3.2. Preparation of Peptides for Spotting

1. All peptides reported in **Table 10.1** were dissolved in the spotting buffer: PBS 1X, at a concentration of 0.13 mg/mL (*see* **Note 5**).
2. Transfer 20 μL of each peptide solution into 384-well polypropylene plates.

3.4. Spotting

The following procedure is optimized onto the SpotBot® 2 Microarrayer robot. In case other arrayers are used, adjust procedure accordingly.

1. Design a microarray pattern to yield at least 10 replicates per different peptide and to accommodate 20 peptidic probes (19 CgA peptides and one unrelated peptide) on every slide according to the scheme reported in **Fig. 10.1A** (*see* **Note 6**).
2. Clean split spotting pins by sonication and place the pins into the spotting head.
3. Set humidity of the robot to 70%.

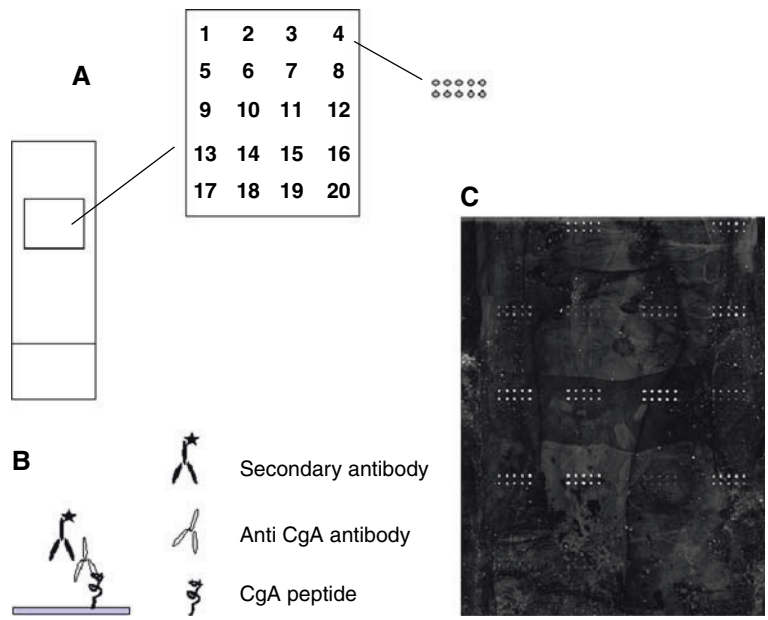


Fig. 10.1. (**A**) Spotting scheme. Each peptide reported in **Table 10.1** was dissolved in PBS buffer at a concentration of 0.13 mg/mL and spotted on copoly(DMA-NAS-MAPS) slides. Each probe was spotted in 10 replicates. (**B**) Scheme of peptide microarray experiment. CgA peptides immobilized on the slides were incubated with anti CgA monoclonal or polyclonal antibodies and revealed by a Cy3-labelled secondary antibody. (**C**) Typical microarray obtained after incubation with anti CgA rabbit polyclonal antibody. Reproduced with permission from Chiari M, Cretich M, Corti A, Damin F, Pirri G, Longhi R. (2005). Peptide microarrays for the characterization of antigenic regions of human chromogranin A. *Proteomics* **5**, 3600–3.

4. Program wash routine of the spotter to wash extensively between different peptides. A wash cycle is composed of a rapid wash step with ethanol 20% of 1.5 s, followed by a dry step of 2.5 s with airflow. The wash cycle should be repeated 6–7 times.
5. Printed slides were incubated overnight in a humid chamber at room temperature (*see* **Note 7**).

3.5. Blocking and Incubation of Microarrays

The most critical step in post-processing is the blocking step, where the remaining active surface is modified to minimize its ability to bind labelled sample.

1. Rinse slides with PBS for 10 min at room temperature and for 10 min with ddH_2O.
2. Incubate the slides in the blocking buffer (50 mM ethanolamine in 0.1 M T Tris/HCl pH 9.0) at room temperature for 60 min (*see* **Note 8**).
3. Rinse slides with ddH_2O.
4. Dilute anti CgA monoclonal antibodies (B4E11 and 5A8) and rabbit polyclonal antibody, obtained as reported in (5), at a concentration of 3 μg/mL in the incubation buffer (0.05 M T Tris/HCl pH 7.6, 0.15 M NaCl, 1% w/V BSA, 0.02% w/V Tween 20) (*see* **Note 9**).
5. Incubate microarrays with the antibody solution in incubation buffer at room temperature for 3 h.
6. Wash microarrays with the washing buffer (0.05 M T Tris/HCl pH 9, 0.25 M NaCl, 0.05% Tween 20) for 10 min at room temperature under stirring.
7. Rinse with ddH_2O.
8. Dry slides by centrifugation at 780 rpm for 3 min.
9. Dissolve the secondary Cy3-labelled antibodies in the incubation buffer (0.05 M T Tris/HCl pH 7.6, 0.15 M NaCl, 1% w/V BSA, 0.02% w/V Tween 20) at a concentration of 0.04 mg/mL.
10. Incubate with secondary antibodies at room temperature for 1 h (*see* **Note 10**). The scheme of the whole experiment in shown in **Fig. 10.1B**.
11. Wash the slides with PBS at room temperature for 10 min under stirring.
12. Wash the slides with ddH2O at room temperature for 10 min under stirring.
13. Repeat step 8.

3.6. Fluorescent Scanning

After binding of the secondary antibody, a fluorescent image of the arrays is acquired by using a confocal laser scanner ScanArray Lite™ (Perkin Elmer Life Sciences) equipped with a 550 nm excitation laser.

1. Scan slides at a resolution of 50 μM and adjust laser power and photomultiplier gain to maximize the dynamic range without getting saturation of the signals.
2. Scan slide with a resolution of 10 μM (*see* **Fig. 10.1C**).

3.7. Data Analysis

The fluorescence intensity of the spots was quantified using the software associated to the scanner. The quantification method chosen was the "fixed circle" method. The spot intensity was calculated as the mean intensity of the pixels located within the spot mask minus the mean intensity of the pixels in the background.

1. Calculate mean value and standard deviation of all signal intensities corresponding to the same peptide.
2. Produce mean signal intensity vs spotted peptide graphs with standard deviation error bars (*see* Fig. **10.2A–C**).

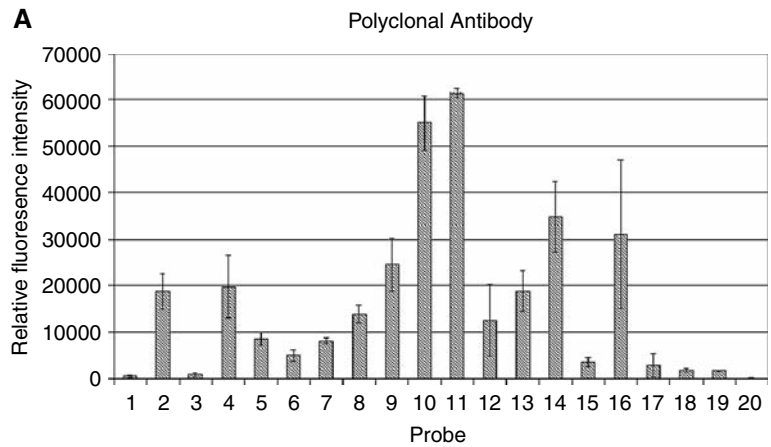


Fig. 10.2. (**A–C)** Plots of fluorescence signals of each probe (as reported in **Table 10.1**) obtained after incubation with anti CgA antibodies. Each signal represents the average fluorescence intensity of 10 replicates. Standard deviations are indicated by error bars. (**A**) Fluorescence signals after incubation with rabbit anti human CgA polyclonal antibody followed by incubation with Cy3 anti Rabbit IgG. The polyclonal antibody was able to bind several peptides to a different extent suggesting the existence of various continuous antigenic sites in CgA. In particular chromogranin residues 25–46, 68–91, 163–187, 275–297, 297–319 (probe 2, 4, 8, 12, 13, respectively) contained weak antigenic sites while residues 222–244, 231–255, 315–337, 374–396 (probe 10, 11, 14, 16, respectively) demonstrated to be strong antigenic sites. (**B**) Fluorescence signals after incubation with anti CgA mouse monoclonal antibody B4E11 followed by incubation with Cy3 anti Mouse IgG. The monoclonal antibody B4E11 strongly recognized the fragment 68–91 (probe 4) with high specificity, in accordance with data obtained by ELISA. (**C**) Fluorescence signals after incubation with anti CgA mouse monoclonal antibody 5A8 followed by hybridization with Cy3 anti Mouse IgG. The monoclonal antibody 5A8 strongly recognized the epitope localized in CgA fragment 47–68 (probe 19) with high specificity, in accordance with data obtained by ELISA. No fluorescence was observed for the unrelated peptide spotted as probe number 20. Reproduced with permission from Chiari M, Cretich M, Corti A, Damin F, Pirri G, Longhi R. (2005). Peptide microarrays for the characterization of antigenic regions of human chromogranin A. *Proteomics* **5**, 3600–3.

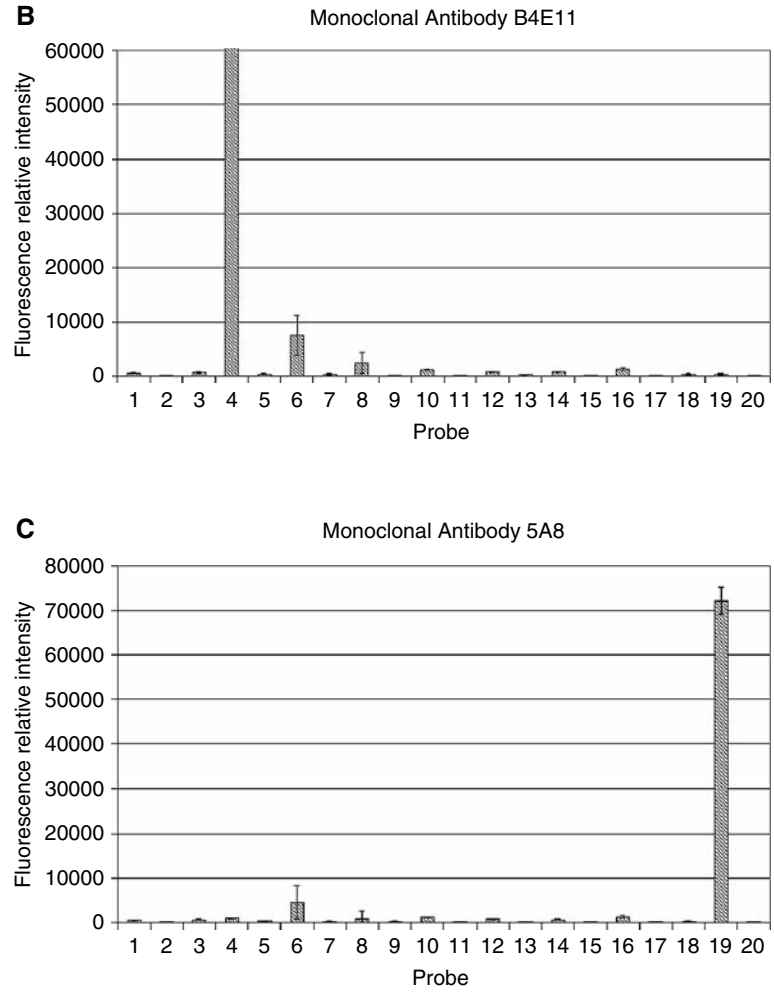


Fig. 10.2. (continued)

4. Notes

1. DMA must be filtered on aluminium oxide to purify the monomer from the stabilizer. THF used for the polymerization has to be distilled under nitrogen, degassed and stored over dry molecular sieves.
2. It is possible to perform the pre-treatment (NaOH and HCl) of the glass slide in glass or plastic chamber indifferently. On the contrary, be sure to use plastic slide chamber for the incubation of the slides in the solution of copoly (DMA-NAS-MAPS) to avoid coating the whole chamber.

3. First prepare a stock solution of ammonium sulphate at a 40% saturation level. To do this, add 242 g of ammonium sulphate to 1 L of ddH_2O. Depending on the volume of the chamber and the number of the slides to be coated, weigh the exact amount of copolymer to obtain a solution 1% w/v, dissolve it in the half volume of ddH_2O, when the copolymer is completely dissolved, and then add the remaining volume of ammonium sulphate at 40%.
4. Glass slides that have been coated with chemical groups are best stored in dry condition in an exsiccator. This is especially important when using the copoly (DMA-NAS-MAPS) because it reacts rapidly with the humidity in the air. The use of copoly (DMA-NAS-MAPS)-coated slides stored in plastic containers at room temperature for 4 weeks gave good results.
5. In our experience peptides can be spotted at variable concentrations (0.1–0.8 mg/mL) depending on their solubility with good results.
6. The number of the replicated spots per peptide is a guideline. Remember a rule from statistics: the more events you have, the more accurate the results.
7. Printed slides were placed in an uncovered storage box, laid in a sealed chamber, saturated with NaCl and incubated at room temperature overnight. To realize a humidity saturated chamber, add as much solid NaCl to water as needed to form a 1 cm deep slurry in the bottom of a plastic container fitted with an airtight lid. In this way, a chamber with a relative humidity of approximately 75% is obtained.
8. The most common strategy to prevent non-specific adsorption in the assay is to treat the microarray slides with BSA (2% w/v BSA in PBS for 1 h to overnight). However, BSA can also mask the spotted peptides and block the interaction with peptides; therefore, a treatment with a non-protein quencher such as ethanolamine is often preferred with peptides arrays.
9. This buffer can be prepared in advance and stored as a stock solution except for BSA that has to be freshly added.
10. An anti Mouse IgG was used for the slides incubated with monoclonal antibodies and an anti Rabbit IgG for the slides incubated with polyclonal antibody. Since the secondary antibodies are fluorescently labelled, protect the slides from light during the incubation and the following steps.

Acknowledgements

The authors gratefully acknowledge the European Commision (TRACEBACK Project n° 036300 and NANOSPAD Project n° 16610) for financial support.

References

1. Helle, K.B., Metz-Boutigue, M.-H., Aunis, D. (2001) Chromogranin A as a calcium-binding precursor for a multitude of regulatory peptides of the immune, endocrine and metabolic systems. *Curr. Med. Chem.: Immunol., Endocr. Metab. Agents* **1**, 119–140.
2. Corti, A., Mannarino, C., Mazza, R., Angelone, T., Longhi, R., Tota, B. (2004) Chromogranin A N-terminal fragments vasostatin-1 and the synthetic CGA 7–57 peptide act as cardiostatins on the isolated working frog heart. *Gen. Comp. Endocrinol.* **136**, 217–224.
3. Feldman, S.A., Eiden, L.E. (2003) The chromogranins: their roles in secretion from neuroendocrine cells and as markers for neuroendocrine neoplasia. *Endocr. Pathol.* **14**, 3–23.
4. Orr, F. O., Chen, T., Johnsen, A.H., Chalk, R., Buchanan, K.D., Sloan, J.M., Pingfan, R., Shaw, C. (2002) The spectrum of endogenous human chromogranin A-derived peptides identified using a modified proteomic strategy. *Proteomics* **2**, 1586–1600.
5. Corti, A., Longhi, R., Gasparri, A., Chen, F., Pelagi, M., Siccardi, A.G. (1996) Antigenic regions of human chromogranin A and their topographic relationships with structural/functiona1 domains. *Eur J Biochem.* **235**, 275–280.
6. Ratti, S., Curnis, F., Longhi, R., Colombo, B., Gasparri, A., Magni, F., Manera, E., Metz-Boutigues, M.H., Corti, A. (2000) Structure–activity relationships of chromogranin A in cell adhesion. *J. Biol. Chem.*, **275**, 29257–29263.
7. Pirri, G., Damin, F., Chiari, M., Bontempo, E., Depero, L.E. (2004) Characterization of a polymeric adsorbed coating for DNA microarray glass slides. *Anal. Chem 76*, 1352–1358.
8. Seong, S.Y. and Choi, C.Y. (2003) Current status of protein chip development in terms of fabrication and application. *Proteomics*, **3**, 2176–2189.
9. Lasaicherre, M. L., Uttamchandani, M., Chen, G. Y. J., Yao, S. Q. (2002) Developing site-specific immobilization strategies of peptides in a microarray. *Bioorg. Med. Chem. Lett.* **12**, 2079–2083.
10. Falsey, J.R., Renil, R., Park, S., Li, S., Lam, K S. (2001) Peptide and small molecule microarray for high throughput cell adhesion and functional assays. *Bioconjugate Chem.* **12**, 346–353.
11. Houseman, B., T., Huh, J.H., Kron, S.J., Mrksich, M. (2002) Peptide chips for the evaluation of protein kinase activity. *Nature Biotech.* **20**, 270–274.
12. Frank, R. (1992) Spot-synthesis: an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. *Tetrahedron* **48**, 9217–9232.
13. Cretich, M., Pirri, G., Damin, F., Solinas, I., Chiari, M. (2004) A new polymeric coating for protein microarrays. *Anal. Biochem.* **332**, 67–74.

Chapter 11

Antimicrobial Peptide Arrays for Detection of Inactivated Biothreat Agents

Chris R. Taitt, Stella H. North, and Nadezhda V. Kulagina

Abstract

Arrays of immobilized antimicrobial peptides are used to detect bacterial, viral, and rickettsial pathogens, including inactivated biothreat agents. These arrays differ from the many combinatorial peptide arrays described in the literature in that the peptides used here have naturally evolved to interact with and disrupt microbial membranes with high affinity but broad specificity. The interaction of these naturally occurring peptides with membranes of pathogens has been harnessed for the purpose of detection, with immobilized antimicrobial peptides acting as "capture" molecules in detection assays. Methods are presented for immobilizing the antimicrobial peptides in planar arrays, performing direct and sandwich assays, and detecting bound targets.

Key words: Biothreat, detection, array, antimicrobial peptide.

1. Introduction

Planar arrays have become an important tool for bioanalytical sensing and detection. A wide range of arrays are commercially available with immobilized recognition molecules as varied as oligonucleotides (Affymetrix, Agilent Technologies, Combimatrix), peptides and proteins (Invitrogen, LC Sciences, New England Peptide), oligo- and polysaccharides (QIAGEN, Glycominds), aptamers (LC Sciences), antibodies (Clontech, Thermo Scientific, and Sigma-Aldrich, BD Biosciences, Schleicher & Schuel, Xyomyx), and various cells/tissues (3H Biomedical AB). In addition to commercial microarrays, researchers can create custom arrays in their own laboratories using techniques such as ink-jet printing, photoimmobilization, and physically isolated patterning. Our laboratory has been producing planar "macro" arrays

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_11,

for over a decade, with the goal of detecting various pathogens, toxins, and physiological markers in rapid biosensor assays (1–5). A subset of these latter arrays comprises those using antimicrobial peptides.

The concept of a peptide-based array is not new. A variety of papers have described the use of peptide arrays for proteomics (6, 7), determination of enzymatic specificity (8–10), therapeutics (7, 11, 12), and screening of ligand-binding interactions (13, 14). Arrays using antimicrobial peptides for recognition differ from most of these other multiplexed peptide-based arrays in the following aspects:

- Antimicrobial peptides are not combinatorially derived or randomly generated but, rather, are found in nature and possess well-documented binding interactions with microbial surfaces;
- The immobilized antimicrobial peptides have defined secondary structures, unlike randomly generated libraries of peptides, aptamers, sugars, etc.;
- The mechanism for detection of the target species is affinity-based, rather than enzymatically based as with arrays using peptide substrates (e.g., arrays for determining kinase specificity (8)); and
- Antimicrobial peptides used in the arrays have overlapping specificities. Therefore, they may be used to detect a larger number and broader variety of microbial targets than more specific recognition species.

Antimicrobial peptides have evolved naturally as part of the "innate" immune system designed to protect the host from microbial invasion; most perform this function by binding to and disrupting the target cells' membranes (15–17), although other modes of killing action have been described (18, 19). Given their natural affinity for microbial membranes, we and others have demonstrated that immobilized antimicrobial peptides can be used for detection of microbial targets and their products (20–25). The ability to detect these targets is presumably mediated through the native interactions between the peptides and invariant components of cell surfaces. Our own work has emphasized development of rapid detection assays for inactivated bacterial, viral, and rickettsial biothreats using these peptides. Here we present methods and materials for preparation and use of antimicrobial peptide-based planar arrays.

Immobilization of peptides onto planar arrays will first be described (**Fig. 11.1**). This procedure consists of activating the surface with silane, followed by treatment with a crosslinker; the crosslinker used possesses a reactive group (*N*-hydroxysuccinimidyl ester, NHS) that is specific for amine moieties and is used to attach the peptides to the surface (**Fig. 11.1**, *left panel*). Patterns of peptides are created on the activated slides through use of a

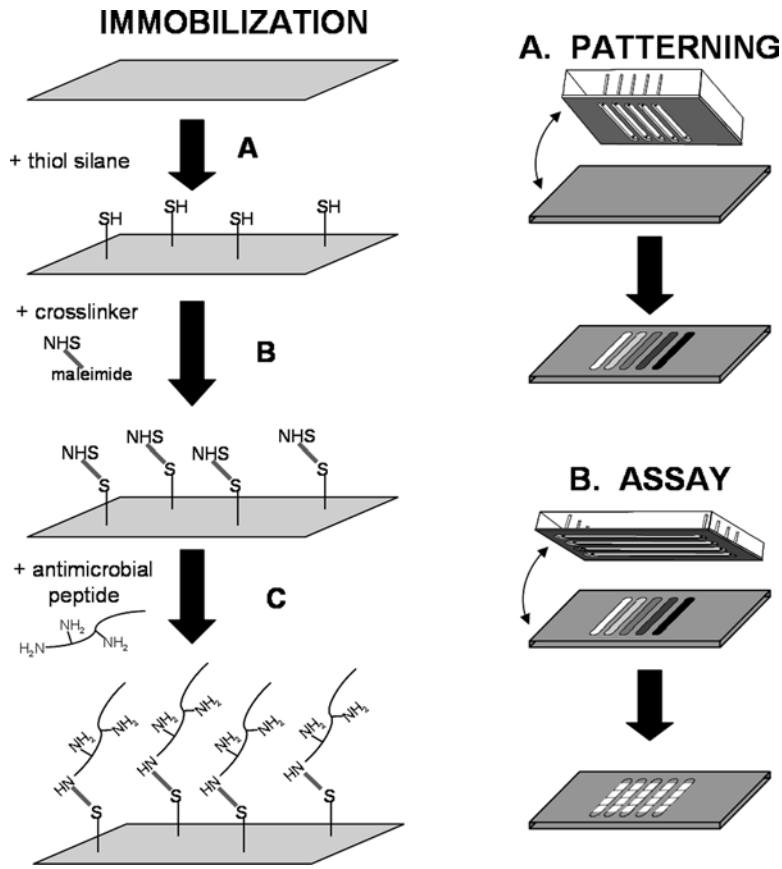


Fig. 11.1. Preparation of antimicrobial peptide arrays. *Left schematic*: overview of the immobilization chemistry. Cleaned slides are treated with a thiol-terminated silane (**A**), leaving the slide endowed with a thiol-decorated surface. Surface thiols are subsequently reacted with the maleimide moiety of GMBS (**B**), leaving the surfaces decorated with NHS moieties. The pendant NHS moieties react with primary amines on the peptides, resulting in direct covalent immobilization of the antimicrobial peptides on the slide surface. Although only the N-terminal amine is shown as the point of immobilization, any primary amine on the peptide can react with the surface-attached NHS groups. *Right schematic*: formation of the array patterns. (**A**) Patterning. Antimicrobial peptides are immobilized on the slide in a series of stripes using patterning templates whose channels are oriented across the short axis of the slide. (**B**) Assay. Patterned slides are placed in contact with assay templates whose channels are oriented orthogonal to the stripes of immobilized peptides. When flowed across the slide, each sample is exposed to each stripe of immobilized peptide, resulting in a cross-hatch pattern.

patterning template molded to contain multiple channels that, when filled with solutions of peptides, result in stripes of immobilized peptides on the surface of the slide (**Fig. 11.1**, *upper right*). Once patterned with these stripes, slides can be used for direct and sandwich assays to detect various categories of bacteria, rickettsiae, viruses, and even toxins. For both types of assays, an assay template is used to flow sample solutions over the patterned arrays (**Fig. 11.1**, *lower right*). The orthogonal arrangement of the assay channels to the patterned stripes allows each sample to encounter each stripe of immobilized peptide, where microbial

components may bind. In sandwich format assays, a fluorescent tracer solution is then used to detect bound targets. Protocols for both types of assays are presented, as well as a subset of results obtained using inactivated biothreat agents.

2. Materials

2.1. Preparation of Patterning and Assay Templates

1. Molds for poly(dimethyl) siloxane (PDMS): reverse molds for PDMS patterning and assay templates can be milled to the user's specifications (26) and are illustrated schematically in **Fig. 11.2**. For the six-channel patterning templates used in our laboratory, for example, molds are typically milled with six pillars of dimensions of 22 mm (L) × 1 mm (W) × 2.5 mm (H), separated by 2 mm; these pillars are oriented across the width of the mold. For six-channel assay templates, molds are milled to final dimensions of 40 mm (L) × 1 mm (W) × 2.5 mm (H), separated by 2 mm; these pillars are oriented along the length of the mold. Analogous molds can be made for patterning and assay templates containing up to 12 (patterning) or 15 (assay) channels; dimensions of the channels are given in (27) (for **Section 3.1**).
2. PDMS: PDMS is made from a two-component silicone system that, once mixed, forms a slurry that is cast and cured within the molds (Nusil Silicone Technology, Carpinteria, CA) (26).

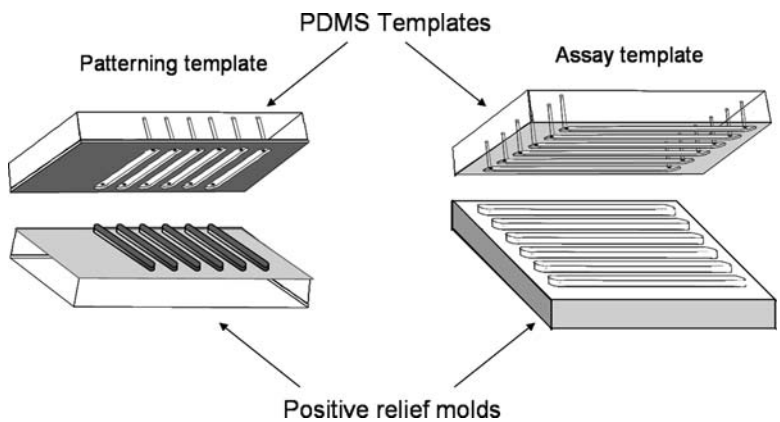


Fig. 11.2. Schematic of the molds used to make PDMS patterning templates (*left*) and assay templates (*right*). Each mold is milled in plastic (most commonly, Plexiglas or poly(methyl) methacrylate) as a positive relief of the channels desired. After the components of the PDMS are mixed, the mixture is poured into the mold and allowed to cure.

2.2. Cleaning of Slides

1. Coplin jar: eight-slot Coplin jar to process 16 slides (for **Section 3.2**).
2. Microscope slides: standard clear (unfrosted) microscope slides can be purchased from any scientific supply source. For real-time, on-line kinetic analyses using the NRL Array Biosensor, slides coated with a patterned silver "cladding" are purchased from Opticoat Associates, Protected Silver (Chelmsford, MA) (*see* **Note 1**).
3. KOH/methanol: dissolve 10 g KOH (Sigma-Aldrich, St. Louis, MO) in 100 mL methanol. This solution is corrosive and should be handled with care (*see* **Note 2**).

2.3. Activation of the Glass Slide Surface (for Section 3.3)

1. Cleaned slides (from **Section 3.2**).
2. Glove bag.
3. Silane–toluene mix: mix 1 mL 3-mercaptopropyl triethoxy silane (Fluka, Sigma-Aldrich) with 49 mL anhydrous toluene (Sigma-Aldrich) (*see* **Note 3**). Prepare this solution immediately before use in a glove bag filled with nitrogen. To prevent polymerization of the silane, 1–2 mL aliquots should be made when the bottle is first opened. These aliquots should be stored in a desiccator at 4°C; each aliquot is warmed to room temperature before opening and is used once, with any remaining silane disposed of as hazardous waste (*see* **Note 4**).
4. Toluene: anhydrous toluene poured into three 150 mL beakers (*see* **Note 4**).
5. GMBS/ethanol: dissolve 12.5 mg *N*-[γ-maleimidobutyryloxy]succinimide ester (GMBS) crosslinker (Pierce, Rockland, IL) in 0.25 mL anhydrous dimethyl sulfoxide (DMSO, Sigma-Aldrich), then add to 43 mL absolute ethanol, mixing well. Use immediately (*see* **Note 5**).

2.4. Immobilization of Antimicrobial Peptides (for Section 3.4)

1. PDMS assay templates (from **Section 3.1**).
2. PLEXIGLAS supports for the slide and PDMS templates: supports are manufactured from PLEXIGLAS or other clear plastic for reproducible placement of the patterning and assay templates on the slides. The lower piece is milled with a single groove to hold the slide in place. Upper pieces are milled to possess two openings to allow insertion of syringe needles (inlets, outlets) into the PDMS templates.
3. Activated slides (from **Section 3.3**).
4. Phosphate-buffered saline (PBS), pH 7.3.

5. Antimicrobial peptides: antimicrobial peptides can be purchased from Anaspec, Sigma, American Peptides or several other commercial sources; amino acid sequences of these peptides are shown in **Table 11.1**. Dissolve the peptides in PBS at final concentrations of 1 mg/mL (most peptides, except as noted), 10 mg/mL (polymyxin B, polymyxin E), 50–250 μg/mL (bactenecin, lactoferricin). Hydrophobic peptides such as alamethicin are dissolved in methanol at 1 mg/mL final concentration. When performing assays on silver-cladded slides, 10 mM phosphate buffer pH 7.4 must be used in place of PBS (26, 28) (*see* **Note 6**).
6. Syringes: syringe barrels (without plungers) will be used as outlets for each channel in the patterning template; syringes equipped with needles and plungers will be used to inject solutions into the channels.
7. Antibodies (controls): antibodies directed against biothreat agents can be obtained from any number of sources, including the Department of Defense Critical Reagent Program (CRP; Edgewood, MD). For an additional positive control, we utilize rabbit anti-chicken IgY

Table 11.1
Amino acid sequences of antimicrobial peptides used in this study

Alamethicin	Ac–UPUAUAQUVUGLUPVUUEQF–OH	U= methylalanine
Bactenecin	RLCRIVVIRVCR	(intramolecular disulfides)
Cecropin A	KWKLFKKIEKVGQNIRDGIIKAGPAVAVVGQATQIAK–$CONH_2$	
Cecropin B	KWKVFKKIEKMGRNIRNGIVKAGPAIAVLGEAKAL	
Cecropin P1	SWLSKTAKKLENSAKKRISEGIAIAIQGGPR	
CecA/Mel hybrid	KWKLFKKIGIGAVLKVLTTG–$CONH_2$	
Lactoferricin	RRWQWRMKKLG	
Magainin I	GIGKFLHSAGKFGKAFVGEIMKS	
Melittin	GIGAVLKVLTTGLPALISWIKRKRQQ–$CONH_2$	
Parasin	KGRGKQGGKVRAKAKTRSS	
Polymyxin B	fa-BTBBBFLBBT	fa=fatty acid; B= diaminobutyrate
Polymyxin B nonapeptide	TBBBFLBBT	B= diaminobutyrate
Polymyxin E (colistin)	fa-BTBBBLLBBT	fa=fatty acid; B= diaminobutyrate

from Jackson Immunoresearch (West Grove, PA). Control capture antibodies are diluted in PBS to a final concentration of 10 μg/mL.

8. PBSTB: PBS supplemented with 0.05% Tween-20 and 1 mg/mL bovine serum albumin (BSA).
9. Blocking buffer: 1 g gelatin (type B, 75 bloom; Sigma-Aldrich) in 100 mL PBS. Let it soak for ~1 h, then heat, with stirring, on a stirplate set to 100°C. After the gelatin is fully dissolved, cool and store at 4°C; remove and warm gently before using.

2.5. Fluorescent Labeling

2.5.1. Labeling of Cells, Viruses (for Section 3.5.1)

1. Borate buffer: 50 mM borate buffer, pH 8.5.
2. Target analytes: inactivated biothreat agents (*Francisella tularensis* Schu4 and LVS strains, *Bacillus anthracis* Ames and Sterne strains, *Yersinia pestis* CO-92, Venezuelan equine encephalitis, vaccinia virus) were obtained from CRP. *B. globigii* spores were obtained from Dugway Proving Ground (Dugway, UT) and *B. thuringiensis* (Kurstaki strain) spores were obtained as Dipel insecticidal powder (Abbott Laboratories, North Chicago, IL). Targets are diluted in borate buffer to 10^8 cells/mL (bacteria, rickettsiae) or 10^9 plaque-forming units (pfu)/mL (viruses).
3. Cy3 bisfunctional *N*-hydroxysuccinimidyl ester (GE Healthcare Life Sciences, Piscataway, NJ).
4. DMSO, anhydrous.
5. Dialysis tubing: dialysis tubing with an appropriate molecular weight cutoff (>1000 Da) can be obtained from many sources including Fisher, VWR, and Sigma-Aldrich.
6. PBS.

2.5.2. Labeling of Antibodies (for Section 3.5.2)

1. Borate buffer: 50 mM borate buffer, pH 8.5.
2. Antibodies: antibodies directed against biothreat agents and other pathogens are commercially available. CRP served as the antibody source for our own studies. Dilute antibodies (1 mg) in borate buffer to an approximate final concentration of 2 mg/mL, ensuring that the final buffer concentration remains at approx. 50 mM (this can be done by using a 10 × buffer stock if necessary) (*see* **Note 7**).
3. Cy3 bisfunctional *N*-hydroxysuccinimidyl ester. Each vial should be sufficient to label 3 mg antibody.
4. DMSO, anhydrous.
5. PBS.
6. Sizing column: chromatography column (dimensions

approx. 1.5 cm diameter × 11–15 cm height), filled with approx. 20–25 mL swollen Bio-Gel P-10 gel (Bio-Rad, Hercules, CA). Pre-equilibrate with PBS.

2.6. Assays

2.6.1. Direct Assays (for Section 3.6.1)

1. PDMS assay templates (from **Section 3.1**).
2. Slides patterned with stripes of immobilized antimicrobial peptides (from **Section 3.4**). For kinetic studies using the NRL Array Biosensor, silver-cladded slides must be used.
3. PBSTB.
4. Cy3-labeled cells, viruses (from **Section 3.5.1**): 10^0–10^8 cells/mL (bacterial, rickettsial cells) or 10^0–10^9 pfu/mL (viruses) diluted in PBSTB. A positive control, Cy3-labeled chicken IgY (Jackson Immunoresearch), was added to each sample to a final concentration of 100 ng/mL.

2.6.2. Sandwich Assays (for Section 3.6.2)

1. PDMS assay templates (from **Section 3.1**).
2. Microscope slides patterned with stripes of immobilized antimicrobial peptides (from **Section 3.4**).
3. PBSTB.
4. Targets: inactivated biothreat agents were obtained from CRP. Dilute targets in PBSTB to concentrations ranging from 10^0 to 10^8 cells/mL (bacteria, rickettsiae) or 10^0–10^9 pfu/mL (viruses).
5. Cy3-labeled "tracer" antibodies (from **Section 3.5.2**): dilute Cy3-labeled tracer antibodies in PBSTB to a final concentration of 10 μg/mL. Antibodies directed against the inactivated biothreat agents were also obtained from the Critical Reagents Program. A positive control, Cy3-labeled chicken IgY (Jackson Immunoresearch), was added to the tracer to a final concentration of 100 ng/mL.

3. Methods

3.1. Preparation of Patterning and Assay Template

The use of different patterning and assay templates – each with channels oriented orthogonal to the other – is essential to the creation of patterned arrays (**Fig. 11.1**, *right schematic*). Patterning and assay templates can be prepared well ahead of time and reused many times over. These templates are created using a reverse mold milled in PLEXIGLAS (**Fig. 11.2**). Upon mixing of a two-component PDMS kit, the slurry is poured into the mold and left to cure for 1–3 days. The patterning (or assay) template is then removed from the mold and is treated with a blocking solution to render the surfaces less hydrophobic and "sticky."

1. In a large disposable plastic beaker, combine part A and part B of the PDMS kit at a 10:1 ratio (26). Mix well with a disposable spatula. The mixture should appear opalescent after mixing.
2. Degas the mixture in vacuum at room temperature until most bubbles have been removed. This first degassing often takes 30 min or longer (*see* **Note 8**).
3. Pour the degassed PDMS mix into the PLEXIGLAS mold, avoiding bubbles whenever possible.
4. Place the PLEXIGLAS mold with PDMS mix into the vacuum and degas until all bubbles are at the surface of the PDMS. This step may take as long as 60 min or as short as only a few minutes.
5. Remove the PLEXIGLAS mold with degassed mix from the vacuum and cure at room temperature for 1–3 days. It may likewise be cured for 30 min at 60°C.
6. After curing, remove the PDMS from the mold and trim the edges.
7. Wash with soap and water.
8. Place the template in a large beaker and cover with a solution of 2% BSA in PBS.
9. Incubate the template in the BSA solution for >2 h, rinse briefly with distilled water, and dry. The PDMS template is now ready to use (*see* **Note 9**).

3.2. Cleaning of Slides

It is imperative that the slide surfaces be kept scrupulously clean for optimal attachment of the peptides.

1. Etch an appropriate marking on all slides using a carbide-tipped pencil. Typically, the upper right corner is marked with a number for later identification and orientation of the arrays.
2. Rinse each slide under a stream of distilled water, rubbing both sides to remove particulates.
3. Dry each slide under an air stream.
4. Place slides back-to-back (etched sides facing out) and insert into an upright Coplin jar.
5. Pour KOH/methanol solution over the slides and incubate 1 h.
6. Remove KOH/methanol solution from Coplin jar and discard into an appropriate waste container.
7. Rinse each slide individually with deionized water, separating the pairs of slides. Ensure that all KOH/methanol solution is removed from the slides by rinsing both sides and all edges of each slide. Cleanliness is indicated by the absence of schlieren lines and sheeting of the rinse water on the slide surface.

8. Dry each slide individually under a stream of air (or nitrogen) (*see* **Note 10**). Ensure that no water remains on the edges of the slides.

3.3. Surface "Activation"

"Activation" of the slide surface to display specific reactive moieties is a two-step process. The first step, silanization (**Section 3.3.1**), deposits a layer of thiol-terminated silane on the slide surface. The second step, attachment of the crosslinker (**Section 3.3.2**), involves reaction of the linker's (thiol-reactive) maleimide group with the thiol-decorated surface, leaving a pendant *N*-hydroxysuccinimyl ester for later binding to the peptides (*see* **Note 11**). A schematic of the overall attachment chemistry for immobilizing the antimicrobial peptides is shown in **Fig. 11.1**.

3.3.1. Silanization

1. Place cleaned slides (from **Section 3.2**) back-to-back and insert into a clean, dry Coplin jar.
2. Place the Coplin jar inside a nitrogen-filled glove bag within a chemical hood and insert all necessary glassware and reagents (glass pipette, flask for mixing, glass-graduated cylinder, silane, toluene) before closing.
3. Prepare the silane–toluene mix within the glove bag and immediately pour over the slides (*see* **Note 4**).
4. Replace the lid on the Coplin jar and incubate for 1 h in the glove bag.
5. Remove the Coplin jar from the glove bag and place elsewhere in the hood.
6. Using forceps to handle the slides, rinse each slide in toluene three times by swishing the slide 3–5 times sequentially in three 150 mL beakers filled with fresh toluene. Great care should be taken to avoid exposure to toluene and its vapors (*see* **Notes 4** and **12**).
7. Dry each slide under a stream of nitrogen or air. Ensure that no solvent remains on the sides of each slide.
8. Place slides back-to-back within a clean, dry Coplin jar and cover. Silanized slides may be kept dried under nitrogen for up to 1 week.

3.3.2. Attachment of Crosslinker

1. Immediately after preparation of GMBS/ethanol solution, pour over dried, silanized slides in the Coplin jar (from **Section 3.3.1**).
2. Replace the lid on the Coplin jar and incubate for 30 min (room temperature) on the benchtop.
3. Rinse each slide separately in three beakers of deionized water as described for the toluene washes.
4. Immediately after rinsing, dry each slide under a stream of nitrogen or air. Proceed immediately with patterning.

3.4. Immobilization of Antimicrobial Peptides

After the slide surface has been activated with silane and the GMBS crosslinker, pendant *N*-hydroxysuccinimidyl active groups react with amines present on the peptides. After the PDMS patterning template is placed atop the GMBS-treated slide, each channel in the template can be filled with a different peptide or antibody solution. Arrays of immobilized peptides are created as a series of stripes oriented across the short axis (3 cm) of the GMBS-treated slide (**Fig. 11.1**, *upper right schematic*).

1. Place the dried GMBS-treated slide (**Section 3.3.2**) into the channel in the PLEXIGLAS support.
2. Place the dried, blocked PDMS patterning template (**Section 3.1**) atop the slide, such that the molded channels are in contact with the slide and run across the width of the slide. Ensure that the ends of each channel do not overlap the slide's edges.
3. Place the upper PLEXIGLAS support atop the PDMS template, such that the ends of each channel are accessible through the windows. Tighten into place.
4. Fill a 1 mL syringe with patterning solution and insert the needle into one end of the appropriate channel. At the other end of the same channel, insert the needle of an open syringe barrel; this syringe barrel serves as an outlet.
5. Inject the patterning solution into the channel, watching the meniscus as the solution fills the channel.
6. Repeat injection of patterning solution(s) until all channels have been filled.
7. Remove all syringe barrels and incubate overnight at 4°C.
8. The following day, connect one end of each channel to a multi-channel peristaltic pump using a needle attached to the end of the pump tubing.
9. To the opposite end of each channel, insert the needle of an open syringe barrel. This syringe barrel will serve as an inlet and reservoir for wash solution.
10. Start the pump flowing at 1 mL/min flow rate, allowing the solution to be sucked out of all channels.
11. After all channels have been evacuated, fill the syringe barrel reservoirs with 1 mL PBSTB and allow the pump to pull this solution through all channels.
12. After all channels have been rinsed and evacuated, stop the pump, remove all needles and disassemble the patterning assembly.
13. Place the patterned slide into blocking buffer for 1 h.
14. Rinse briefly, and dry under a stream of nitrogen or air. Slides patterned with peptides can be stored at 4°C for 2 weeks.

3.5. Fluorescent Labeling

Depending on the type of assay format used – direct or sandwich – one component of the assay must be labeled with an appropriate dye for use in a fluorobiosensor such as the NRL Array Biosensor or a confocal scanner. For direct assays, the target itself is labeled and binding to surface-immobilized peptides is measured directly. For sandwich assays, an unlabeled target cell or virus is first captured onto the surface and is subsequently detected using a fluorescent "tracer" antibody.

3.5.1. Labeling of Cells, Viral Particles

Cells and viral particles are labeled using an NHS–ester derivative of Cy3; other NHS–ester dyes can be used with other detection platforms. These derivatives label surface amines on the cells or viral targets.

1. Dissolve the packet of Cy3 NHS–ester in 25 μL DMSO.
2. Add the entire vial of dissolved dye to the diluted cells or viruses (10^8 cells/mL bacteria or 10^9 pfu/mL virus in borate buffer).
3. Incubate 30–60 min at room temperature with agitation.
4. Place labeled cells/viruses into dialysis tubing and dialyze overnight at 4°C with at least three changes of buffer.
5. Store the labeled cells/viruses in the dark at 4°C until used.

3.5.2. Labeling of Antibodies

Antibodies are labeled with dye in the same manner as cells and viruses. However, unincorporated dye is more easily (and more rapidly) removed by gel filtration on a sizing column.

1. Dissolve the packet of Cy3 NHS–ester in 25 μL DMSO.
2. Add approx. 8 μL dissolved dye to 1 mg of antibody (diluted to 2 mg/mL in borate buffer).
3. Incubate for 35 min in the dark at room temperature.
4. During this incubation, prepare the sizing column by rinsing with at least one column volume of PBS. Ensure that the buffer is drained from the gel bed before the incubation is complete.
5. Pipette the entire volume of dye/antibody onto the top of the sizing column, ensuring that the top of the gel bed is not disturbed.
6. Allow the dye/antibody to enter the gel and wash the sides of the column and top of the gel bed with a small (500 μL) aliquot of PBS. Allow the PBS to drain into the gel bed.
7. Pipette PBS onto the top of the column, filling the reservoir. Allow buffer to run into the column, filling the reservoir as necessary.
8. As the PBS is running through the column, watch the dye front as it proceeds down the column. The dye should separate out into two distinct bands. The first colored band consists of labeled protein (desired product), whereas the second colored band consists of unincorporated dye.

9. Collect the first colored band as a single fraction. The second colored band can be collected as waste and discarded.
10. Determine the antibody concentration and extent of labeling by UV/Vis spectroscopy. Typically, absorbances are taken at 280 and 553 nm and concentrations of antibody and incorporated dye determined using equations provided by the manufacturer.
11. Store the labeled antibodies in the dark at 4°C until use.

3.6. Assays

Two assay formats are most commonly used for detection of cells, viruses, and relatively large molecules (at least several kDa in molecular weight): direct and sandwich. "Direct" assays are most commonly used to demonstrate initial proof-of-concept that a binding interaction occurs. However, since most targets are not naturally fluorescent, a "sandwich" format is generally used by researchers who do not wish to process their samples before analysis. Each assay type utilizes a PDMS assay template that has been molded to possess channels oriented to run along the length of the slide (perpendicular to the stripes of immobilized peptide) (**Fig. 11.1**, *lower right schematic*; **Fig. 11.3**).

3.6.1. Direct Assays

This assay format utilizes naturally fluorescent or fluorescently labeled targets and directly measures binding of the target to the immobilized peptide (or other capture species). This format can also be used to determine kinetic parameters of the binding

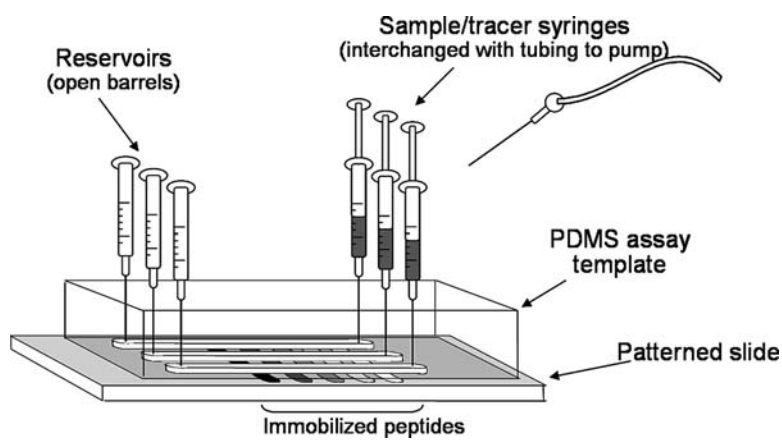


Fig. 11.3. Schematic of the slide setup for assays. The patterned slide is placed underneath the assay template, such that the channels run perpendicular to the stripes of immobilized peptides. Syringe barrel reservoirs are placed at the end of the channels, whereas samples or tracers are injected from the opposite ends using syringes with plungers. For all steps requiring flow, the syringe/plungers are replaced with tubing to a peristaltic pump that is used to pull solutions from the reservoirs over the slide. An analogous setup is used for patterning, although the channels are oriented across the short axis of the slide.

interaction (25, 29), but the researcher must be cautioned that any modification of the target analyte may affect binding characteristics (25, 29).

1. Place the patterned slide (**Section 3.4**) into the channel in the PLEXIGLAS assay support.
2. Place the PDMS assay template (**Section 3.1**) atop the slide, such that the molded channels are in contact with the slide and run along the length of the slide. Ensure that the ends of each channel do not overlap the slide's edges.
3. Place the upper PLEXIGLAS support atop the PDMS template, such that the ends of each channel are accessible through the windows. Tighten into place.
4. At one end of each channel, insert the needle of an open syringe barrel; this syringe barrel serves as a reservoir.
5. Hook the opposite end of each channel to a multi-head peristaltic pump (**Fig. 11.3**).
6. Add 1 mL PBSTB to each of the syringe barrel reservoirs and start the pump flowing at 0.8 mL/min flow rate. Rinse the buffer through each channel and allow the pump to evacuate all channels.
7. Disconnect the slide from the pump, keeping the syringe barrel reservoirs in place.
8. Inject sample solution into the appropriate channel, watching the meniscus as the solution fills the channel. Typically, the buffer blank is injected first to ensure that there are no leaks.
9. Repeat sample injections until all channels have been filled.
10. Incubate 1 h at room temperature.
11. After 1 h, connect the opposite end of each channel to a multi-channel peristaltic pump using a needle attached to the end of the pump tubing.
12. Start the pump flowing at 0.3 mL/min flow rate, allowing the sample solution to be sucked out of all channels (*see* **Note 13**).
13. After all channels have been evacuated, fill the syringe barrel reservoirs with 1 mL PBSTB and allow the pump to pull this solution through all channels.
14. After all channels have been rinsed and evacuated, stop the pump, remove all needles, and disassemble the assay assembly.
15. Rinse the slide briefly with deionized water to remove salts, and dry under a stream of nitrogen or air.
16. Keep slide dark until imaging.

3.6.2. Sandwich Assays

Sandwich assays are performed essentially in two steps, separated typically by one or more washes. The target analyte is first bound to the immobilized peptide. After the non-bound sample

components are washed away, the fluorescent "tracer" antibody is then incubated with the capture surface to measure the amount of bound target.

1. Place the patterned slide (**Section 3.4**) into the channel in the PLEXIGLAS assay support.
2. Place the PDMS assay template (**Section 3.1**) atop the slide, such that the molded channels are in contact with the slide and run across the length of the slide. Ensure that the ends of each channel do not overlap the slide's edges.
3. Place the upper Plexiglas support atop the PDMS template, such that the ends of each channel are accessible through the windows. Tighten into place.
4. At one end of each channel, insert the needle of an open syringe barrel; this syringe barrel serves as a reservoir (**Fig. 11.3**).
5. Inject sample solution into the appropriate channel, watching the meniscus as the solution fills the channel. Typically, the buffer blank is injected first to ensure that there are no leaks.
6. Repeat sample injections until all channels have been filled.
7. Incubate 1 h at room temperature.
8. After 1 h, connect the opposite end of each channel to a multi-channel peristaltic pump using a needle attached to the end of the pump tubing.
9. Start the pump flowing at 1 mL/min flow rate, allowing the sample solution to be sucked out of all channels (*see* **Note 13**).
10. After all channels have been evacuated, fill the syringe barrel reservoirs with 1 mL PBSTB and allow the pump to pull this solution through all channels.
11. After all channels have been rinsed and evacuated, stop the pump and disconnect the slide assembly from the pump. Keep the syringe barrel reservoirs in place.
12. Inject fluorescent tracer antibody into all lanes. Incubate 30 min in the dark at room temperature.
13. Connect one end of each channel to the pump again and start the pump flowing at 0.3 mL/min flow rate (*see* **Note 14**).
14. After the tracer antibody has been flowed through and each channel is filled with air, add 1 mL PBSTB to each syringe barrel reservoir.
15 Increase the flow rate to 1 mL/min and rinse each channel with 1 mL PBSTB as above.
16. Stop the pump, remove all needles, and disassemble the assay assembly.

17. Rinse the slide briefly with deionized water to remove salts, and dry under a stream of nitrogen or air.
18. Keep slide dark until imaging.

3.7. Data Collection and Analysis

Data can be collected using a standard confocal microarray scanner or alternatively a rapid analysis system such as the NRL Array Biosensor (1–5) if real-time binding analyses are desired. Each system is equipped with its own imaging and data extraction software. The researcher is directed to the appropriate resources for this information and is cautioned that each system may have limitations on the specific fluorophores detected.

3.8. Examples of Bacteria, Viruses Detected

Immobilized antimicrobial peptides can be used for rapid (<120 min) detection of Gram-positive and Gram-negative bacteria, rickettsiae, viruses, and surprisingly, toxins. The data shown below demonstrate the potential for these arrays to detect various targets considered as select agents. In each case shown, an inactivated target was used to demonstrate proof-of-concept; final confirmation with live, active agents will be required before these tests can be transitioned to the field.

3.8.1. Detection of Gram-Negative Threat Agents

Radiation-inactivated cell preparations of *Y. pestis* CO-92 and *F. tularensis* were obtained from CRP and were confirmed non-viable prior to shipment. These samples were diluted in PBSTB to final concentrations ranging from 10^0 to 10^8 cells/mL and assayed using a sandwich format using Cy3-labeled rabbit anti-*Y. pestis* and goat anti-*F. tularensis* tracer antibodies, respectively (*see* **Note 15**). Assayed slides were imaged using a Packard ScanArray Lite confocal microarray scanner (Packard Biochip Technologies, Billerica, MA) and fluorescence data were extracted from the images using the associated microarray analysis program (QuantArray). **Figures 11.4** and **11.5** show representative images from assays for *Y. pestis* and *F. tularensis*, respectively, (left panels) and dose–response curves from data extracted from these images (right panels).

3.8.2. Detection of Gram-Positive Threat Agents

Radiation-inactivated cell preparations of *B. anthracis* Ames and Sterne strains were obtained from CRP and were confirmed non-viable prior to shipment. Samples were diluted in PBSTB to final concentrations ranging from 10^0 to 10^8 cells/mL and assayed using a sandwich format using Cy3-labeled rabbit anti-*B. anthracis* antibodies. Assayed slides were imaged using a Packard ScanArray Lite and data extracted as indicated above (**Fig. 11.6**). Sterne strain cells were readily detectable using antibody as an immobilized capture molecule, but these cells did not bind to any of the peptides tested. Ames strain cells, on the other hand, bound readily to immobilized peptides cecropin P1 and parasin. Dose–response curves could be determined for *B. anthracis* Ames binding to immobilized peptides

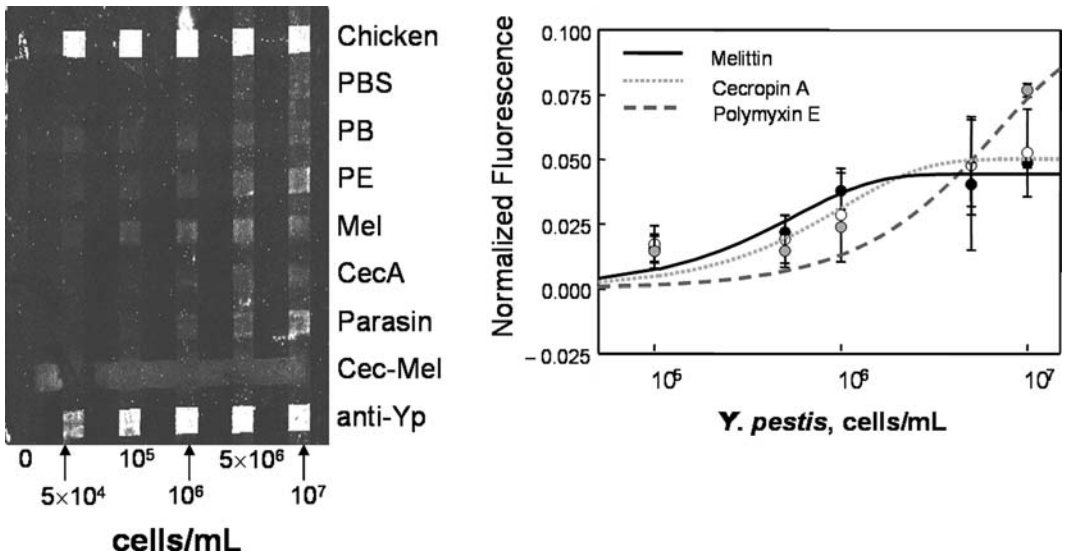


Fig. 11.4. Detection of inactivated *Y. pestis* CO-92 using sandwich assays. *Left*: representative image from array of immobilized antimicrobial peptides. *Right*: dose–response curve generated from data extracted from images. Solutions patterned included controls – anti-chicken IgY (Chicken), buffer (PBS), and anti-*Y. pestis* (anti-Yp); and antimicrobial peptides – polymyxin B (PB), polymyxin E (PE), melittin (Mel), cecropin A (CecA), parasin, and cecropin-melittin hybrid peptide (Cec-Mel).

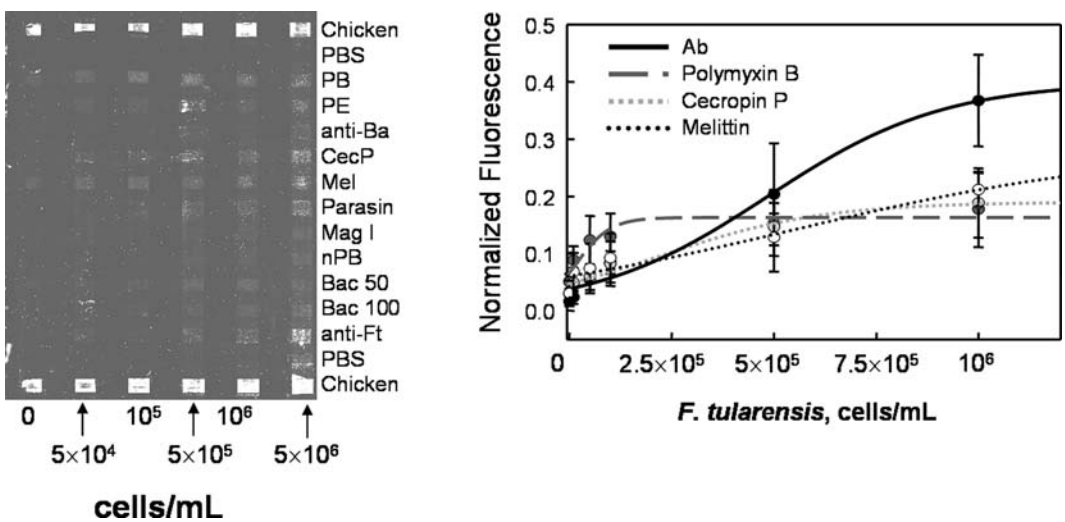


Fig. 11.5. Detection of inactivated *F. tularensis* using sandwich assays. *Left*: representative image from array of immobilized antimicrobial peptides. *Right*: dose–response curve generated from data extracted from images. Solutions patterned included controls – anti-chicken IgY (Chicken), buffer (PBS), anti-*B. anthracis* (anti-Ba), and anti-*F. tularensis* (anti-Ft); and antimicrobial peptides – polymyxin B (PB), polymyxin E (PE), cecropin P1 (CecP), melittin (Mel), parasin, magainin-I (Mag 1), polymyxin B nonapeptide (nPB), and bactenecin (Bac).

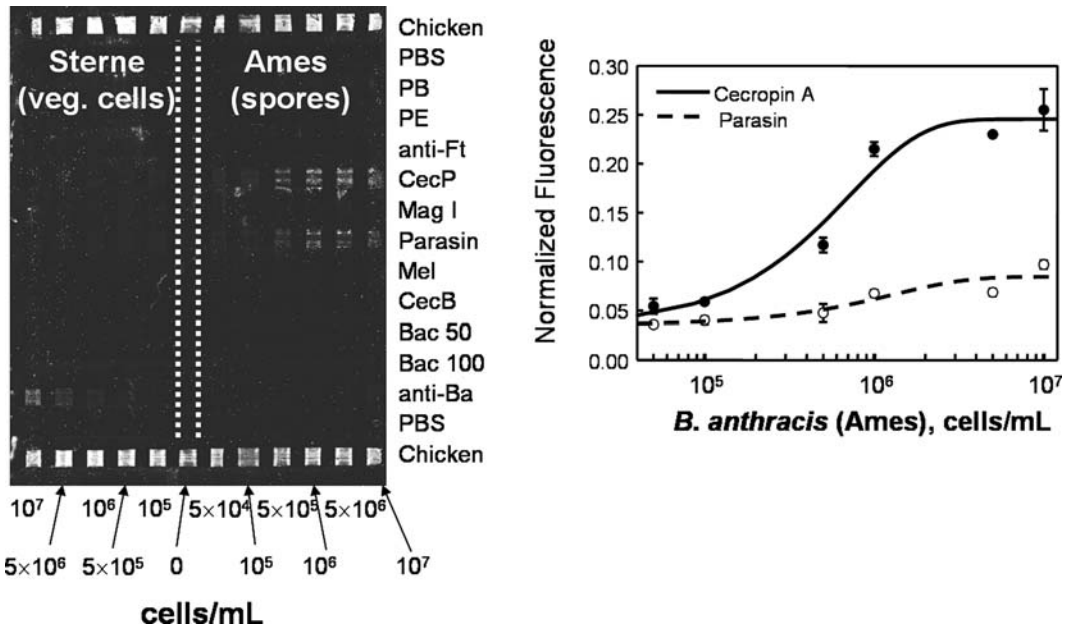


Fig. 11.6. Detection of inactivated *B. anthracis* using sandwich assays. *Left*: representative image from array of immobilized antimicrobial peptides. *Right*: dose–response curves generated for *B. anthracis* Ames strain from data extracted from images. Solutions patterned included controls – anti-chicken IgY (Chicken), buffer (PBS), anti-*F. tularensis* (anti-Ft), and anti-*B. anthracis* (anti-Ba); and antimicrobial peptides – polymyxin B (PB), polymyxin E (PE), cecropin P1 (CecP), magainin-I (Mag 1), parasin, melittin (Mel), cecropin B (CecB), and bactenecin (Bac). The cecropin A dose–response curve shown on the right is derived from other slides patterned with this peptide (not shown).

(**Fig. 11.6**, *right panel*), but no curve could be generated with the antibody control, since no signal was detected on these spots with Ames strain cells. Shown on this graph are summary dose–response curves ($n = 3$ slides) for binding to cecropin A and parasin. Although not shown on the image to the left, binding of *B. anthracis* Ames to cecropin A was the strongest response, similar in magnitude to cecropin P1. *B. globigii* and *B. thuringiensis* were also tested on similar arrays using Cy3-labeled rabbit anti-*B. globigii* (obtained from the Naval Medical Research Command, Bethesda, MD) and goat anti-*B. anthracis* (a kind gift from Dr. John Ezzell, USAMRIID, Frederick, MD), respectively. Somewhat different patterns of binding were observed for these species when compared to *B. anthracis* (**Fig. 11.7**). *B. globigii* bound to immobilized cecropin B and melittin and only marginally to cecropins A and P, whereas signals for *B. thuringiensis* bound to cecropin A were higher than those for the other peptides.

3.8.3. Detection of Enveloped Viruses

Radiation-inactivated preparations of Venezuelan equine encephalitis (VEE) and vaccinia virus were obtained from CRP and were certified inactive prior to shipment. The inability of the supplied antibodies to bind strongly to these viral preparations precluded development of sandwich format

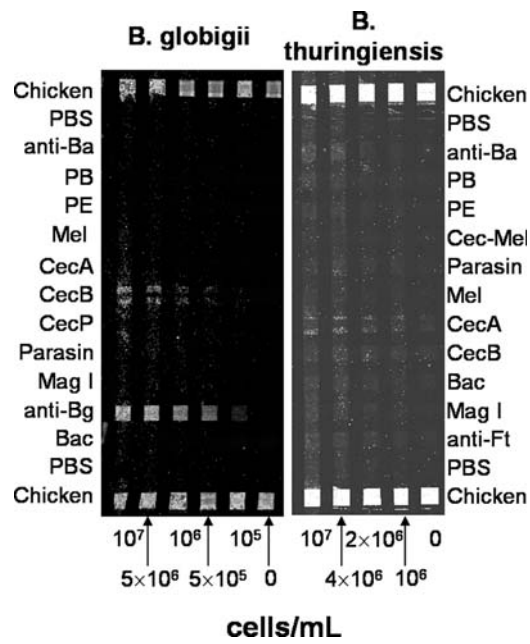


Fig. 11.7. Analysis of other *Bacillus* species using sandwich assays. Representative images from assays for *B. globigii* (*left panel*) and *B. thuringiensis* (*right panel*). Solutions patterned included controls – anti-chicken IgY (Chicken), buffer (PBS), anti-*B. anthracis* (anti-Ba), anti-*B. globigii* (anti-Bg), anti-*F. tularensis* (anti-Ft); and antimicrobial peptides – polymyxin B (PB), polymyxin E (PE), melittin (Mel), cecropins A, B and P1 (CecA, CecB, CecP), cecropin-melittin hybrid peptide (Cec-Mel), parasin, magainin-I (Mag 1), and Bactenecin (bac). The lack of strong signal in the image for *B. thuringiensis* may be due to the low affinity of the tracer antibody for this strain; the positive control (anti-Ba) also gave low fluorescence.

assays, and it was necessary to label the viral particles with Cy3 and perform direct assays (24). **Figure 11.8** shows representative images from direct assays for these two viruses; dose–response curves showed that the *patterns* of binding were virtually identical (24).

4. Notes

1. The silver "cladding" is necessary to confine excitation light within the slide/waveguide when using the NRL Array Biosensor. The pattern and dimensions of the cladding are given in (26). If using the NRL Array Biosensor, the fluorescent component (cells, viruses, antibodies) must be labeled with Cy5, rather than Cy3; all procedures described herein are appropriate for use with Cy5 derivatives also.

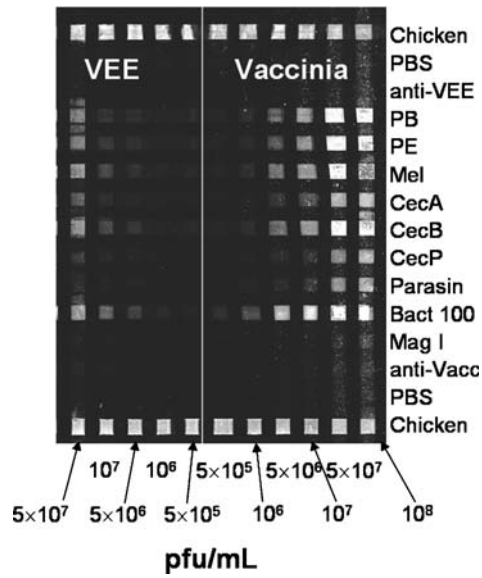


Fig. 11.8. Detection of enveloped viruses in a direct assay (labeled viruses). Representative image from direct assay for Venezuelan equine encephalitis (VEE) and vaccinia viruses; the viral particles were labeled with Cy3 to allow detection. Solutions patterned included controls – anti-chicken IgY (Chicken), buffer (PBS), anti-VEE, and anti-vaccinia (anti-Vacc); and antimicrobial peptides – polymyxin B (PB), polymyxin E (PE), melittin (Mel), cecropins A, B and P1 (CecA, CecB, CecP), parasin, bactenecin (Bact100), and magainin-I (Mag 1).

2. 10% KOH in isopropanol was originally described as one of the easiest and most effective methods for cleaning glass slides (30). Replacement of isopropanol with methanol expedites dissolution of the KOH without significantly affecting the cleanliness of the slides.
3. 3-Mercaptopropyl triethoxy and 3-mercaptopropyl trimethoxy silanes can be used interchangeably. However, as the methoxy derivative poses a more significant health hazard (it is an eye hazard), we typically prefer to use the ethoxy derivative.
4. Toluene is a reproductive hazard and hepatotoxin. All manipulations of solutions containing toluene should be performed in a chemical hood by personnel equipped with an appropriate respirator and other personal protective equipment.
5. GMBS is moisture-sensitive and hydrolyzes rapidly when exposed to water. After opening, tightly cap the vial when not in use, seal it with parafilm, and store in a desiccator at 4°C. Warm both the desiccator and capped vial to room temperature before opening the vial to avoid condensation.
6. Low-salt buffers must be used with silver-cladded slides to prevent delamination of the silver coating (28).

7. Antibodies can also be labeled at lower concentrations, as well as at neutral pH. However, the degree of labeling (dyes/antibody) is often lower under these non-optimal conditions.
8. Place a Kimwipe or paper towel under the beaker prior to applying vacuum and ensure that the beaker used is large enough to support a large increase in volume while degassing. During the first few minutes of degassing, it is prudent to observe the mix, decreasing the vacuum applied if the PDMS mix appears as though it will overflow the container.
9. PDMS patterning and assay templates can be reused many times. After each use, they should be scrubbed with soap and water, then incubated for 30 min in bleach. After bleach treatment, the templates are then rinsed exhaustively under running water and treated for >60 min with 2% BSA in PBS to render the surfaces more hydrophilic and decrease non-specific binding of different assay components. After blocking with BSA, the templates are then rinsed briefly with water and dried; they are then ready to reuse.
10. When processing large numbers of slides, rinsed slides should be placed into a fresh Coplin jar filled with deionized water until after the entire batch of cleaned slides are rinsed. The slides can then be individually removed and dried immediately, preventing any spotting due to non-uniform drying.
11. Although our own studies have shown that GMBS is optimal for most peptides, crosslinkers other than GMBS may be used, provided that maleimide and NHS–ester terminal moieties are present. The researcher should be cautioned that, even though their reactive moieties are identical, use of crosslinkers such as SMPB, SMCC, and sulfo-LC-SMPT (Pierce) may result in greatly altered binding affinities of the immobilized capture molecule (peptides, sugars, other small molecules) (31).
12. The rinsing step can be performed within the nitrogen-filled glove bag. Manipulating large numbers of slides within the glove bag is very cumbersome however, and we have found that moving this step elsewhere in the chemical hood is more convenient. Care must be exercised to ensure that adequate airflow is maintained within the hood to prevent airborne exposure to the toluene. See especially **Note 3**.
13. When using any potential biohazard, the waste should be inactivated/disinfected before disposal. For solutions flowing through the peristaltic pump, the flowthrough is pumped directly into a beaker containing bleach.
14. This rate of flow was used for all samples except *Bacillus* spp. For *Bacillus* samples, a rate of 0.1 mL/min was used.

15. We have found that use of Cy3-labeled tracer antibodies often results in lower background signals and non-specific sticking than Cy5-labeled equivalents (23). For this reason, preparation and use of Cy3 conjugates are described in this manuscript, although all protocols are appropriate for Cy5 also. The final choice of fluorophore, however, rests on the requirements of the imaging system.

Acknowledgements

The authors would like to acknowledge the financial support of Joint Science and Technology Office for Chemical and Biological Defense/Defense Threat Reduction Agency (#AA04DET005, 8.10016_08_NRL_B) and the National Institutes of Health (R01 EB00680). NVK was a recipient of a National Research Council fellowship. The views expressed herein are those of the authors and do not represent those of the US Navy, the US Department of Defense, or the US government.

References

1. Wadkins, R. M., Golden, J. P., Pritsiolas, L. M., and Ligler, F. S. (1998) Detection of multiple toxic agents using a planar array. *Biosens Bioelectron* **13**, 407–415.
2. Rowe, C.A., Scruggs, S.B., Feldstein, M.J., Golden, J.P., and Ligler, F.S. (1999) An array immunosensor for simultaneous detection of clinical analytes. *Anal Chem* **71**, 433–439.
3. Sapsford, K.E., Charles, P.T., Patterson, C.H. Jr., and Ligler, F.S. (2002) Demonstration of four immunoassay formats using the array biosensor. *Anal Chem* **74**, 1061–1068.
4. Ligler, F.S., Sapsford, K.E., Golden, J.P., Shriver-Lake, L.C., Taitt, C.R., Hanson, W., Barone, S., Myatt, C.J. (2007) The Array Biosensors: Portable, automated systems. *Anal Sci* **23**, 5–10.
5. Shriver-Lake, L.C., Erickson, J.S., Sapsford, K.E., Ngundi, M.M., Shaffer, K.M., Kulagina, N.V., Hu, J.E., Gray, S.A. III, Golden, J.P., Ligler, F.S., and Taitt, C.R. (2007) Blind laboratory trials for multiple pathogens in spiked food matrices. *Anal Lett* **40**, 3219–3231.
6. Rodriguez, M., Li, S.S.-C., Harper, J. W., and Songyan, Z. (2004) An oriented peptide array library (OPAL) strategy to study protein–protein interactions. *J Biol Chem* **279**, 8802–8807.
7. Chen, R., Pan, S., Brentnall, T.A, and Aebersold, R. (2005) Proteomic profiling of pancreatic cancer for biomarker discovery. *Mol Cell Proteom* **4.4**, 523–533.
8. Rychlewski, L., Kschischo, M., Dong, L., Schutkowski, M., and Reimer, U. (2004) Target specificity analysis of the Abl kinase using peptide microarray data. *J Mol Biol* **336**, 307–311.
9. Beernink, H.T.H. and Nock, S. (2005) Challenges facing the development and use of protein chips to analyze the phosphoproteome. *Exp Rev Proteomics* **2**, 487–497.
10. Diaz-Mochón, J.J., Bialy, L, and Bradley, M. (2006) Dual colour, microarray-based, analysis of 10000 protease substrates. *Chem Commun* **38**, 3984–3986.
11. Robinson, W.H., Steinman, L., and Utz, P.J. (2002) Protein and peptide array analysis of autoimmune disease. *BioTechniques* **33**, S66–S69.
12. Okochi, M., Nakanishi, M., Kato, R., Kobayashi, T., and Honda, H. (2006) High-throughput screening of cell death inducible short peptides from TNF-related apoptosis-inducing ligand sequence. *FEBS Lett* **580**, 885–889.

13. Bolger, G.B., Baillie, G.S., Li, X., Lynch, M.J., Herzyk, P., Mohamed, A., Mitchell, L.H. McCahill, A., Hundsrucker, C., Klussmann, E., Adams, D.R., and Houslay, M.D. (2006) Scanning peptide array analyses identify overlapping binding sites for the signalling scaffold proteins, β-arrestin and RACK1, in cAMP-specific phosphodiesterase PDE4D5. *Biochem J* **398**, 23–36.
14. Rigter, A., Langeveld, J.P.M, Timmers-Parohi, D., Jacobs, J.G., Moonen, P.L.J.M., and Bossers, A. (2007) Mapping of possible prion protein self-interaction domains using peptide arrays. *BMC Biochemistry* **8**, 6-1–6-14.
15. Hancock, R.E.W., and Chapple, D.S. (1999) Peptide antibiotics. *Antimicrob Agents Chemother* **43**, 1317–1323.
16. Matsuzaki, K. (2001) Why and how are peptide–lipid interactions utilized for self defence? *Biochem Soc Trans* **29**, 598–601.
17. Papo, N. and Shai, Y. (2003) Exploring peptide membrane interaction using surface plasmon resonance: differentiation between pore formation versus membrane disruption by lytic peptides. *Biochemistry* **42**, 458–466.
18. Haukland, H. H., Ulvatne, H., Sandvik, K. and Vorland, L. H. (2001) The antimicrobial peptides lactoferricin B and magainin 2 cross over the bacterial cytoplasmic membrane and reside in the cytoplasm. *FEBS Lett* **508**, 389–393.
19. Hsu, C.H., Chen, C., Jou, M.L., Lee, A.Y., Lin, Y.C., Yu, Y.P., Huang, W.T., and Wu, S.H. (2005) Structural and DNA-binding studies on the bovine antimicrobial peptide, indolicidin: evidence for multiple conformations involved in binding to membranes and DNA. *Nucleic Acids Res* **33**, 4053–64.
20. James, E.A., Schmeltzer, K., and Ligler, F.S. (1996) Detection of endotoxin using an evanescent wave fiber-optic biosensor. *Appl Biochem Biotechnol* **60**, 189–202.
21. Gregory, K. and Mello, C.M. (2005) Immobilization of *Escherichia coli* cells by use of the antimicrobial peptide cecropin P1. *Appl Environ Microbiol* **71**, 1130–1134.
22. Kulagina, N.V., Lassman, M.E., Ligler, F.S., and Taitt, C.R. (2005) Use of an antimicrobial peptide for detection of bacteria in biosensor assays. *Anal Chem* **77**, 6504–6508.
23. Kulagina, N.V., Shaffer, K.M., Anderson, G.P., Ligler, F.S., and Taitt, C.R. (2006) Antimicrobial peptide-based array for *E. coli* and *Salmonella* screening. *Anal Chim Acta* **575**, 9–15.
24. Kulagina, N.K., Shaffer, K.M., Ligler, F.S., and Taitt, C.R. (2007) Antimicrobial peptides, new recognition molecules for challenging targets. *Sens Actuators B* **121**, 150–157.
25. Kulagina, N.V., Anderson, G.P., Ligler, F.S., Shaffer, K.M., and Taitt, C.R. (2007) Antimicrobial peptides: New recognition molecules for detecting botulinum toxins. *Sensors* **7**, 2808–2824.
26. Feldstein, M.J., Golden, J.P., Rowe, C.A., MacCraith, B.D., and Ligler, F.S. (1999) Array Biosensor: Optical and fluidics systems. *J Biomed Microdevices* **1**, 139–153.
27. Taitt, C.R., Anderson, G.P., Lingerfelt, B.M., Feldstein, M.J., and Ligler F.S. (2002) Nine-analyte detection using an array-based biosensor. *Anal Chem* **74**, 6114–6120.
28. Feldstein, M.J., Golden, J.P., Ligler, F.S., Rowe, C.A. (2000) Optical waveguide-Flow cell integration method. US Patent 6,192,168, issued 24 Oct 2000.
29. Ngundi, M.M., Taitt, C.R., and Ligler, F.S. (2006) Simultaneous determination of kinetic parameters for the binding of cholera toxin to immobilized sialic acid and monoclonal antibodies using an array biosensor. *Biosens Bioelectron* **22**, 124–130.
30. Cras, J. J., Rowe-Taitt, C. A., Nivens, D. A., and Ligler, F. S. (1999) Comparison of chemical cleaning methods of glass in preparation for silanization. *Biosens Bioelectron* **14**, 683–688.
31. Ngundi, M.M., Taitt, C.R., Ligler, F.S. (2007) Crosslinkers modify affinity of immobilized carbohydrates for cholera toxin. *Sens Lett* **5**, 621–624.

Chapter 12

Mapping Functional Prion–Prion Protein Interaction Sites Using Prion Protein Based Peptide-Arrays

Alan Rigter, Jan Priem, Drophatie Timmers-Parohi, Jan PM Langeveld, and Alex Bossers

Abstract

Protein–protein interactions are at the basis of most if not all biological processes in living cells. Therefore, adapting existing techniques or developing new techniques to study interactions between proteins are of importance in elucidating which amino acid sequences contribute to these interactions. Such new insights may in turn lead to improved understanding of the processes underlying disease and possibly provide the basis for new therapeutic approaches. Here we describe the novel use of an ovine prion protein-based peptide-array normally used for determining prion-specific antibody epitopes, with the prospect that this would yield information on interaction sites between its PrP moiety and the ovine prion protein derived linear peptides. This adapted application of the peptide-array shows, by incubating the mature part of ovine (ARQ) PrP^{C} fused to maltose binding protein (MBP), binding with between the PrP moiety and the ovine prion derived peptides occurs and indicates that several specific self-interactions between individual PrP molecules can occur; hereby illustrating that this adapted application of a peptide-array is a viable method to further specify which distinct amino acid sequences are involved in protein–protein interaction.

Key words: Protein–protein interaction, binding site, mapping, amino acid sequence, prion, peptide-array.

1. Introduction

Central in many biological processes is the interaction between proteins. This encompasses interaction with other as well as "self" proteins, resulting in a multitude of effects. One example is signal transduction, in which extracellular signals are conferred to the inside of a cell by protein–protein interactions of the signalling molecules. Another is the long-term interaction between proteins as part of a protein complex, in which one protein helps another from one to another cellular compartment (i.e. importins). Also

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_12,

brief interactions between proteins occur, resulting in modification of the target protein (i.e. post-translational modification), which in itself can change protein–protein interactions. Thus, protein–protein interactions are of importance for most if not all biological processes in living cells. Therefore, studying these interactions and elucidating the contributing amino acid sequences involved in these interactions may lead to improved understanding of the processes underlying disease and may provide the basis for new therapeutic approaches.

The common event in transmissible spongiform encephalopathies (TSEs) or prion diseases is the conversion of host-encoded protease sensitive cellular prion protein (PrP^{C}) into strain-dependent isoforms of scrapie-associated protease resistant isoform (PrP^{Sc}) of prion protein (PrP). Formation of PrP^{Sc} is a post-translational process and involves refolding (conversion) of the host-encoded prion protein (PrP^{C}) into partially protease resistant forms (PrP^{Sc}) (1).

These processes are determined by similarities as well as strain-dependent variations in the PrP structure (2–11). Selective self-interaction between PrP molecules is the most probable basis for initiation of these processes, potentially influenced by chaperone molecules; however, the mechanisms behind these processes are far from understood. Here we describe the utilization of a peptide-array consisting of 15-mer overlapping peptides systematically covering the whole mature part of the cellular prion protein (**Fig. 12.1**) in order to elucidate the interaction domains involved in self-interaction of PrP^{C} (12). To this end, an ovine PrP peptide-array consisting of 15-mer overlapping peptides was probed with recombinant sheep PrP^{C} fused to maltose binding protein (MBP–PrP).

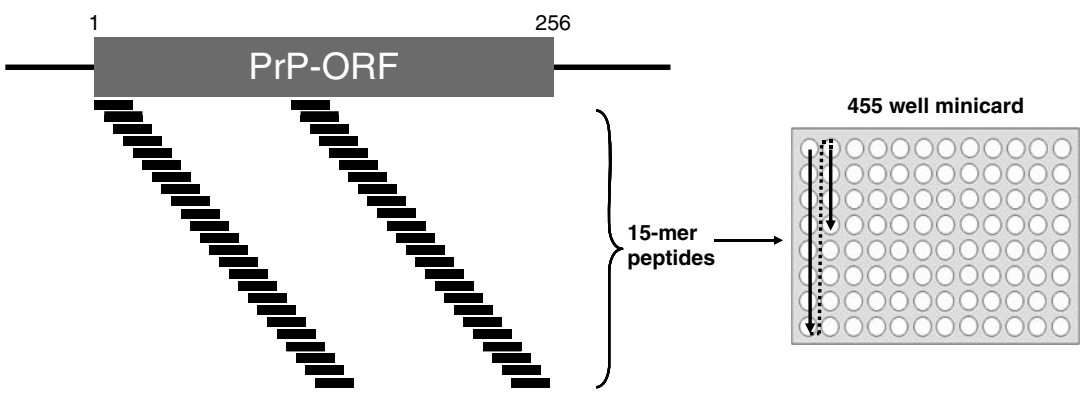


Fig. 12.1. Basic setup of the prion peptide-array. The mature PrP-ORF (excluding N- and C-terminal signal sequences) is divided into 15-mer overlapping peptides, resulting in a grid where in each well the amino acid sequence has shifted by one amino acid.

2. Materials

2.1. MBP–PrP Construction and Verification

1. Difco™ LB Broth (Miller)
2. Difco™ LB Agar (Miller)
3. 100 mg/ml Ampicillin stock (Sigma, –20°C)
4. Amplitaq® Gold DNA polymerase (Applied Biosystems)
5. GeneAmp® PCR buffers (Applied Biosystems)
6. GeneAmp® dNTP blend (Applied Biosystems)
7. Agarose (Merck)
8. TA cloning vector (Invitrogen, –20°C)
9. One Shot® TOP10 Chemically Competent *E. coli* (Invitrogen, –20°C)
10. Wizard® *Plus* Minipreps DNA Purification System (Promega)
11. Enzymes *Dra*I and *Eco*RI (New England Biolabs, –20°C)
12. pMAL-c2X expression vector (New England Biolabs, –20°C)
13. Monoclonal PrP-specific antibody (9A2; in house, –20°C, *see* **Note 1**)
14. Monoclonal MBP-specific antibody (New England Biolabs, –20°C, *see* **Note 1**)
15. Factor Xa (New England Biolabs)
16. 10% SDS stock (Merck)
17. NuPAGE gel-electrophoreses system (Invitrogen):
    a. XCell SureLock® Mini-Cell
    b. 12% NuPAGE® Novex Bis-Tris Gel (4°C)
    c. 20X NuPAGE® MOPS SDS Running Buffer (4°C)
    d. NuPAGE® Antioxidant (4°C)
    e. MagicMark™ XP Western Protein Standard (4°C)
    f. XCell II™ Blot Module CE Mark
    g. NuPAGE® Transfer Buffer (20X) (4°C)
    h. Ureum sample buffer (4M ureum + 10% β-mercaptoethanol, –20°C)
18. Hybond™-C Extra (Amersham, optimized for protein transfer)
19. 0.05% PBS–Tween tablets (Calbiochem, RT)
20. SYPRO® Orange protein gel stain (Invitrogen, 5000X concentrate in DMSO, 4°C)
21. STORM 840/Typhoon (GE Healthcare, Life Sciences)
22. ECF™ substrate for Western blotting (GE Healthcare, –20°C, *see* **Note 2**)

2.2. MBP–PrP Expression and Analysis

1. Difco™ LB Broth (Miller, freshly made and autoclaved prior to use)
2. 100 mg/ml Ampicillin stock (Sigma, –20°C)
3. 2 g/l Glucose (Merck)
4. 0.1 M IPTG stock (Sigma, –20°C)
5. *E. coli* containing pMAL–MBP–PrP (glycerol-stock, –80°C)
6. Protease inhibitors (1000X stock) (–20°C, *see* **Note 3**):
   a. 25 mg/ml Pefabloc-Sc (AEBSF) in dH_2O (500 mg, Roche)
   b. 700 μg/ml Pepstatine (1 μM) in MeOH (10 mg, Roche)
   c. 500 μg/ml Aprotinine (0.15 μM) in dH_2O (10 mg, Roche)
   d. 500 μg/ml Leupeptine (1 μM) in dH_2O (25 mg, Roche)
   e. 0.5 M EDTA, pH 8.0 (Merck, RT)
7. Demineralized water (MiliQ)
8. Methanol, analysis grade (Merck)
9. Column-buffer (RT): 20 mM Tris-HCl pH 7.4, 200 mM NaCl, 1 mM EDTA pH 8.0, 10 mM β-mercaptoethanol (added fresh prior to use).
10. Cuphorn sonicator (Branson)
11. 10% Triton X-100 stock (BDH, RT)
12. Amylose slurry (New England Biolabs, supplied pre-swollen in 20% ethanol, 4°C)
13. 5 ml Disposable columns (Qiagen)
14. 1 M Maltose stock (Merck, 4°C)
15. 0.5 ml siliconized screw-cap tubes (BioPlastics)
16. NuPAGE gel-electrophoreses system (Invitrogen)
17. Hybond™-C Extra (Amersham, optimized for protein transfer)
18. 0.05% PBS–Tween tablets (Calbiochem, RT)
19. SYPRO® Orange protein gel stain (Invitrogen, 5000X concentrate in DMSO, 4°C)
20. 10% non-immune goat serum (Zymed laboratories inc.)
21. Monoclonal MBP-specific antibody (New England Biolabs, –20°C, *see* **Note 1**)
22. STORM 840/Typhoon Trio (GE Healthcare, Life Sciences)
23. ECF™ substrate for Western blotting (GE Healthcare, –20°C, *see* **Note 2**)

2.3. Peptide-Array Analysis

Materials needed for synthesis and coupling of 15-mer peptides to form the ovine prion protein peptide-array are to a certain extent described in (13–16); however, the exact methodological details are the intellectual property of Pepscan Systems BV, who perform these experiments on a contractual basis. Measurement of colouration of the substrate and subsequent translation to the optical density requires a specialized digital camera and analysis software setup as described by Slootstra et al. (15). The whole procedure, except for subsequent analysis of the final density values, was performed by Pepscan Systems BV.

1. SQ-buffer (4°C): 5% (v/v) horse serum, 5% (v/v) hen albumin, 1% (v/v) Tween 80 in PBS
2. Monoclonal MBP-specific antibody (New England Biolabs, –20°C, *see* **Note 1**)
3. Rabbit anti-mouse-IgG antibody coupled horse radish peroxidase antibody (Dako, 4°C)
4. 0.05% PBS–Tween (Calbiochem, RT)
5. ABTS peroxidase substrate (Sigma-Aldrich)
6. Hydrogen peroxide (Merck, H_2O_2)
7. CCD camera and image processing system:
   a. Sony CCD Video Camera XC-77RR
   b. Nikon micronikkor 55 mm f/2.8 lens
   c. Camera adaptor (Sony Camera adaptor DC-77RR)
   d. Image processing softwarepackage TIM, v. 3.36 (Difa Measuring Systems, Breda, The Netherlands).

3. Methods

3.1. MBP–PrP Construction and Verification

Polymerase chain reaction (PCR) was performed conforming to standard conditions. Cloning of the PCR-product into the TA-cloning vector and sub-cloning into pMAL-c2X were performed conforming to the protocols supplied by the manufacturers.

1. In order to obtain the PrP gene suitable for cloning into the pMAL protein fusion and purification system, the mature part of the ovine PrP (ARQ) open reading frame (ORF) was PCR amplified using platinum *Taq* polymerase and PrP specific primers ShBo-F-DraI (5′-GGTGGTTTTAAAAAGCGAC-CAAAACCTGG-3′) and Sh-R-STOP (5′-GGTGGTCTA

TGCCCCCCTTTGGTAATAAGCC-3′) using standard 50 µl PCR reaction containing: 100 ng DNA template, 5 µl PCR buffer (10X), 0.5 µl dNTPs (50 mM), 1.0 µl of each primer (10 mM), sterile water to a total volume of 49 µl and 1 unit of *Taq* polymerase.

2. In order to check the integrity and purity of PCR product, 5.0 µl of the PCR reaction was run on a 1% agarose gel. The PCR produced a single band of appropriate size; therefore, the PCR product was directly cloned into the TA-vector (*see* **Note 4**).
3. The resulting PrP (AA25-233), without its N- and C-terminal signal sequences, was cloned into a general TA-cloning vector using the TA Cloning® Kit, by adding 1.0 µl of fresh PCR-product to 50 ng of the TA-vector, 1.0 µl of 10X ligation buffer, 1.0 µl of T4 DNA Ligase (4 Weiss units) and make up to a total volume of 10.0 µl with sterile water.
4. After ligation of the PCR-product in the TA-vector, the complete ligation reaction was transformated in the commercial TOP10 chemically competent *E. coli* and the transformated cells spread on LB-agar plates containing 100 µg/ml ampicillin.
5. Single colonies were picked and grown overnight at 37°C while shaking in 2 ml LB-broth containing 100 µg/ml ampicillin.
6. Plasmid DNA was isolated using the Wizard® *Plus* Minipreps DNA Purification System for further use.
7. Before sub-cloning the mature PrP ORF in the pMAL-c2X expression plasmid, the PrP ORF was sequenced to exclude PCR artefacts.
8. The PrP fragment was subsequently sub-cloned using *Dra*I and *Eco*RI into the pMAL-c2X expression vector, resulting in the maltose binding protein (MBP) fusion to the N-terminus of PrP (MBP–PrP).

In order to assess whether the complete mature part of PrP^C was expressed correctly and whether the fusion protein had accessible folding, MBP–PrP was expressed and analysed as described in **Section 3.2**. MBP–PrP was digested with factor Xa as described in the manual of the pMAL protein fusion and purifications system. Expression of maltose binding protein N-terminally fused to PrP (MBP–PrP) revealed a mainly soluble recombinant MBP–PrP of approximately 70 kDa (**Fig. 12.2**, lanes 1 and 2) and the naked PrP protein could be obtained by digestion with protease Factor Xa, indicating accessible folding (**Fig. 12.2**, lanes 3–6). After 24 h approximately 45% of MBP–PrP was digested by factor Xa; however, when aided by addition of 0.01% SDS

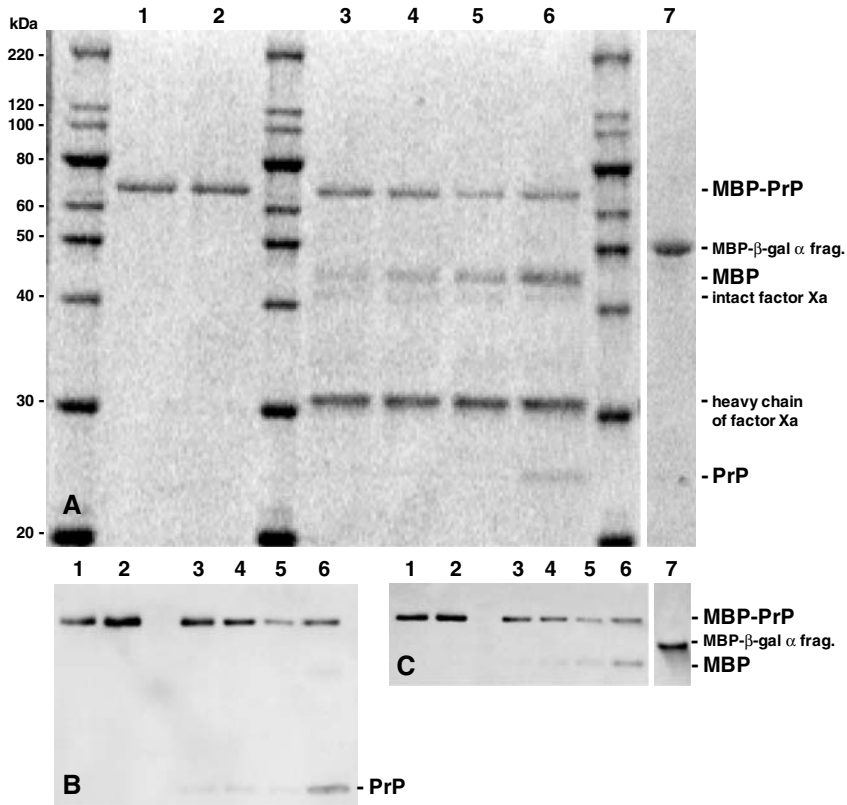


Fig. 12.2. Analysis of MBP–PrP and MBP expression and MBP–PrP digestion by Factor Xa. Lane 1 contains untreated MBP–PrP, whereas lane 2 contains a mock digestion of MBP–PrP. MBP–PrP was digested with 1% w/w factor Xa and during digestion samples were taken at 2, 4, 7 and 24 h (lanes 3,4, 5 and 6, respectively). All samples were run on SDS-PAGE and the gel was stained with Sypro Orange (total protein stain, *panel A*) before Western blotting and subsequent immunodetection using either a PrP-specific monoclonal antibody (9A2, *panel B*) or MBP-specific monoclonal antibody (α-MBP, *panel C*). Expression of MBP, expressed from the pMAL-c2X vector with no insert (MBP-β-gal α fragment), was analysed by Western blot using either 9A2 or α-MBP (lane 7, *panels A* and *C*, respectively). Reproduction of **Fig. 1**, published in BMC Biochemistry by Rigter et al. (12).

factor Xa completely digested MBP–PrP within 24 h (data not shown). Furthermore, MBP–PrP is readily detected in Western blot using both a PrP-specific antibody (9A2, **Fig. 12.2B**) and a MBP-specific antibody (**Fig. 12.2C**). Several monoclonal and polyclonal antibodies with recognition sites dispersed throughout the PrP protein detected MBP–PrP in Western blot, indicating complete expression of the mature part of sheep PrP^C (data not shown). MBP expressed without additional fusion protein (PrP), which frequently served as negative control in this study, was also of homogeneous quality (**Fig. 12.2**, lane 7) and of expected size (MBP-β-gal α fragment, 50.8 kDa) which is somewhat larger (as expected) than MBP cleaved from the fusion protein after factor Xa digestion (42.5 kDa).

3.2. MBP–PrP Expression and Analysis

Expression and purification by affinity chromatography was performed as described in the manual of the pMAL protein fusion and purifications system (method I) with the following adaptations:

1. The litre of rich broth (LB-broth containing 2 g/l glucose and 50 μg/ml ampicillin) was inoculated with 50 ml overnight culture.
2. To improve binding of MBP–PrP by preventing formation of interchain disulphide bonds upon lysis (as suggested in the protocol), 10 mM β-mercaptoethanol was added to the column buffer.
3. Before sonication of the cell suspension, an inhibitor cocktail (pefabloc sc, pepstatine, aprotinine, leupeptine and EDTA) was added and the cell suspension was sonicated eight times for 15 s, resting the cell-extract on ice for 45 s between sonication rounds.
4. To ensure optimal lyses of the cells 1% (v/v) Triton X-100 was added to the sonicated cell suspension and incubated for 20 min at 4°C, while gently rocking.
5. After lyses the dilution of the crude extract with column buffer was not necessary.
6. A column was prepared using 2 ml of 20% amylase slurry and the MBP–PrP fusion protein was eluted after binding with column buffer containing 10 mM maltose in 10 fractions of 250 μl.

Quantity and quality of the eluted MBP–PrP was determined by SDS-PAGE as follows.

1. From each collected fraction a 10 μl sample was taken and run on a 12% NuPAGE® Novex Bis-Tris Gel together with a bovine serum albumin (BSA) concentration curve samples (2.0, 1.0, 0.5 and 0.25 μg, *see* **Note 5**).
2. After separation by electrophoresis, the gel was taken out of its casing, trimmed and washed in 1X NuPAGE® transfer buffer for 10 min to remove excess unbound SDS from the gel (*see* **Note 6**).
3. The gel was then transferred to fresh 1X NuPAGE® transfer buffer containing 1:5000 diluted Sypro® Orange and incubated for 1 h (*see* **Note 7**).
4. The gel was then rinsed in 1X NuPAGE® transfer buffer and scanned using the STORM 840 using standard settings (*see* **Note 8**).
5. After scanning, the gel was blotted for 1 h at constant voltage of 30 V onto nitrocellulose (Hybond™-C Extra) using the XCell II™ Blot Module CE Mark and NuPAGE® transfer buffer (*see* **Note 9**).

6. The resulting blot was transferred to a 50 ml falcon tube containing 40 ml blocking buffer (5% milk protein in PBS) and incubated for 1 h at room temperature while gently shaking. Alternatively the blot can be incubated overnight at 4°C.
7. After blocking the blot was rinsed once in 40 ml of 0.05% PBS–Tween.
8. The blot was then incubated in 5 ml PBS–Tween containing 1% non-immune goat serum (NGS) and either 1:1500 diluted MBP-specific antibody or 1:1000 diluted PrP-specific antibody 9A2 as primary antibody, for 1 h at room temperature while gently shaking.
9. After incubating with the primary antibody, the blot was rinsed once in 40 ml of PBS–Tween (0.05%).
10. After rinsing the blot was washed two times in 25 ml 0.05% PBS–Tween for 10 min while gently shaking.
11. The blot was then incubated in 5 ml 0.05% PBS–Tween containing 1% NGS and 1:2000 diluted goat-anti-mouse-IgG coupled with alkaline phosphatase as the secondary antibody, for 1 h at room temperature while gently shaking.
12. After incubating with the secondary antibody, the blot was rinsed once in 40 ml of 0.05% PBS–Tween.
13. After rinsing the blot was washed two times in 25 ml 0.05% PBS–Tween for 10 min while gently shaking.
14. The 0.05% PBS–Tween of the final wash is discarded and 50 ml of fresh 0.05% PBS–Tween was added for transport and storage at 4°C (*see* **Note 10**).
15. The surface area of the blot was determined and 24 µl/cm^2 ECF was pipetted directly onto the glass plate of the STORM 840.
16. The blot was taken out of the 0.05% PBS–Tween in which it was stored, and gently dried before placing the blot onto the ECF, making sure the ECF spreads underneath the whole blot (*see* **Note 11**).
17. The blot was immediately scanned using the STORM 840 scanning software and standard settings (*see* **Note 12**) for ECF and the scan was repeated after 5, 10, 15 and if necessary 20 min (*see* **Note 13**) of the start of the initial scan.
18. Using the ImageQuant software supplied by the manufacturer, the band density was determined for further analysis.

3.3. Peptide-Array Analysis

Synthesis of complete sets of overlapping 15-mer peptides (**Fig. 12.1**) were carried out on grafted plastic surfaces, covering the ovine PrP amino acid sequence of mature PrP^C (residues 25–234 of ovine and 25–242 of bovine PrP) (17). The plastic surface consisted of 455, 3 µl wells on a credit-card size plastic (minicard) carrier. In each well the peptides were coupled to the grafted plastic surface in two ways: either by direct step-by-step coupling of amino acids or by coupling complete 15-mer peptides at their C-terminus (13–16) and subsequent ELISA analyses were performed. The established procedures for peptide-array analysis are used to determine antibody recognition domains (**Fig. 12.3A**); however, when using the prion protein peptide-array to determine interaction domains of PrP, direct detection with α-PrP antibodies can possibly recognize peptides that did not bind PrP and thus frustrate the results. However, by incubating MBP–PrP as the antigen that binds to the peptide, detection of the fusion protein is possible using α-MBP antibody, followed by rabbit α-mouse-HRP (**Fig. 12.3B**), preventing frustration of the results due to cross-reaction between the primary antibody and the ovine prion protein derived peptides. The procedure described below is not extensive; however, points 2–4 entail the necessary adaptation of the standard peptide-array analysis protocol, which allows for elucidation of the amino acid sequences involved in protein–protein interaction.

1. Support-coupled peptides were pre-coated with SQ-buffer for 1 h at 37°C to block non-specific absorption of antibodies.

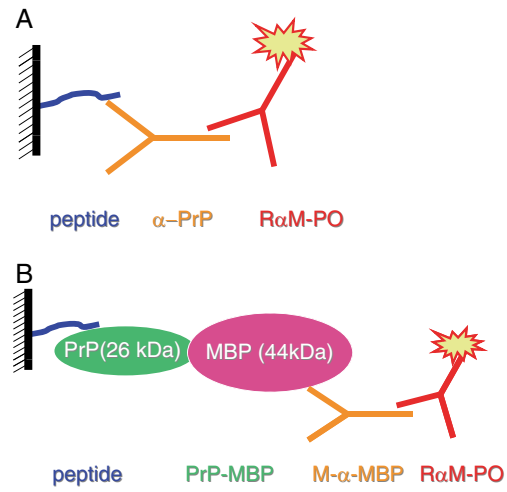


Fig. 12.3. ELISA-based detection principle of peptide-array analysis. (**A**) Standard detection setup for a peptide-array in order to determine antibody epitope; (**B**) Alternative setup used to determine peptide–protein interaction in this study.

2. MBP–PrP was diluted in SQ-buffer (standard buffer used by Pepscan B.V.) to a concentration resulting in a sufficient relative density value, allowing differentiation between peaks (**Fig. 12.4**, column graph), without affecting the background and preventing the relative density value reaching its plateau. In the case of MBP–PrP a concentration of 7.5–10 µg/ml was optimal (*see* **Note 14**).
3. MBP–PrP was incubated overnight at 4°C as an antigen on the peptide-array minicard.
4. The peptide-array minicard was washed three times in 0.05% PBS–Tween and all buffer was removed from the wells.
5. The peptide-array minicard was incubated with the primary MBP-specific antibody diluted 1:2000 in SQ-buffer for 1 h at 25°C.

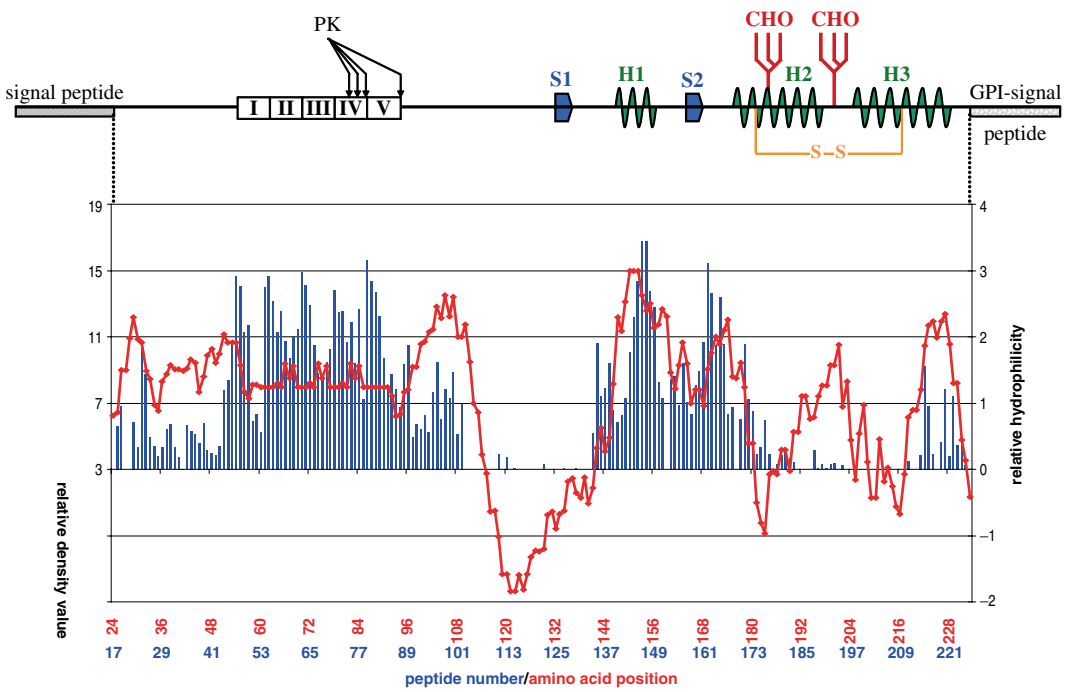


Fig. 12.4. Overview of PrP^C secondary structures and antibody epitopes versus peptide-array binding pattern and Kyte–Doolittle hydrophilicity plot. Schematic representation of PrP^C showing signal sequences, β-sheets (S1, S2), α-helices (H1, H2, H3), disulphide bridge site (S–S) and glycosylation sites (CHO). The sequence of PrP is lined up with both the Kyte–Doolittle hydrophilicity plot (*line graph*; negative is hydrophobic and positive is hydrophilic) and the relative binding pattern found (*column graph*) with the ovine prion peptide-array. Adaptation of **Fig. 5**, published in BMC Biochemistry by Rigter et al. (12).

6. The peptide-array minicard was washed three times in 0.05% PBS–Tween and the buffer was removed from all the wells.
7. The peptide-array minicard was incubated with the secondary rabbit anti-mouse-IgG antibody coupled horse radish peroxidase antibody diluted 1:1000 in SQ-buffer for 1 h at 25°C.
8. The peptide-array minicard was washed three times in PBS–Tween (0.05%) and the buffer was removed from all the wells.
9. After washing the peroxidase substrate ABTS and 2 µl/ml 3% H_2O_2 was added and incubated for 1 h at room temperature.
10. The optical density of the signals in the separate wells we measured using a CCD camera and an image processing system as described by Slootstra et al.(15).
11. The measured density values were further analysed as follows:
    The background was determined by calculating the mean value of 20 peptides with low density values of which at least 5 peptides were in consecutive order. The relative density value (r.d.v.) was calculated by dividing the optical density value (o.d.v.) by the background and binding was considered relevant when at least three consecutive peptides showed binding values of at least three times the background.

Interaction between the individual PrP sequences (peptides) and MBP–PrP was sufficient for immunodetection, resulting in a reproducible binding pattern (**Fig. 12.4**, column graph). This binding pattern, expressed in relative density values was characterized by two distinct high binding areas (peptides 35–102 and 134–102, respectively) as well as some lower binding areas. Analysis of the correlating peptide sequences revealed that these areas usually were characterized by consensus sequences which suggested the existence of distinct interaction domains for the mature part of PrP^C (for detailed analysis *see* (12)). To further assess the extent of the specificity of the binding pattern found, MBP–PrP was also tested against a bovine PrP peptide-array. This yielded a similar binding pattern comparable to the results with ovine PrP peptide-array but with some slight differences, which can be correlated to differences in amino acid differences between the ovine and bovine prion protein (12) (data not shown).

To find a correlation with structural properties, the relative binding pattern of MBP PrP on the peptide-array was compared to the Kyte–Doolittle hydrophilicity plot of mature PrP^C (**Fig. 12.4**, line graph) revealing a high correlation between hydrophilic (exposed) regions of PrP^C and the observed binding pattern regions.

4. Notes

1. Storage of the antibody stocks is done best at –20°C. Stocks are best divided in aliquots before storage. When an aliquot is taken into use it is best to store this aliquot at 4°C to prevent antibody degradation by repeated thawing and freezing until finished.
2. ECF is best stored as aliquots of 1 ml (about two blots) to minimize decline of the substrate as a result of repeated thawing and freezing.
3. For ease of use it is best to aliquot the protease inhibitors (except EDTA) and store at –20°C, for prolonged storage. The aliquot in use can be stored at 4°C until finished.
4. When the PCR does not produce a single band it is pertinent to run the whole PCR reaction on an agarose gel and excise the band that correlates to the specific PCR product wanted. Extract the DNA from the gel size using a commercially available kit. However, the amount of PCR product recovered is usually low and if possible it is advised to optimize the PCR reaction in order to get a single band. Then the procedure described can be followed, with greater chance of success.
5. After staining with Sypro Orange and visualization using the STORM 840, the ImageQuant software supplied by the manufacturer can be used to determine the (relative) density of the BSA bands, these densities ("volumes") can be extrapolated against the concentration of BSA loaded in the lane, resulting in a (linear) concentration curve. From this curve the concentration of the protein present in the fractions can be estimated using the determined band volumes for these fractions.
6. Wash at least for 10 min, but no more than 30 min. Sypro Orange binds to protein-bound SDS. Washing helps eliminate background colouring; however, prolonged washing diminishes the protein staining.
7. Sypro Orange is light sensitive; therefore, incubating in a closed light-impermeable container is advised for optimal results.
8. Sypro Orange has an excitation max of 300 and 470 nm, and an emission max of 570 nm; use Blue fluorescence mode (automatically coupled to emission filter) on STORM 840.
9. Correct stacking of the gel, nitrocellulose and blot-pads is described in the information provided by the manufacturer.

10. The blot can be stored at 4°C in PBS–Tween for several weeks without loss of signal. However, storage may cause the signal to come up later compared to direct measuring using ECF and the STORM 840.
11. When placing the blot onto the ECF, make sure no bubbles are trapped underneath the blot. It is easier to add extra ECF when necessary afterwards, when to much ECF is present this will flow out from underneath the blot and will create smearing during scanning.
12. ECF has an excitation max of 440 nm and an emission max of 560 nm; use the Blue fluorescence mode (automatically coupled to emission filter) on STORM 840.
13. Scanning at different time-intervals allows for choosing the scan in which good signal is present, but in which the highest signal has not yet reached its plateau. This ensures that the subsequent density measurements of the bands can be compared in further analysis.
14. During the experiments we discovered that soluble MBP–PrP protein interacts with the peptides on the peptide-array. Aggregation of MBP–PrP resulted in disappearing of signal; therefore, slight differences in optimal MBP–PrP concentration are likely due to differences in solubility between different batches of MBP–PrP used.

Acknowledgements

This work was supported by a grant 903-51-177 from the Dutch Organization for Scientific Research (NWO), by a grant from the Dutch Ministry of Agriculture, Nature Management and Fisheries (LNV) and by EU NeuroPrion project FOOD-CT-2004-506579. We kindly thank Saskia Ruiter for her work on the MBP–PrP expression construct.

References

1. DeArmond, S. J., and Prusiner, S. B. (2003) Perspectives on prion biology, prion disease pathogenesis, and pharmacologic approaches to treatment. *Clin Lab Med*, **23**, 1–41.
2. Atarashi, R., Sim, V. L., Nishida, N., Caughey, B., and Katamine, S. (2006) Prion strain-dependent differences in conversion of mutant prion proteins in cell culture. *J Virol*, **80**, 7854–62.
3. Bossers, A. (1999) PhD-Thesis: "Prion diseases: Susceptibility and transmissibility. In vivo and in vitro studies with sheep scrapie". June 10th1999, University of Utrecht, The Netherlands. ISBN 90–393–2108–6, http://library.wur.nl/WebQuery/wurpubs/lang/309746.
4. Bossers, A., Belt, P., Raymond, G. J., Caughey, B., de Vries, R., and Smits, M. A. (1997) Scrapie susceptibility-linked

polymorphisms modulate the in vitro conversion of sheep prion protein to protease-resistant forms. *Proc Natl Acad Sci USA*, **94**, 4931–6.

5. Bossers, A., Rigter, A., de Vries, R., and Smits, M. A. (2003) In vitro conversion of normal prion protein into pathologic isoforms. *Clin Lab Med*, **23**, 227–47.
6. Bossers, A., Schreuder, B. E., Muileman, I. H., Belt, P. B., and Smits, M. A. (1996) PrP genotype contributes to determining survival times of sheep with natural scrapie. *J Gen Virol*, 77(Pt 10), 2669–73.
7. Kocisko, D. A., Priola, S. A., Raymond, G. J., Chesebro, B., Lansbury, P. T., Jr., and Caughey, B. (1995) Species specificity in the cell-free conversion of prion protein to protease-resistant forms: a model for the scrapie species barrier. *Proc Natl Acad Sci USA*, **92**, 3923–7.
8. Raymond, G. J., Bossers, A., Raymond, L. D., O'Rourke, K. I., McHolland, L. E., Bryant, P. K., 3rd, Miller, M. W., Williams, E. S., Smits, M., and Caughey, B. (2000) Evidence of a molecular barrier limiting susceptibility of humans, cattle and sheep to chronic wasting disease. *Embo J*, **19**, 4425–30.
9. Raymond, G. J., Hope, J., Kocisko, D. A., Priola, S. A., Raymond, L. D., Bossers, A., Ironside, J., Will, R. G., Chen, S. G., Petersen, R. B., Gambetti, P., Rubenstein, R., Smits, M. A., Lansbury, P. T., Jr., and Caughey, B. (1997) Molecular assessment of the potential transmissibilities of BSE and scrapie to humans. *Nature*, **388**, 285–8.
10. Sabuncu, E., Petit, S., Le Dur, A., Lan Lai, T., Vilotte, J. L., Laude, H., and Vilette, D. (2003) PrP polymorphisms tightly control sheep prion replication in cultured cells. *J Virol*, 77, 2696–700.
11. Rigter, A., and Bossers, A. (2005) Sheep scrapie susceptibility-linked polymorphisms do not modulate the initial binding of cellular to disease-associated prion protein prior to conversion. *J Gen Virol*, **86**, 2627–34.
12. Rigter, A., Langeveld, J. P., Timmers-Parohi, D., Jacobs, J. G., Moonen, P. L., and Bossers, A. (2007) Mapping of possible prion protein self-interaction domains using peptide arrays. *BMC Biochem*, **8**, 6.
13. Geysen, H. M., Meloen, R. H., and Barteling, S. J. (1984) Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. *Proc Natl Acad Sci USA*, **81**, 3998–4002.
14. Piccardo, P., Langeveld, J. P., Hill, A. F., Dlouhy, S. R., Young, K., Giaccone, G., Rossi, G., Bugiani, M., Bugiani, O., Meloen, R. H., Collinge, J., Tagliavini, F., and Ghetti, B. (1998) An antibody raised against a conserved sequence of the prion protein recognizes pathological isoforms in human and animal prion diseases, including Creutzfeldt–Jakob disease and bovine spongiform encephalopathy. *Am J Pathol*, **152**, 1415–20.
15. Slootstra, J. W., Puijk, W. C., Ligtvoet, G. J., Langeveld, J. P., and Meloen, R. H. (1996) Structural aspects of antibody–antigen interaction revealed through small random peptide libraries. *Mol Divers*, **1**, 87–96.
16. Timmerman, P., Van Dijk, E., Puijk, W., Schaaper, W., Slootstra, J., Carlisle, S. J., Coley, J., Eida, S., Gani, M., Hunt, T., Perry, P., Piron, G., and Meloen, R. H. (2004) Mapping of a discontinuous and highly conformational binding site on follicle stimulating hormone subunit-beta (FSH-beta) using domain scan and matrix scan technology. *Mol Divers*, **8**, 61–77.
17. Goldmann, W., Hunter, N., Martin, T., Dawson, M., and Hope, J. (1991) Different forms of the bovine PrP gene have five or six copies of a short, G-C-rich element within the protein-coding exon. *J Gen Virol*, **72(Pt 1)**, 201–4.

Chapter 13

A Designed Peptide Chip: Protein Fingerprinting Technology with a Dry Peptide Array and Statistical Data Mining

Kenji Usui, Kin-ya Tomizaki, and Hisakazu Mihara

Abstract

There has recently been increased interest in the potential for microarray technologies to study protein networks in a whole cell system within a single experiment. Protein-detecting microarrays are composed of numerous agents immobilized within a tiny area on solid surfaces to capture targeted proteins and to detect interactions in a high-throughput fashion. In this chapter, in order to extend the usability of peptide microarrays, we describe a novel dry peptide microarray format to obtain protein fingerprint (PFP) data sets and a statistical PFP data manipulation technique to quantitatively analyze targeted proteins.

Key words: Protein-detecting chip, designed peptide, dry peptide array, fluorimetric assay, statistical data mining.

1. Introduction

In the past decade, protein-detecting microarray technology has been widely studied as a promising tool for understanding interactions among biomolecules (1–5). The microarray is composed of numerous agents immobilized onto solid supports such as glasses and membranes (sometimes embedded in gels). Capturing agents that selectively bind to targeted proteins, such as proteins, peptides, antibodies, small molecules, or carbohydrates, are immobilized onto solid surfaces. This technology allows us to obtain important information on molecular interactions for a whole cell lysate within a single experiment.

The use of peptides as capturing agents for the preparation of protein-detecting microarrays has the following advantages: (i) peptides can be designed to form secondary structures such as α-helices, β-strands, β-sheets, or β-turns by arranging amino acid

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_13,

sequences by request; (ii) peptides can be prepared in large quantities by well-established chemical synthetic procedures; (iii) peptides are stable against oxidative degradation and dryness compared with proteinic capture agents; and (iv) peptides can be site-selectively modified with functional groups such as unnatural amino acid residues and fluorophores.

In general, peptide microarrays have been employed as substrates to detect enzymatic activities and to screen for inhibitors of targeted enzymes because peptides usually bind to proteins more weakly than antibodies, not enough to "fish" targeted proteins.

However, we have been focusing on the use of synthetic peptides as capturing agents for obtaining information about the interactions between a targeted protein and fluorophore-labeled synthetic peptides by a protein fingerprint (PFP) technique (6–12). The PFP is the pattern of changes in fluorescence intensity of the fluorophore-labeled synthetic peptides upon interacting with a targeted protein and allows us to discriminate target proteins from contaminants in sample solutions by comparing their PFP patterns. This technique does not require capturing agents such as antibodies that strongly and selectively bind to the targeted proteins.

In this chapter, in order to extend the usability of the peptide microarray, we describe the preparation of PFPs to discriminate targeted proteins with libraries of fluorophore-labeled synthetic peptides arrayed on a novel dry peptide array format which does not require covalent immobilization of peptides onto the supports, but is instead spotted onto the surfaces and dried in preparation. We also describe a technique for manipulating PFP data to statistically analyze protein–peptide interactions.

2. Materials

2.1. Construction of a Designed Peptide Library

1. Fmoc peptide synthesis (13) reagents: TentaGel S RAM resin is used as a Rink amide resin for peptide synthesis. The coupling cocktail contains Fmoc amino acids (Fmoc-AA-OH), 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), 1-hydroxybenzotriazole (HOBt), and diisopropylethylamine (DIEA) in *N*-methyl pyrrolidone (NMP) (*see* **Note 1**). Fmoc removal solution is 25% piperidine in NMP (v/v).
2. Peptide synthesizer: Advanced ChemTech Model 348 MPS.
3. Mtt (4-methyltrityl) protecting group removal solution: dichloromethane/triisopropylsilane/trifluoroacetic acid (TFA) (94/5/1, v/v).

4. Fluorophore coupling reagents: 5-(and-6)-carboxytetramethylrhodamine (TAMRA), *O*-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU), and DIEA in NMP (*see* **Note 2**).
5. Cleavage solution: TFA/H_2O (20/1, v/v) (*see* **Note 3**).
6. HPLC: HPLC is performed on the Hitachi L7000 or the Shimadzu LC2010C system using a Wakosil 5C18 or a YMC-Pack ODS-A (4.6 × 150 mm) for analysis, and a YMC ODS A323 (10 × 250 mm) for preparative purification with a linear gradient of acetonitrile/0.1% TFA at a flow rate of 1.0 ml/min for analysis and 3.0 ml/min for preparative separation, respectively.
7. Amino acid analysis is carried out using a Wakopak WS-PTC column (4.0 × 200 mm; Wako Pure Chemical Industries) after hydrolysis in 6 M HCl at 110°C for 24 h in a sealed tube followed by phenyl isothiocyanate labeling.
8. MALDI-TOFMS is measured on a Shimadzu KOMPACT MALDI III with 3,5-dimethoxy-4-hydroxycinnamic acid as the matrix.
9. Peptide absorbance is measured on a Benchmark Multiplate Reader (Bio-Rad Laboratories) with a 490 nm filter using microtiter plates (Assay Plate, IWAKI) in 50 mM KH_2PO_4 containing 5.0 M guanidine hydrochloride (pH 9.0).

2.2. Dry Peptide Array System and Assay

1. Peptide solution: 2.0 μM peptide in H_2O (*see* **Note 4**).
2. Flat glass plate for the array: SuperAmine Microarray substrates (TeleChem International).
3. Microarrayer: SpotBot Personal Microarrayer (TeleChem International) (**Fig. 13.1A**).

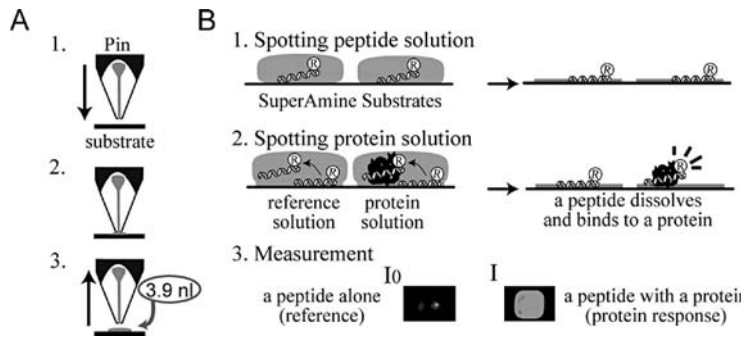


Fig. 13.1. (**A**) Illustration of printing mechanism using a Micro Spotting Pin of SpotBot (CMP10, TeleChem International) for sample spotting. (**B**) Scheme of the dry peptide array system. Reproduced from (6) with permission from the Royal Society of Chemistry.

4. Protein solution: various concentrations of proteins of interest with 100 μM $CaCl_2$ in H_2O (*see* **Note 5**).
5. Reference solution: 100 μM $CaCl_2$ in H_2O (*see* **Note 5**).
6. Fluorescence intensities of the fluorescent-labeled peptides in nano-spotted plates are measured by a CRBIO IIe (Hitachi Software Engineering) at room temperature; 532 and 573 nm filters are used for excitation and emission, respectively.

2.3. Generation of PFP from Raw Data and Data Analyses from PFPs

1. Data sets obtained from the microarray assay.
2. PC applications: a spreadsheet application, a text editor, a graphic viewer, and data analysis and graphing software (*see* **Note 6**).

3. Methods

To obtain information on protein function and structure, as well as binding properties, the designed peptide microarray is constructed. First, we should choose an appropriate peptide library that can detect proteins of interest properly. The synthetic strategy is then optimized to acquire highly pure library peptides. After synthesizing the peptide library, these library peptides should be arrayed in nanoliter amounts by the dry peptide array method (6). The dry peptide array system is a novel system which contains an array preparation and assay procedures with dried samples including non-immobilized capturing agents (designed peptides) (**Fig. 13.1**). The system has the following advantages: (1) peptides are tolerant of dry conditions; (2) the array is easily portable; (3) the preparation and assay procedures are much easier than those of immobilization methods; (4) target proteins need not be labeled, since peptide capture agents are fluorescently labeled; (5) plate washing processes are not required; and (6) vaporization of the sample solution need not be taken into account for nanoliter solution measurements. In **Sections 3.1** and **3.2**, protocols for the above are described in detail.

Upon adding various proteins to these peptide arrays, the library peptides with fluorescent probes for protein detection will show different responses against different proteins in fluorescent increments depending on the peptide sequences. The patterns of these responses can be regarded as "protein fingerprints" (PFPs), which establish the identity of target proteins (7–10). In addition, the PFPs correlate with the recognition properties of the target proteins to peptides. Furthermore, the data generated from PFPs have been

analyzed using several statistical methods such as the hierarchical clustering analysis (HCA) and the principal component analysis (PCA) (11). These PFP data analysis methods enable high-throughput peptide array-based analyses of protein functions and structural properties as well as binding properties. These protocols are described in detail in **Sections 3.3** and **3.4**.

3.1. Construction of a Designed Peptide Library

1. The peptide library is designed. We have recently constructed designed peptide libraries with loop (7), α-helical (8, 12), or β-sheet (9) structures and also constructed these designed peptide libraries with functional groups such as an alkyl chain (9) or sucrose (10). A suitable secondary structure and/or functional group can be selected for effective detection and analysis of proteins of interest. For example, the loop library (7) is composed of peptides having random sequences in turn structures, so it can effectively detect proteins that preferably bind to turn structures. The α-helical libraries (8, 12) consist of peptides with systematically changed sequences. Consequently, proteins such as EF-hand motif families can be powerfully monitored. Also, if a functional group is added to the peptides, the library may distinguish proteins that interact with that functional group.
2. The fluorescent detection method is selected. This choice may influence the synthetic strategy. For example, we have already shown that a dual-labeled α-helical peptide library gave efficient FRET responses in detection (8). However, the synthetic strategy was more complicated than that of single-labeled peptides.
3. The designed library is synthesized. The synthetic strategies are different depending on which library you choose. We describe a mini-library consisting of 20 designed α-helical peptides with various charges and/or hydrophobicities (**Fig. 13.2A**,**B**) to detect a variety of proteins including calmodulin (CaM) as reported in the literature (10) as an example (*see* **Note 7**).
4. The designed peptide library sequences (Ac-Lys(Mtt)-[various sequences]-Gly-resins) are synthesized on Rink amide resin by a combination of automatic and manual syntheses with Fmoc chemistry. Initially, Fmoc-Gly-resin is manually synthesized with Fmoc-Gly-OH (1/3 eq.) using HBTU (1/3 eq.), HOBt (1/3 eq.), and DIEA (2/3 eq.) to reduce the loading amount of resin (0.25 mmol/g). Then, Ac-Lys(Mtt)-[various sequences]-Gly-resins are synthesized using the synthesizer with Fmoc-AA-OH (10 eq.), HBTU (10 eq.), HOBt (10 eq.), and DIEA (12 eq.).

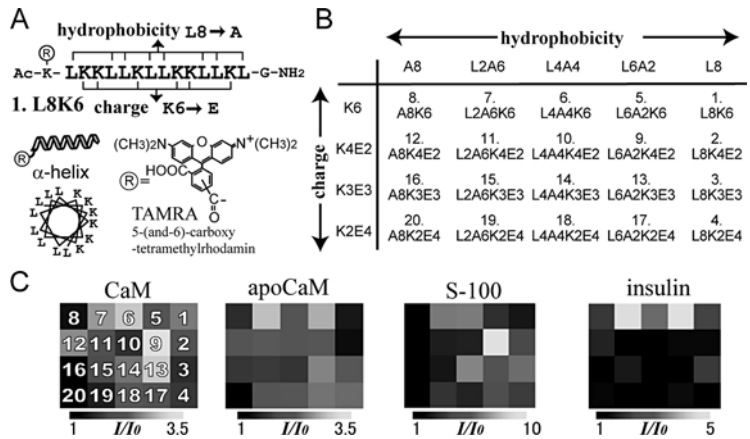


Fig. 13.2. (**A**) Strategy for the construction of the α-helical peptide mini-library, showing the sequence of No. 1 L8K6 peptide and construction of the α-helical peptide mini-library (*upper side*), diagrams of the α-helix structure with a wheel representation of L8K6 (*lower-left side*), and the structure of TAMRA (*lower-right side*) (*see* **Note 7**). (**B**) Numbers and names of peptides in the α-helical peptide mini-library (*see* **Note 7**). (**C**) The "protein fingerprints" (PFPs) of calmodulin (CaM), apoCaM (CaM with 150 pmol EDTA), S-100 proteins (S-100), and insulin (75 ng, 4.5 pmol). In this figure, the obtained PFPs were converted into gray scale PFPs. Reproduced from (6) with permission from the Royal Society of Chemistry.

5. The fluorophore is attached to the peptides. The Mtt group is removed by treatment with the Mtt removal solution (2 min × 8 times). Then, the TAMRA moiety is introduced by using TAMRA (1.5 eq.) with HATU (1.5 eq.) and DIEA (3 eq.).
6. The peptides with fluorescent probes are cleaved from the resin and the side chain protections are removed with the cleavage solution for 2 h. The peptides are precipitated by the addition of diethylether and collected by centrifugation.
7. The crude peptides are purified by RP-HPLC (*see* **Note 8**) and characterized by MALDI-TOFMS and amino acid analysis. The purified samples are lyophilized.
8. The purified peptides are dissolved in methanol at ca. 1 mM and stored at 4°C. In order to estimate the concentration of the stock solution of the library peptides, fluorophore absorbance of the library peptides can be measured.

3.2. Dry Peptide Array System and Assay

1. The following protocol is illustrated (*see* **Fig. 13.1**).
2. The peptide solution (3.9 nl) is spotted onto flat glass plates using the microarrayer.

3. The spotted solution is allowed to dry for a few minutes.
4. The protein solution (3.9 nl) and the reference solution (3.9 nl) are arrayed on the spots of the peptides.
5. The spotted solution is again allowed to dry for a few minutes.
6. The fluorescence intensities of the spots containing peptide alone (I_0) and the spots containing peptide and protein (I) are measured. The peptide response to the proteins is expressed using the intensity ratio, I/I_0.

3.3. Generation of PFP from Raw Data

1. The raw data is converted to PFP images. If you have an appropriate data analysis application and graphing software, you may skip the following protocol. In this section, we describe the manual data conversion (14) and the concept.
2. A color scale type is chosen. There are different kinds of color scale candidates such as black–red–yellow (warm colors), black–green–blue (cold colors), and gray scale. We used warm colors for increasing the data set intensity ratio ($I/I_0 > 1$) and/or cold colors for decreasing it ($I/I_0 < 1$).
3. The lower limits and upper limits of the color scale are defined. The color image is changed significantly depending on these values. Generally, the upper limit should be around the maximum value within the data set, and conversely, the lower limit should be around the minimum value. But there will be a certain amount of trial and error before an impressive PFP is obtained.
4. PFP dimensions and the library peptide distribution are determined. These also influence the characterization. For example, in the α-helical peptide library (8), 2D PFP, whose column heading and row heading represent, respectively, hydrophobicity and charge, was mostly applied because this library consists of designed α-helical peptides with systematically changed charges and/or hydrophobicities. Consequently, we may analyze the protein binding properties from the 2D PFPs.
5. Here is a description of the warm colors scale conversion using the dry array data (6) (**Fig. 13.2C**). The RGB color code numbers of the grid are saved as a comma-separated-value (.csv) file. Each grid position is first assigned three whole numbers corresponding to RGB color codes from (0, 0, 0) (full black, minimum value) to (255, 0, 0) (red) to (255, 255, 0) (yellow, maximum value), which correspond to the range from the lower limit to upper limit divided into 511

levels. Thus, a grid with X columns and Y lines is coded with $3X$ columns and Y lines of whole numbers between 0 and 255.

6. This file is then opened in a text editor and the following three (or four) lines are inserted at the top of the file:

   P3

   # (optional line with identifier)

   X Y

   255

   . . .where X is the number of columns in the image table and Y is the number of lines in the image table.

7. The file is saved from the text editor in the portable-pixelmap format by simply adding ".ppm" to the filename.
8. This file is then opened by graphic viewer software, resized, and saved in other formats such as the bitmap file format (.bmp).

3.4. Data Analyses from PFPs

1. Data analyses are conducted. In this section, we use the standard protocol described in the literature (11) as an example.
2. If the library size is big (>50 library peptides), you can select peptides for the analyses in order to make the analyses simpler and more effective. For example, as reported in the literature (11), we removed some peptides that show low responses to all proteins to filter out the background noise. Then the representative peptides (15 peptides), which had higher standard deviations, were selected (*see* **Note 9**).
3. The PFP is normalized before the Euclidean distance analyses are conducted. The normalization formula is as follows:

   $(P_{ori}-P_{ave})/d = P_{norm}$

   (P_{ori}, a value of a cell in a PFP; P_{ave}, the average of all values of a PFP; d, the standard deviation of all values of a PFP; P_{norm}, a normalized value of a cell in a PFP).
4. Euclidean distance is calculated (**Fig. 13.3A**,**B**). This is a common measure when considering the distance between two vectors. Similarity between the normalized PFP patterns is measured by Euclidean distance in multidimensional space defined by each PFP. These can be represented by color coding (yellow for the highest similarity and black for the lowest) as described in the previous section.
5. The HCA among the normalized PFPs is conducted (**Fig. 13.3C,D**). Ward's clustering algorithm is used and the dendrogram is obtained with the analyses of Euclidean distances using a statistical processing application.

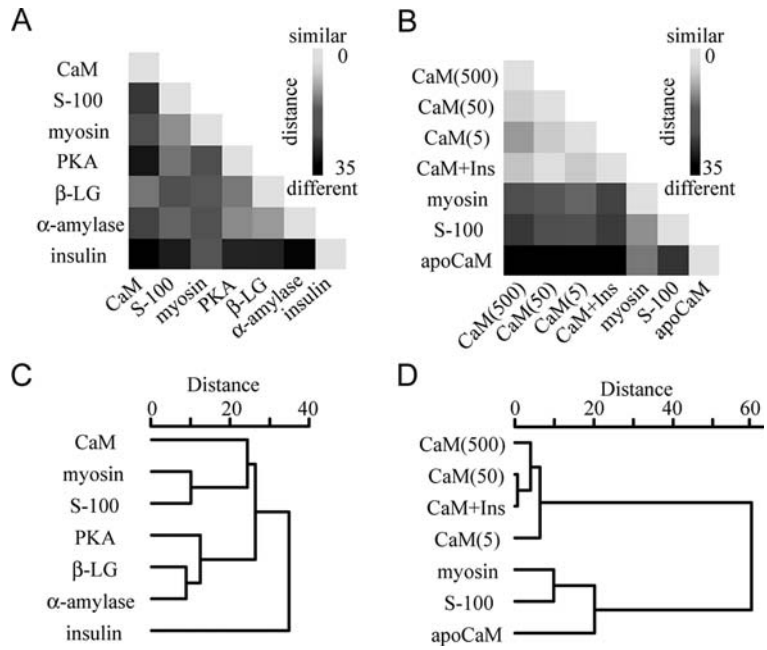


Fig. 13.3. (**A** and **B**) Euclidean distance matrices among various PFPs with selected peptides. (**A**) PFP differences among various proteins (CaM, S-100, myosin, protein kinase A (PKA), β-lactoglobulin (β-LG), α-amylase, and insulin). (**B**) PFP differences among EF-hand proteins and CaM at various concentrations. Numbers with CaM denote the concentrations, 500, 50, and 5 μg/ml (these are solution assay results). "CaM+Ins" denotes a mixture of CaM (50 μg/ml) and insulin (50 μg/ml). In this figure, the obtained matrices were converted into gray scale. (**C** and **D**) The clustering dendrograms of various PFPs. (**C**) Various proteins. (**D**) EF-hand proteins and CaM at various concentrations generated by analysis of Euclidean distances with selected peptides. The horizontal axis indicates the distances among PFPs (left for PFPs with the highest similarity and right for PFPs with the lowest similarity). All analyses showed (1) proteins within an EF-hand family are clustered successfully; (2) the PFP pattern analysis can be conducted regardless of the concentration of a protein sample; (3) concentration of a protein sample can be evaluated from the fluorescence intensities; and (4) structural change could be monitored by PFPs, as with the apoCaM PFP being completely distinguishable from CaM PFPs (with Ca^{2+}) (Euclidean distance > 30). Reproduced from (11) with permission from the Royal Society of Chemistry.

The horizontal axis represents the distance among normalized PFPs (left for PFPs with the highest similarity and right for PFPs with the lowest similarity).

6. Data manipulation using PCA is conducted (**Fig. 13.4**). PCA is a dimension reduction technique, i.e., PCA reduces the number of variables (features) to a more manageable size. In the previous study, the Varimax rotation algorithm was used.

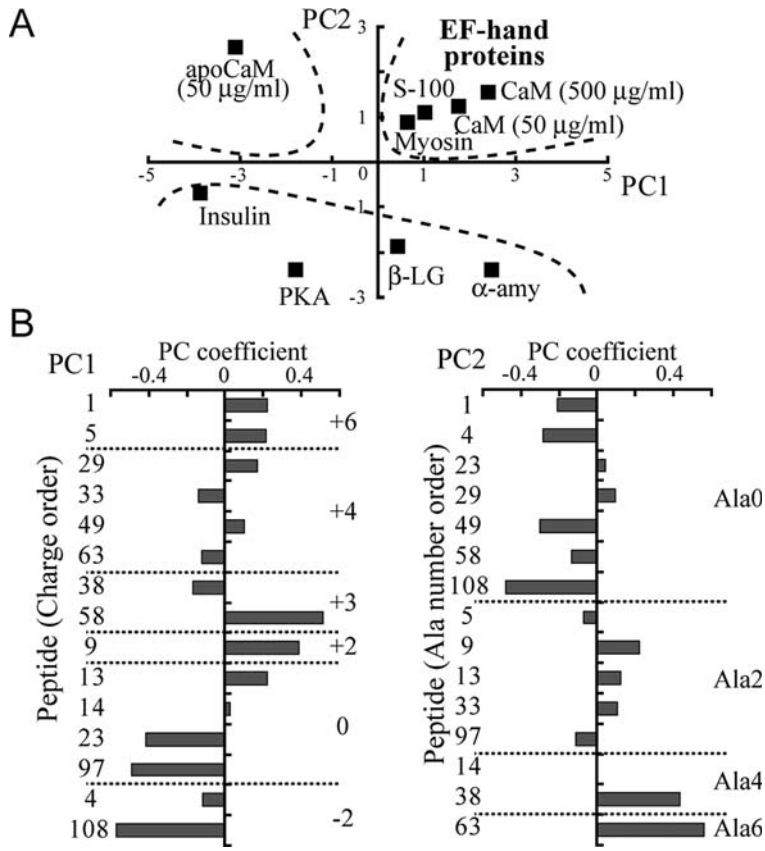


Fig. 13.4. (**A**) Projection of proteins according to principal components PC1 and PC2. Three groups are identified visually (*dashed lines*). In this analysis, EF-hand proteins were also grouped successfully and apoCaM PFP was distinguished from EF-hand family proteins. These results were highly correlated with the results of the HCA (**Fig. 13.3C**). (**B**) Coefficients per selected peptides of the first principal components (PC1 and PC2), which account for 33% (PC1) and 20% (PC2) of observed variability. Numbers represent peptide numbers. Positive coefficients represent positive effects on the PC axis of the score plots. Conversely, negative numbers represent negative effects on the PC axis, and "0" represents non-effects. The analysis showed that the pattern of PC1 coefficients reflected positive charges of peptides and that the pattern of PC2 coefficients reflected numbers of Ala, i.e., hydrophobicity (lower hydrophobicity with higher numbers of Ala). Reproduced from (11) with permission from the Royal Society of Chemistry.

4. Notes

1. These are standard peptide synthetic reagents. If necessary, HBTU–HOBt can be replaced with stronger reagents such as HATU depending on the library peptide sequences.

2. These are examples of fluorophore coupling reagents. The fluorophore and its coupling method are changeable depending on the strategy for the peptide microarray system.
3. Other scavengers such as triisopropylsilane should be added if this does not work well for the library peptides.
4. This is just an example of peptide concentration. The concentration should be determined depending on what is needed to detect fluorescence. The concentration can be made much lower after the optimization of the detection.
5. A solution that does not contain any salts is preferable because salt may be precipitated under dry conditions which may influence the assay. In this case, because the proteins of interest include CaM, which needs Ca^{2+} to bind to peptides, $CaCl_2$ is added to the protein and reference solutions.
6. All of these are not absolutely necessary to obtain the PFPs. Having a standard spreadsheet application, a standard text editor, and a standard graphic viewer is enough to make the PFPs. To acquire many PFPs and to do all of the following analyses, data analysis software and graphing software are needed.
7. The strategy for the construction of the α-helical peptide mini-library is shown in **Fig. 13.2A,B**. On the basis of the L8K6 sequence, which is a cationic amphiphilic α-helical peptide with the sequence LKKLLKLLKKLLKL (15) and is known to bind to CaM in the presence of Ca^{2+} (16, 17), the eight Leu residues were replaced systematically with Ala, and the six Lys residues were replaced systematically with Glu. This mini-library was a subset of the α-helical peptide library described previously (8).
8. This step can be skipped if the library size is too big and the synthetic protocol is optimized enough to obtain highly pure peptides. In this case, we recommend that instead of HPLC purification, the crude peptides should be quickly purified by gel permeation chromatography using Sephadex LH-20 (Amersham Biosciences KK) swelled with methanol in order to remove the byproducts and reagents from the cleavage step.
9. How to select the peptides (11) is briefly described as follows. The similarities among the responses of each peptide upon addition of the proteins were analyzed quantitatively by the HCA with Euclidean distances. As a result, the library could be sorted into eight groups (Groups A–H). The averages and standard deviations of the PFP values of each peptide against the seven proteins were also obtained. Three peptides, which were ranked in the first–third for highest values in standard deviations within a group, were chosen from each group (Groups A–E) and no peptides were chosen from Groups F–H, which have lower averages and standard deviations of the PFP values for each peptide.

References

1. Kambhampati, D. (2003) Protein Microarray Technology. Wiley-VCH, Weinheim.
2. Fung, E. T. (2004) Protein Arrays: Methods and Protocols; Methods in Molecular Biology vol 264. Humana Press, New Jersey.
3. Tomizaki, K.-Y., Usui, K., Mihara, H. (2005) Protein-detecting microarrays: Current accomplishments and requirements. *ChemBioChem* **6**, 782–799.
4. Kodadek, T. (2001) Protein microarrays: Prospects and problems. *Chem. Biol.* **8**, 105–115.
5. Uttamchandani, M., Wang, J., Yao, S. Q. (2006) Protein and small molecule microarrays: Powerful tools for high-throughput proteomics. *Mol. BioSyst.* **2**, 58–68.
6. Usui, K., Tomizaki, K.-Y., Ohyama, T., Nokihara, K., Mihara, H. (2006) A novel peptide microarray for protein detection and analysis utilizing a dry peptide array system. *Mol. Biosyst.* **2**, 113–121.
7. Takahashi, M., Nokihara, K., Mihara, H. (2003) Construction of a protein-detection system using a loop peptide library with a fluorescence label. *Chem. Biol.* **10**, 53–60.
8. Usui, K., Takahashi, M., Nokihara, K., Mihara, H. (2004) Peptide arrays with designed alpha-helical structures for characterization of proteins from FRET fingerprint patterns. *Mol. Divers.* **8**, 209–218.
9. Usui, K., Ojima, T., Takahashi, M., Nokihara, K., Mihara, H. (2004) Peptide arrays with designed secondary structures for protein characterization using fluorescent fingerprint patterns. *Biopolymers* **76**, 129–139.
10. Usui, K., Ojima, T., Tomizaki, K.-Y., Mihara, H. (2005) A designed glycopeptide array for characterization of sugar-binding proteins toward a glycopeptide chip technology. *NanoBiotechnology* **1**, 191–200.
11. Usui, K., Tomizaki, K.-Y., Mihara, H. (2006) Protein-fingerprint data mining of a designed alpha-helical peptide array. *Mol. Biosyst.* **2**, 417–420.
12. Usui, K., Tomizaki, K.-Y., Mihara, H. (2007) Screening of alpha-helical peptide ligands controlling a calcineurin-phosphatase activity. *Bioorg. Med. Chem. Lett.* **17**, 167–171.
13. Chan, W. C., White, P. D. (2000) Fmoc solid phase peptide synthesis: A practical approach. New York, Oxford University Press.
14. Wahler, D., Badalassi, F., Crotti, P., Reymond, J. L. (2002) Enzyme fingerprints of activity, and stereo- and enantioselectivity from fluorogenic and chromogenic substrate arrays. *Chemistry* **8**, 3211–3228.
15. Cox, J. A., Comte, M., Fitton, J. E., DeGrado, W. F. (1985) The interaction of calmodulin with amphiphilic peptides. *J. Biol. Chem.* **260**, 2527–2534.
16. O'Neil, K. T., DeGrado, W. F. (1990) How calmodulin binds its targets: sequence independent recognition of amphiphilic alpha-helices. *Trends. Biochem. Sci.* **15**, 59–64.
17. Maulet, Y., Cox, J. A. (1983) Structural changes in melittin and calmodulin upon complex formation and their modulation by calcium. *Biochemistry* **22**, 5680–5686.

Section III

Technological Advancements

Chapter 14

Peptide Microarrays on Bisphenol A Polycarbonate

Vianney Souplet, Clément Roux, and Oleg Melnyk

Abstract

Microarrays are frequently prepared on microscope glass slides. However, glass substrates can break or cut and thus can lead to the contamination of the manipulator during the analysis of biological samples. Alternately, bisphenol A polycarbonate (PC) is shock-resistant and, in addition, is easily eliminated by incineration. We show here that PC is a useful substrate for peptide microarray preparation. We describe in particular the preparation of peptide microarrays on PC using semicarbazide-functionalized silica nanoparticles and in situ semicarbazone ligation with glyoxylyl-peptides. The microarrays were used for the detection of antibodies using fluorescence detection.

Key words: Peptide microarrays, bisphenol A polycarbonate, silica, nanoparticles, semicarbazide, alpha-oxo aldehyde, glyoxylyl, semicarbazone.

1. Introduction

Peptide microarrays are often used for the specific detection of proteins such as antibodies (1–8). The parallel detection of antibodies in biological samples has a wide range of potential applications in the diagnosis of allergies, autoimmune and infectious diseases as well as in epitope-mapping studies and the development of vaccines (9–12). They are also used for the serodetection of infections (13–16). Microarrays are frequently prepared on microscope glass slides or silica-based materials. However, glass substrates can break or cut and thus can lead to the contamination of the manipulator during the analysis of biological samples. The widespread use of microarrays for diagnostic applications needs the development of safer substrates.

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_14,

We show here that PC can be used instead of microscope glass slide for peptide microarray preparation. Bisphenol A polycarbonate (PC) is shock-resistant, cheap and can be easily eliminated by incineration. We describe in particular the preparation of peptide microarrays on PC using semicarbazide-functionalized silica nanoparticles and in situ semicarbazone ligation with glyoxylyl-peptides. The microarrays were used for the detection of antibodies using fluorescence detection (17–19).

2. Materials

2.1. Preparation of Silica Nanoparticles

1. 9-Fluorenylmethoxycarbonyl–$NHNH_2$ (Fmoc–$NHNH_2$, *see* **Note 1**)
2. Absolute ethanol (EtOH, ACROS Organics, Geel, Belgium)
3. 3-(Triethoxysilyl)propyl isocyanate (Sigma-Aldrich)
4. Freshly distilled dichloromethane (CH_2Cl_2, ACROS Organics)
5. Pentane (ACROS Organics)
6. Ludox AS-40 silica nanoparticles (40% w/v aqueous suspension, Sigma-Aldrich)
7. Benzyltrimethylammonium hydroxide solution at 40% v/v in MeOH (Triton B, Sigma-Aldrich)
8. *N,N′*-dimethylformamide (DMF, JT Baker, Phillisburg, USA)
9. Piperidine (ACROS Organics)
10. Dowex 1 × 4 ion-exchange resin (100–200 Mesh, Cl^-, Fluka)
11. Dowex 50 W × 4 ion-exchange resin (100–200 Mesh, H^+, Fluka)
12. Ammonia 6% v/v in water (Sigma-Aldrich)
13. Methanol (MeOH, ACROS Organics)

2.2. Preparation of PC Slides

1. Polycarbonate sheets 300 × 300 × 1 mm (Goodfellow, Huntingdon, UK)
2. Absolute ethanol (ACROS Organics)

2.3. Printing

1. Semicarbazide silica nanoparticles (2% w/v in DMF), 20 µL for each CHOCO–peptide
2. CHOCO–peptides (*see* **Note 3**)
3. 5 mM Sodium acetate buffer (pH 5.5)

4. V-bottomed 96-well microtitre plates (ABgene, Surrey, UK, AB-1058)
5. BCA1 Microarrayer (Perkin Elmer, MA, USA)

2.4. Incubations and Scanning

1. 0.01 M Sodium phosphate buffer pH 7.2 containing 150 mM NaCl (PBS)
2. Tween® 20 (Sigma-Aldrich)
3. Bovine serum albumin (BSA)
4. Primary antibodies: anti-HA murine antibody (HA.11 from ascites, Eurogentec, Belgium), anti-FLAG® M2 murine antibody (F 1804, affinity purified, Sigma-Aldrich, MI, USA), mouse IgG (ref 015-000-003, Jackson Immunoresearch, PA, USA)
5. Tetramethylrhodamine-labelled goat anti-murine antibody (Sigma-Aldrich)
6. Absolute ethanol (EtOH, ACROS Organics)
7. Microplate microarray hardware 96 wells (4 × 24) (ref. MMH4 × 24) equipped with a silicone gasket (ref. GMMH4 × 24, TeleChem International, Inc. CA, USA). The microarray hardware can accept four slides
8. Standard incubator/agitator for 96-well microtitre plates (ref. PST-60HL-4, Biosan, MI, USA)
9. Standard 96-well microtitre plate washing station (wellwash AC, Thermo Electron, Finland)
10. Fluorescence microarray scanner (532 nm, Innopsys scanner, France) and Mapix® image analysis software (Innopsys, France)

3. Methods

Polycarbonate (PC) is a useful substrate for the preparation of microfluidic devices (20, 21). Recently, its utility for bioanalysis has attracted much attention owing to the possibility of using compact discs as platforms for the high-throughput analysis of biomolecular interactions (20, 22–24). PC is known to be degraded by aqueous bases, swelled or solubilized by various organic solvents. It cannot be exposed to temperatures exceeding 115°C. The method described here does not damage the PC surface despite the use of organic solvents such as DMF.

The method used for making peptide microarrays on PC is described in **Fig. 14.1**. Semicarbazide-functionalized silica nanoparticles (Ø 27 nm) are prepared as a suspension in DMF (*see* **Note 2**). This suspension is mixed with an equal volume of the CHOCO–

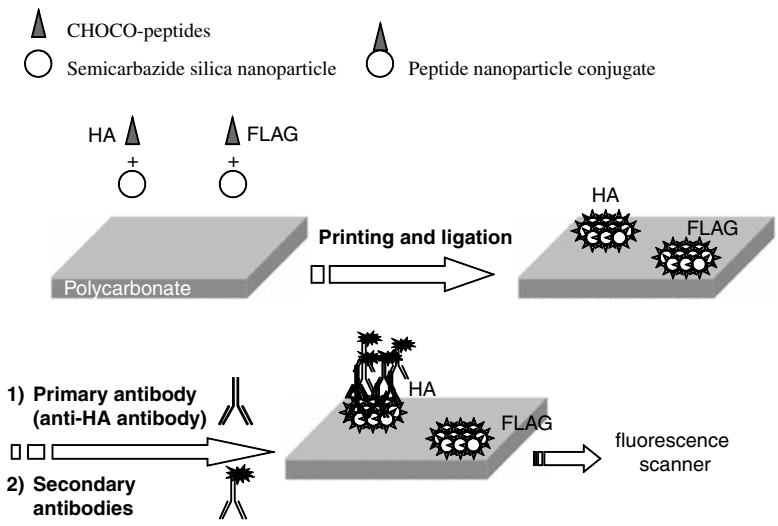
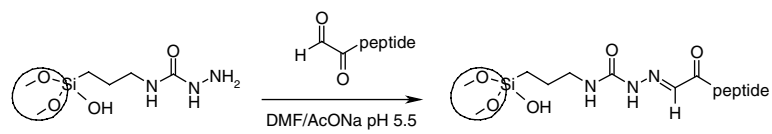


Fig. 14.1. Preparation of peptide microarrays on PC by in situ site-specific reaction of semicarbazide-functionalized silica nanoparticles with CHOCO–peptides (COCHO–HA: CHOCO–GYPYDVPDYAGYPYDVPDYAGYPYDVPDYAS–NH_2, COCHO–FLAG: CHOCO–DYKDH DGDYKDHDIDYKDDDDKGGS–NH_2). Semicarbazide-functionalized Ø 27 nm nanoparticles in suspension in DMF are mixed with CHOCO–peptides dissolved in pH 5.5 acetate buffer and printed on PC. In the model study described here, the microarrays were incubated with primary murine antibodies and then tetramethylrhodamine-labelled anti-murine goat secondary antibody. Finally, the PC slides are scanned using a standard microarray fluorescence scanner at 532 nm.

peptide (*see* **Note 3**). We used CHOCO–HA and COCHO–FLAG peptides in this model study (*see* **Note 4**). The mixture is printed on PC substrates using a microarrayer (*see* **Note 5**). Ligation results in the formation of an alpha-oxo semicarbazone bond –NHCONHN=CHCO– between the semicarbazide group –$NHCONHNH_2$ on the nanoparticles and the glyoxylyl group CHOCO– on the peptide (25–27) (Scheme **14.1**). Once the microarray is printed, it can be used for the detection of antibodies using standard protocols. The surface between the spots is left unchanged. In the work presented here, the microarrays were incubated with commercially available murine anti-HA, anti-FLAG® M2 monoclonal antibodies or murine IgG used as negative control. We used a tetramethylrhodamine-labelled anti-murine goat secondary antibodies to reveal captured primary antibodies. Detection was achieved using a standard fluorescence scanner. The results are presented in **Fig. 14.2**.



Scheme 14.1. α-Oxo semicarbazone ligation between semicarbazide groups on the silica nanoparticles and a glyoxylyl peptide.

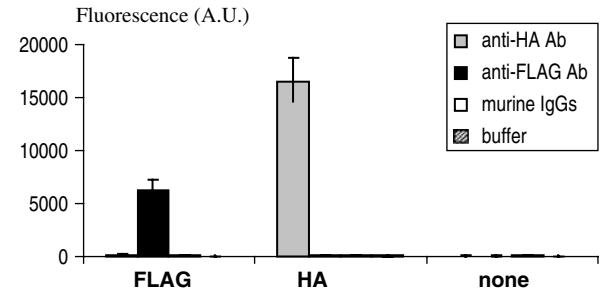


Fig. 14.2. Peptide patterns on PC (FLAG or HA peptides printed at 10^{-5} M in DMF/5 mM sodium acetate pH 5.5 buffer, three drops/spot, 1 nL overall, peptides printed in triplicate) were incubated in triplicate with anti-FLAG®, anti-HA, murine IgGs (10 μg/mL) or the buffer (PBS-B) at 37°C during 1 h and then with tetramethylrhodamine-labelled goat anti-murine antibodies (10 μg/mL, 1 h, 37°C). Incubations and washings were performed in a microplate microarray hardware (four PC slides, 96 wells). Slides were scanned at 532 nm and 16-bit images were analysed with the Mapix® software. Data represent the median and inter-quartile range after background subtraction.

3.1. Preparation of Silica Nanoparticles

1. Suspend Fmoc–$NHNH_2$(2 g, 7.86 mmol) in 60 mL of EtOH in a round-bottomed flask equipped with a septum, and flush with argon. Add 3-(triethoxysilyl)propyl isocyanate (2.28 mL, 8.80 mmol) using a syringe and a needle, and the reaction mixture is stirred for 1 h. The completion of the reaction is indicated by the disappearance of solid Fmoc–$NHNH_2$. After 1 h, concentrate the reaction mixture under reduced pressure to obtain a white solid. Dissolve the solid in 2 mL of freshly distilled CH_2Cl_2, and precipitate by adding 50 mL of dry pentane. The suspension is filtered on a sintered funnel and dried in vacuum, yielding 3.1 g of silane 1 $(EtO)_3$-$Si(CH_2)_3NHCONHNH$–Fmoc as a white solid (78% yield). ^1H NMR (300 MHz, $CDCl_3$ *see* **Note 6**): d=0.66 (t, J=8.1 Hz, 2H), 1.24 (t, J=6.7 Hz, 9H), 1.6 (m, J=6.7 Hz, 2H), 3.26 (q, J=6.6 Hz, 2H), 3.82 (q, J=7 Hz, 6H), 4.25 (t, J=6.82 Hz, 1H), 4.51 (d, J=6.79 Hz, 2H), 5.46 (m, 1 H), 6.46 (s, 1 H), 6.71 (s, 1 H), 7.34 (t, J=7.61 Hz, 2H), 7.43 (t, J=7.01 Hz, 2 H), 7.61 (d, J=7.28 Hz, 2H), 7.79 (d, J=7.39 Hz, 2 H) (*see* **Note 6**).
2. Rinse 1.5 g of Dowex 1 × 4 resin (Cl^-) and 1.5 g of Dowex 50 W × 4 resin (H^+) with 5 mL of water (*see* **Note 7**) and 15 mL of MeOH (*see* **Note 8**). Transfer the 1 × 4 resin to a stoppered glass Erlenmeyer containing 200 mL of Triton B. After 2 h of vigorous shaking, filter the resin and shake with another 200 mL of fresh Triton B solution. Filter the resin and rinse with 2 × 150 mL of water. Meanwhile, rinse the 50 W × 4 resin with 3 mL of ammonia (6% in water) and 20 mL of water. Once both resins are ready, Ludox AS-40 silica nanoparticles (3 mL, initial concentration ~40% w/v, *see*

Note 9) are diluted by dropwise addition of 3 mL of water under continuous shaking. The suspension is then filtered successively through the 1 × 4 resin and the 50 W × 4 resin (*see* **Note 10**). The concentration of the resulting solution is usually close to 20% w/v.

3. Dilute this 20% w/v colloidal suspension with EtOH so as to obtain a 1% w/v suspension, and further dilute it with enough DMF (*see* **Note 11**) to obtain a 0.5% w/v suspension. Transfer the suspension to an argon-flushed distillation set-up, and heat to 150°C for 1 h to evaporate water and EtOH. Allow the suspension to cool to room temperature.

4. Dissolve silane 1 (1.725 g, 3.44 mmol) in a solution of piperidine at 10% v/v in EtOH (40 mL). After 15 min stirring, add the silane solution to the colloidal suspension in the distillation unit, and heat the mixture to 150°C for 2 h under vigorous stirring. Allow the suspension to cool to room temperature. The colloidal suspension is split into four equal parts and centrifuged in 50 mL centrifuge tubes at 21,000*g* for 30 min at 10°C. Replace the supernatant by the same amount of fresh DMF. Resuspend the nanoparticles by alternating vortexing and sonication (*see* **Note 12**). The purification process is repeated three times. The nanoparticles are stored in a hermetical polypropylene container (*see* **Note 13**) at 4°C until use. The density of the suspension is estimated as described above, and is usually close to 1% w/v.

The nanoparticle suspension can be concentrated at will by centrifugation and resuspension in a lower volume of DMF.

3.2. Preparation of PC Slides

Commercially available polycarbonate sheets 300 × 300 × 1 mm (Goodfellow, Huntingdon, UK) are cut by high-pressure water jet at the size of microscope glass slides 75 × 25 × 1 mm (Economos, Dunkerque, France). Polycarbonate slides are washed twice with deionized water and absolute ethanol, respectively, before use (*see* **Note 14**).

3.3. Printing

1. Prepare nanoparticle–CHOCO–peptide mixtures in microtitre plates by mixing semicarbazide silica nanoparticles with CHOCO–peptides dissolved at 20 μM in acetate buffers. Add the peptide solution to the nanoparticles. The final peptide concentration is 10 μM.

2. Print the mixtures (three drops, 300 pL each, three replicates per peptide) on PC slides using the microarrayer. Up to 24 patterns can be printed per PC slide. The position of these patterns must be carefully controlled to allow the correct positioning of the patterns in the microplate microarray hardware 96 wells (*see* **Section 3.4**).

3. Incubate the slides overnight in an oven at 37°C at 60% relative humidity to favour ligation of the peptides to the nanoparticles.
4. Printed slides are stored under nitrogen atmosphere at room temperature until use.

3.4. Incubations (see Note 15)

1. Prepare PBS-A buffer by dissolving 0.05% v/v Tween®20 in PBS.
2. Prepare PBS-B buffer by dissolving 2% w/v BSA in PBS.
3. Insert the PC slides in the microplate microarray hardware 96 wells and place the assembled plate in the incubator at 37°C for 15 min.
4. Use three patterns per primary antibody. The patterns are first saturated 5 min at 37°C in PBS-B (150 µL per well).
5. In the meantime, dilute primary antibody in PBS-B (typically from 10 µg/mL to 10 ng/mL).
6. Wash the microarrays three times with PBS-A and incubate immediately with the primary antibody (120 µL per well, 1 h at 37°C, *see* **Note 16**).
7. In the meantime, dilute secondary antibody in PBS-B (final concentration 10 µg/mL).
8. The microarrays are washed three times with PBS-A and incubated immediately with secondary antibody (120 µL per well, 1 h at 37°C).
9. Wash three times with PBS-A and water.
10. Remove the slides from the microplate hardware and wash the slides with absolute ethanol. Dry the slides under a flush of nitrogen.
11. Analyse with the microarray scanner at 532 nm.

4. Notes

1. Fmoc–$NHNH_2$ can be purchased from various suppliers, or easily prepared from Fmoc–chloroformate and hydrazine hydrate, according to (30).
2. In this work, the diameter of the silica semicarbazide nanoparticles was 27 nm. Other diameters can be used as well up to 300 nm. The influence of surface curvature on protein stability is a subject of intense research today and the advantage of the method is the possibility to adjust the diameter of the nanoparticles to the detection issues.

3. Experimental conditions were optimized to avoid aggregation of the nanoparticles after mixing. The best results in terms of nanoparticle stability were obtained at around 10 μM CHOCO–peptide concentration. Higher peptide concentrations lead usually to nanoparticle aggregation and sedimentation.
4. The use of a CHOCO–Pro motif at the N-terminus must be avoided due to unwanted cyclization (28). Lyophilized CHOCO–peptides are stable for months when stored at −20°C. CHOCO–peptides can be synthesized using solution or solid-phase protocols. CHOCO–peptides can be prepared by periodic oxidation of a serine or threonine residue present at the N-terminus of the peptide. This method was used for the preparation of CHOCO–HA and COCHO–FLAG peptides used in this study. However, periodate can damage cysteine, methionine or sometimes tryptophane residues depending on the experimental conditions. Clearly, it is usually difficult to avoid the partial oxidation of methionine (29). In this case, we recommend the use of masked forms of glyoxylic acid which can be coupled to the peptide chain on the solid phase, and which can be converted into CHOCO group using mild experimental conditions. The later method is compatible with the parallel and automated synthesis of COCHO–peptides as needed in peptide microarray projects.
5. The experiment described here was performed using a non-contact piezoelectric microarrayer (BCA1, Perkin Elmer). The parameters of the piezoelectric tips must be adjusted to allow the use of a DMF/water mixture. Water content lower than 50% in DMF is not compatible with this printing technique (and could damage the PC slides). Typically, three drops (300 pL each) were printed per spot. Other printing machines can be used as well. We have obtained excellent results using a Qarray2 contact microarrayer (Genetix, New Milton, UK). The extent of the doughnut effect due to the migration of the nanoparticles to the periphery of the spot seems to depend on the printing technique. The Qarray2 arrayer led to negligible doughnut effects compared to the BCA1 piezoelectric arrayer may be due to the smallest volume printed.
6. Care must be taken as to perform the analysis quickly after dissolution of the silane in $CDCl_3$ in order to avoid its hydrolysis in the NMR tube. Alternately, the analysis can be performed in DMF-d7.
7. Deionized water has a resistivity of 18 MΩcm.
8. Commercially available Ludox AS-40 nanoparticles contain traces of Na^+ and Cl^- ions. The thorough removal of these ions was crucial for the successful grafting of silane 1 on the silica nanoparticles.

9. The density of the colloidal suspension is estimated by evaporating under vacuum 100 µL aliquots of the suspension in tared microtubes and weighting the resulting solid.
10. High resistance is encountered when filtering the suspension through the Cl^- resin. This is normal and the resin bed must not be disturbed. On the contrary, the H^+ resin offers no such resistance.
11. In our experience, DMF other than that supplied by JT Baker has led to colloidal instability and sedimentation with time. It is, therefore, essential that fresh JT Baker DMF is used throughout the whole preparation process. The origin of this instability is unknown.
12. One cycle of resuspension of the nanoparticles can take up to 1 day. It is important to vortex the suspensions on a very frequent basis (e.g. every 10 min).
13. Although we have never checked, it is probable that any glassware that had been in contact with the colloidal suspension will be covered with an invisible layer of nanoparticles.
14. PC sheets are covered by a protective film which remains on the slides after the cutting step. This film must be removed before washing steps. Be careful when removing this film to avoid the bending and deformation of the slides.
15. It is of prime importance to avoid the drying of the microarrays during the incubations, otherwise high background levels will be obtained.
16. Incubating at 37°C rather than at 25°C accelerates antigen–antibody complex formation. The microarray hardware is metallic and thus problems of heat transfer between metallic parts of the hardware and plastic slides were sometimes observed. This leads to temperature heterogeneities within the device and thus to a greater variability of the results. To avoid this problem, it is better to perform the incubations at 25°C. For the automated washings, standard washing protocols are usually satisfactory.

Acknowledgements

This work was partly supported by the French Ministry of Research, grants from the EEC and from the Région Nord Pas-de-Calais (Puces nano3D, Fansbamed and Interpoly3D projects). We also used the microarray platform facilities from the Institute of Biology (Lille, France).

References

1. Glokler, J., and Angenendt, P. (2003) Protein and antibody microarray technology. *J Chromatogr B Analyt Technol Biomed Life Sci* **797**, 229–40.
2. Graham, K. L., Robinson, W. H., Steinman, L., and Utz, P. J. (2004) High-throughput methods for measuring autoantibodies in systemic lupus erythematosus and other autoimmune diseases. *Autoimmunity* **37**, 269–72.
3. Hueber, W., Kidd, B. A., Tomooka, B. H., Lee, B. J., Bruce, B., Fries, J. F., Sonderstrup, G., Monach, P., Drijfhout, J. W., van Venrooij, W. J., Utz, P. J., Genovese, M. C., and Robinson, W. H. (2005) Antigen microarray profiling of autoantibodies in rheumatoid arthritis. *Arthritis Rheum* **52**, 2645–55.
4. Neuman de Vegvar, H. E., Amara, R. R., Steinman, L., Utz, P. J., Robinson, H. L., and Robinson, W. H. (2003) Microarray profiling of antibody responses against simian–human immunodeficiency virus: postchallenge convergence of reactivities independent of host histocompatibility type and vaccine regimen. *J Virol* **77**, 11125–38.
5. Robinson, W. H., DiGennaro, C., Hueber, W., Haab, B. B., Kamachi, M., Dean, E. J., Fournel, S., Fong, D., Genovese, M. C., de Vegvar, H. E., Skriner, K., Hirschberg, D. L., Morris, R. I., Muller, S., Pruijn, G. J., van Venrooij, W. J., Smolen, J. S., Brown, P. O., Steinman, L., and Utz, P. J. (2002) Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat Med* **8**, 295–301.
6. Robinson, W. H., Fontoura, P., Lee, B. J., de Vegvar, H. E., Tom, J., Pedotti, R., DiGennaro, C. D., Mitchell, D. J., Fong, D., Ho, P. P., Ruiz, P. J., Maverakis, E., Stevens, D. B., Bernard, C. C., Martin, R., Kuchroo, V. K., van Noort, J. M., Genain, C. P., Amor, S., Olsson, T., Utz, P. J., Garren, H., and Steinman, L. (2003) Protein microarrays guide tolerizing DNA vaccine treatment of autoimmune encephalomyelitis. *Nat Biotechnol* **21**, 1033–9.
7. Robinson, W. H., Steinman, L., and Utz, P. J. (2002) Protein and peptide array analysis of autoimmune disease. *Biotechniques* **Suppl**, 66–9.
8. Robinson, W. H., Steinman, L., and Utz, P. J. (2003) Protein arrays for autoantibody profiling and fine-specificity mapping. *Proteomics* **3**, 2077–84.
9. Templin, M. F., Stoll, D., Schrenk, M., Traub, P. C., Vohringer, C. F., and Joos, T. O. (2002) Protein microarray technology. *Drug Discov Today* **7**, 815–22.
10. Talapatra, A., Rouse, R., and Hardiman, G. (2002) Protein microarrays: challenges and promises. *Pharmacogenomics* **3**, 527–36.
11. Ryan, T. E., and Patterson, S. D. (2002) Proteomics: drug target discovery on an industrial scale. *Trends Biotechnol* **20**, S45–51.
12. Xu, Q., and Lam, K. S. (2002) An efficient approach to prepare glyoxylyl functionality on solid-support. *Tetrahedron Lett.* **43**, 4435–4437.
13. Duburcq, X., Olivier, C., Malingue, F., Desmet, R., Bouzidi, A., Zhou, F., Auriault, C., Gras-Masse, H., and Melnyk, O. (2004) Peptide–protein microarrays for the simultaneous detection of pathogen infections. *Bioconjug Chem* **15**, 307–16.
14. Duburcq, X., Olivier, C., Desmet, R., Halasa, M., Carion, O., Grandidier, B., Heim, T., Stievenard, D., Auriault, C., and Melnyk, O. (2004) Polypeptide semicarbazide glass slide microarrays: characterization and comparison with amine slides in serodetection studies. *Bioconjug Chem* **15**, 317–25.
15. Mezzasoma, L., Bacarese-Hamilton, T., Di Cristina, M., Rossi, R., Bistoni, F., and Crisanti, A. (2002) Antigen microarrays for serodiagnosis of infectious diseases. *Clin Chem* **48**, 121–30.
16. Petrik, J. (2001) Microarray technology: the future of blood testing? *Vox Sang* **80**, 1–11.
17. Schaferling, M., and Nagl, S. (2006) Optical technologies for the read out and quality control of DNA and protein microarrays. *Anal Bioanal Chem* **385**, 500–17.
18. Brecht, A., and Gauglitz, G. (1995) Optical probes and transducers. *Biosens Bioelectron* **10**, 923–36.
19. Ekins, R., Chu, F., and Micallef, J. (1989) High specific activity chemiluminescent and fluorescent markers: their potential application to high sensitivity and 'multi-analyte' immunoassays. *J Biolumin Chemilumin* **4**, 59–78.
20. Soper, S. A., Hashimoto, M., Situma, C., Murphy, M. C., McCarley, R. L., Cheng, Y. W., and Barany, F. (2005) Fabrication of DNA microarrays onto polymer substrates using UV modification protocols with integration into microfluidic platforms for the sensing of low-abundant DNA point mutations. *Methods* **37**, 103–13.

21. Liu, J., Roussel, C., Lagger, G., Tacchini, P., and Girault, H. H. (2005) Antioxidant sensors based on DNA-modified electrodes. *Anal Chem* **77**, 7687–94.
22. La Clair, J. J., and Burkart, M. D. (2003) Molecular screening on a compact disc. *Org Biomol Chem* **1**, 3244–9.
23. La Clair, J. J., and Burkart, M. D. (2006) Geometry in digital molecular arrays. *Org Biomol Chem* **4**, 3052–5.
24. Alexandre, I., Houbion, Y., Collet, J., Hamels, S., Demarteau, J., Gala, J. L., and Remacle, J. (2002) Compact disc with both numeric and genomic information as DNA microarray platform. *Biotechniques* **33**, 435–6, 438–9.
25. Olivier, C., Perzyna, A., Coffinier, Y., Grandidier, B., Stievenard, D., Melnyk, O., and Durand, J.-O. (2006) Detecting the chemoselective ligation of peptides to silicon with the use of cobalt–carbonyl labels. *Langmuir* **22**, 7059–65.
26. Olivier, C., Hot, D., Huot, L., Ollivier, N., El-Mahdi, O., Gouyette, C., Huynh-Dinh, T., Gras-Masse, H., Lemoine, Y., and Melnyk, O. (2003) Alpha-oxo semicarbazone peptide or oligodeoxynucleotide microarrays. *Bioconjug Chem* **14**, 430–9.
27. Melnyk, O., Duburcq, X., Olivier, C., Urbes, F., Auriault, C., and Gras-Masse, H. (2002) Peptide Arrays for Highly Sensitive and Specific Antibody-Binding Fluorescence Assays. *Bioconjugate Chemistry* **13**, 713–720.
28. Rose, K., Chen, J., Dragovic, M., Zeng, W., Jeannerat, D., Kamalaprija, P., and Burger, U. (1999) New cyclization reaction at the amino terminus of peptides and proteins. *Bioconjug Chem* **10**, 1038–43.
29. Geoghegan, K. F., and Stroh, J. G. (1992) Site-directed conjugation of nonpeptide groups to peptides and proteins via periodate oxidation of a 2-amino alcohol. Application to modification at N-terminal serine. *Bioconjug Chem* **3**, 138–46.
30. Zhang, R. E., Cao, Y. L., and Hearn, M. W. (1991) Synthesis and application of Fmoc–hydrazine for the quantitative determination of saccharides by reversed-phase high-performance liquid chromatography in the low and subpicomole range. *Analytical biochemistry* **195**, 160–167.

Chapter 15

Self-Assembly of PNA-Encoded Peptides into Microarrays

Francois Debaene and Nicolas Winssinger

Abstract

Several technologies have been described to immobilize libraries of small molecules or peptides in a microarray format. Herein, we describe protocols for an alternative strategy whereby each small molecule or peptide within a library is labeled with a peptide nucleic acid (PNA) tag such that they self-assemble in a microarray format upon hybridization with readily available DNA arrays. An important asset of the method is that it allows the library to be used in solution prior to hybridizing and as such offers the opportunity to separate the inhibitors bound to the protein from the rest of the library. Two methods based on size exclusion filtration or gel electrophoresis to separate protein-bound inhibitors from the remaining library are described.

Key words: Microarray, peptide nucleic acid (PNA), proteases, multiplexed screening, combinatorial libraries.

1. Introduction

The success of DNA microarrays to measure the expression level of thousands of genes simultaneously has inspired researchers from other disciplines to embrace this powerful analytical format. A number of methods have been described to immobilize small molecules, including peptides, to the surface of the array chemoselectively (1). Two applications of small molecule microarrays have already yielded important results beyond the proof of concept: enzyme profiling and ligand discovery. In both cases, the microarrays are incubated with the enzyme or protein of interest. We have developed an alternative approach where the members of a library are labeled with a PNA tag which enables a library present in solution as a mixture to be converted to the microarray format by a simple hybridization to a DNA microarray (2).An important

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_15,

asset of this approach is that libraries can be used in solution prior to hybridization. Aside from minimizing potential problems associated with the display of ligands or substrates on a surface (unspecific interactions with the surface, high local concentration), it also allows separation of ligands that are bound to a protein from unbound ones prior to hybridization and as such offers a detection method which is not possible with covalently immobilized compounds. A second asset of the PNA tag is that it can be used to encode libraries prepared by mix-and-split combinatorial synthesis using a unique PNA codon for every building block in the library. Upon cleavage from the solid phase, the library is obtained as a mixture in solution; however, it sorts itself into an addressable microarray upon hybridization (**Fig. 15.1**).

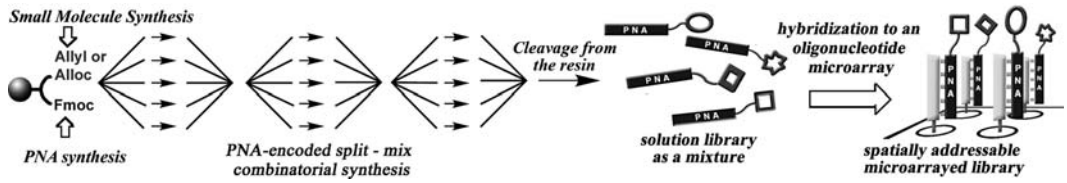


Fig. 15.1. Starting from a bifunctional linker such as a lysine with orthogonal protecting groups (such as Fmoc and Alloc), a library is prepared by mix-and-split combinatorial syntheses where a specific PNA codon is used to encode every amino acid or building block of the library. The library is then cleaved from the solid phase to obtain a mixture in solution which is converted to the microarray format by hybridization to an oligonucleotide microarray.

As the ligands bound to a protein can be separated from the rest of the library prior to hybridization, the ligand can be labeled with a fluorophore thus allowing the screen to be performed with crude protein extracts without the necessity of a secondary antibody or other means of detecting the protein on the microarray. The PNA-encoded strategy has been used to screen cysteine protease inhibitors directly from allergenic dustmite extracts (3) as well as to detect substrate specificities of purified enzyme samples (4). Identified inhibitors were shown to be sufficiently specific to discriminate closely related enzyme from the same family. For screens against crude proteomic mixtures, the inhibitor's target can be identified by affinity column coupled to mass spectrometry. Thus far, two approaches have been validated to separate protein–inhibitor–PNA adducts: a simple filtration through a size exclusion filter and gel electrophoresis. Both techniques require the use of denaturing conditions for efficient separation and are thus limited to inhibitors which form a covalent bond to their target. The advantage of the gel separation is that the formation of a protein–inhibitor complex is clearly visible by band shift on the gel (**Fig. 15.2**).

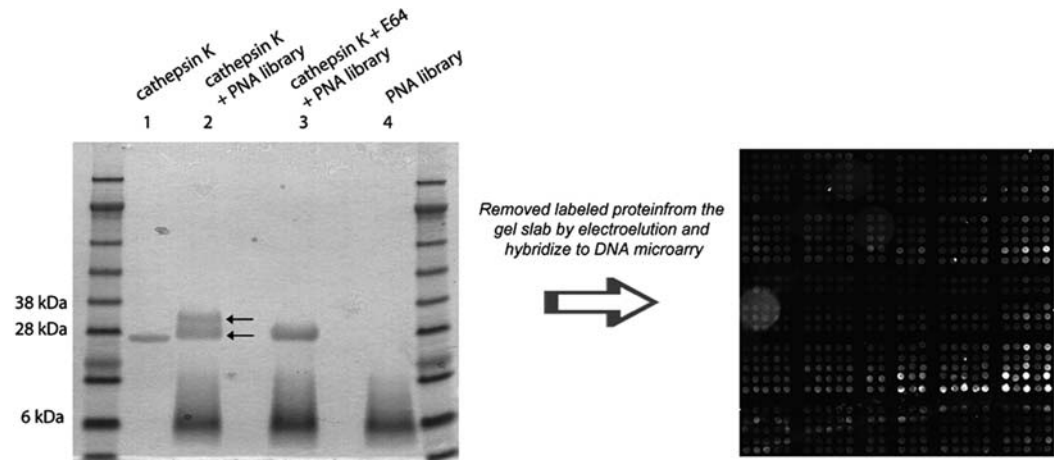


Fig. 15.2. Screening a PNA-encoded library and isolation of inhibitors by gel electrophoresis. A protease (cathepsin K) is incubated with the PNA-encoded library of inhibitors and loaded on a 4–12% SDS-PAGE gel. Lane 1: cathepsin K alone; Lane 2: cathepsin K incubated with an inhibitor library; Lane 3: cathepsin K inactivated with an indiscriminate inhibitor (E64); Electroelution of the protein–inhibitor complex followed by hybridization reveals the structure of the fittest inhibitor.

2. Materials

2.1. Optimized Codon System for Libraries

The codons used to encode the library were optimized such that the solution set of all permutations of final PNA tags fall within a narrow distribution of T_m to insure a homogeneous detection of every library member at a given concentration. The codon system shown below was thoroughly validated for homogeneity, fidelity of hybridization, and sensitivity (4). The codon system shown below can be used to encode a library of 625 compounds with four elements of diversity (R_1–R_4) such as a tetrapeptide.

sequence of the PNA codon for each element of diversity (R) in a PNA-encoded combinatorial library

	R4	R3	R2	R1	
PNA-*N*-terminus	GGAA	CCG	GCA	GAGA	PNA-*c*-terminus
	AAGG	CGA	CGA	GACG	
	CGGC	GCA	ACG	CAGG	
	GCCG	TGG	GTG	CGAA	
	GAAC	GGC	AGC	AGGC	

2.2. Microarray Spotting

1. The oligonucleotides were purchased as desalted without further purification and diluted to 100 μM in 50% DMSO using a Biomek 2000 pipetting robot (using BioArchimed sofware, TDZ ingenierie) in a 1536-well spotting plate (Greiner).
2. Spotting was performed on aminosilane-coated slides subdivided into 16 subarrays (Schott Nexterion mpx 16, slide A) using a Microgrid II arrayer (Genomic Solutions) with Microspot 10 K Microarray Pins (Genomic Solutions).
3. Crosslinking was performed with UV at 600 mJ (Stratalinker, Stratagene).

2.3. Library Synthesis

1. The Rink (0.2–0.7 mmol/g, 1%DVB) peptide synthesis reagents can be purchased from Novabiochem (http://www.emdbiosciences.com/html/NBC/home.html). Solvents and general chemicals can be purchased from Sigma-Aldrich.
2. Fmoc-protected PNA monomers may be purchased from ASM research chemicals, Germany (http://www.asm-research-chemicals.com/index.html) or Panagene, Korea (http://www.panagene.com/).
3. Aminoacids protected in the form of an allyl ester can be obtained from Bachem (http://www.bachem.com/).
4. The library synthesis can be performed in an automated fashion using an Applied Biosystems Expdite 9000 with screw top cartridges from Glen Research (G51226), in a semi-automated fashion using an Argonaut Quest 210 or in a manual fashion using disposable filter cartridges from Supelco. All reactions are performed at room temperature (21–25°C).

2.4. Microarray Hybridization

All solvents and buffers used for microarray experiment were filtered through a 0.2 μm filter.

1. Blocking solution: 50% formamide, 25% 20X SSC pH 7, 0.002% BSA, and 0.002% SDS in deionized water. (20X SSC: 175.3 g NaCl, 88.2 g sodium citrate in 800 mL of H_2O. Adjust the pH to 7.0 with a few drops of HCl. The volume was then adjusted to 1 L with H_2O).
2. Hybridization solution: 40% formamide in deionized water.
3. Washing solutions: *First washing solution*: SSC 2X, SDS 0.02%; *Second washing solution*: SSC 2X, *Third washing solution*: SSC 0.2%.

2.5. Microarray Scanning and Analysis

1. The arrays were scanned with a Genepix microarray scanner at 10 μm resolution using filter corresponding to Cy3 fluorophore (550/570 nm) with a 0.4 s constant exposure time.

2. Array images were analyzed and quantified using Imagene Software.

2.6. Separation of Enzyme-Bound Inhibitor–PNA Adduct by Gel Electrophoresis

1. Centrifugations can be performed on an Eppendorf 5417R at 14,000 rfc (20,000g) without refrigeration or equivalent instrument for 10 min to pellet precipitates.
2. Samples were concentrated using an iCON filter (Pierce) with a molecular weight cutoff of 20 kDa on an Eppendorf 5417R at 14,000 rfc (20,000g) without refrigeration.
3. 12% SDS-PAGE denaturing gel (Invitrogen) was used according to the manufacturer's instruction using MES buffer (MES, 2-(*N*-morpholino)ethanesulfonic acid).
4. The gel was stained with Gel Code Blue Stain (Pierce).
5. Electroelution of the band of interest was performed with D-Tube dialyzer mini dialysis device with a molecular weight of 12–14 kDa (Novagen).

2.7. Separation of Enzyme-Bound Inhibitor–PNA Adduct from the Rest of the Library

1. Centrifugations can be performed on an Eppendorf 5417R at 14,000 rfc (20,000*g*) or an equivalent instrument for 10 min to pellet precipitates.
2. An ultra-free 30 kDa molecular weight cutoff filter (Millipore, MA) was used to separate unbound library members from the protein–inhibitor–PNA adduct.

3. Methods

3.1. Microarray Spotting

1. The oligonucleotides (100 μM in 50% DMSO) were transferred (3 μL) to a 1536-well spotting plate. In order to facilitate subsequent analysis by plain visualization, the oligonucleotides were organized such that each codon represents a column or a row in the array (25 × 25 matrix). Trends regarding the activity of particular residues within a library, thus emerge as horizontal or vertical lines in the array.
2. The oligonucleotides were spotted on aminosilane-coated slides with a Microgrid II arrayer spotting robot at 40–45% relative humidity and 19–21°C.
3. After spotting, aminosilane-coated slides were dried for 48 h in a desiccator before crosslinking with UV at 600 mJ.

3.2. Library Synthesis

1. *General procedure for Alloc or Allyl deprotection*: the resin is treated with a solution of $Pd(PPh_3)_2Cl_2$ (0.2 eq), Ph_3P (0.8 eq), and $TMSN_3$ (10 eq) as a CH_2Cl_2 solution (0.1 M)

followed by a solution of nBu$_3$SnH (5.0 eq) in CH_2Cl_2 (1 M) for 15 min. The reaction is repeated twice and the resin is then washed three times with CH_2Cl_2 andthree times with DMF.

2. *General procedure for amino acid coupling*: the polymer bound acid is pre-activated with DIC (4.5 eq) and HOBt (4.5 eq) in DMF (0.15 M) for 5 min followed by the addition of the amine (5.0 eq) and 2,6-lutidine (10 eq) as a DMF solution (1 M for the amine) and the reaction is continued for 10 h. The resin is then washed three times with DMF.
3. *General procedure for PNA encoding*: the resin is treated with 20% piperidine in DMF for 5 min, washed five times with DMF, and treated with a pre-mixed (5 min) solution of PNA monomer (5.0 eq), EtiPr$_2$N (5.0 eq), HBTU (4.4 eq), and 2,6-lutidine (7.5 eq) in NMP (0.17 M) for 1 h. The resin is filtered without washing and acetic anhydride (5.0 eq) followed by 2,6-lutidine (5.0 eq) in DMF (0.17 M) is added. The resin is agitated for 5 min and then washed three times with DMF. The procedure is repeated three times with the appropriate PNA monomers for the first and last codons and twice for the second and third codons.
4. Rink resin is loaded with Fmoc–Lys(Alloc)–OH at a loading of 0.2 mmol/g (*see* **Note 1**).
5. The Alloc group is removed according to the general deprotection condition and the resulting amine is reacted with diglycolic anhydride (0.1 M), 2,6-lutidine (0.1 M) in DMF (0.17 M). The resin is then washed three times with a wet solution of DMF (0.1 M H_2O) with 2,6-lutidine (0.1 M) followed by three washes with a solution trichloroacetic acid (0.1 M) in CH_2Cl_2 followed by three washes of CH_2Cl_2.
6. The resin is then split into five pools, and the first residue is added according to the general procedure for amino acid coupling followed by the PNA encoding according to the general procedure (*see* **Note 2**).
7. The resin is then removed from all the cartridges, mixed, and suspended in DMF/CH_2Cl_2. After thorough mixing, the resin is redistributed to new cartridges with a pipette (*see* **Note 3**).
8. The allyl is removed according to the general procedure and another amino acid is introduced followed by the PNA encoding according to the general procedures (steps 2 and 3, respectively).
9. Steps 7 and 8 are reiterated twice. The terminal coupling for the inhibitor may be performed with any amine.

10. The final Fmoc is removed from the PNA and Boc–Lys (Fmoc)–OH is coupled to the resin according to the same procedure as for the PNA couplings. Removal of the final Fmoc with 20% piperidine in DMF or DBU (5 eq) in CH_2Cl_2 for 10 min afforded a free amine which can be labeled with Cy3 using Cy-3-NHS (10 eq) in DMF with 2,6-lutidine (10 eq) for 3 h (*see* **Note 4**).
11. The library is released from the resin by cleaving with a 4:1 solution of TFA:*m*-cresol for 1–3 h. The resin is removed by filtration and the TFA solution is added to Et_2O (10–30 times the TFA volume) to obtain a dark pink solution which is pelleted by centrifugation. The pellet is washed three times with Et_2O, redissolved in a acetonitrile:water solution (1:4) and lyophilized.

3.3. Microarray Hybridization and Detection

1. Slides are blocked by incubation with the blocking solution in a 50 mL conical tube for 1 h at 50°C.
2. Slides are then washed four times for 5 min in 200 mL of deionized water and dried with a nitrogen spray or by centrifugation (**Note 5**).
3. The silicon superstructure is then glued to the array and 40 μL of sample in the hybridization buffer (40% formamide with deionized water) are added to each well of the microarray. The slides are hybridized with gentle shaking at 50°C during 4–12 h (**Note 6**).
4. Samples are then removed with a multichannel pipette and the subarrays are washed with 100 μL of the first washing solution before removing the silicon superstructure. The glue from the superstructure must be cautiously removed.
5. Microarrays are then washed by immersion in two succeeding baths of the second washing solution at room temperature. Each slide is then was placed in a 50 mL conical tube where it was washed two times with pure water (**Note 7**).
6. The slides are spun dry by centrifugation at 2000 rpm for 10 s.
7. The arrays are scanned and analyzed.

3.4. Gel Isolation of Enzyme-Bound PNA-Encoded Inhibitors

1. The enzyme of interest (0.1–50 nmol) is incubated with the library at such a concentration that enzyme inhibition is observed. For convenience, the enzyme may be incubated with three-fold dilutions of the library to find the appropriate inhibitory concentration (IC_{50}), the optimal concentration can be assessed by gel-shift of the protein.
2. The samples are then centrifuged at 14,000 rfc (20,000g) for 10 min to remove all precipitates.

3. The soluble fraction is diluted in 5 mL of 1X MES buffer and the sample is concentrated by size exclusion filtration.
4. Concentrated samples are diluted in denaturing buffer and incubated for 5 min at 80°C before being loaded on a SDS-PAGE.
5. The gel is run for 10 min at 100 V and 35 min at 180 V in 1X MES buffer.
6. The gel is then removed from its plastic structure and washed two times for 1 min in 200 mL bath of pure water.
7. Comassie staining of the gel (1 h at room temperature), followed by a 1 h wash with water enable the visualization of the protein–inhibitor–PNA adduct. A shift in molecular weight corresponding to the inhibitor–PNA adduct (3–5 kDa) should be observed relatively to a control band containing the protein alone.
8. The corresponding band is cut out and electroeluted in an electrophoresis chamber at a constant current of 100 V for 1 h in MES 1X buffer.
9. The recovered sample is then diluted in hybridization buffer (60 μL) and serially diluted by taking 20 μL of this solution in 40 μL of hybridization buffer.
10. The samples are hybridized on a microarray for 4–12 h.

3.5. Enzyme Activity Profile from Crude Cell Lysates

1. A mixture of enzymes or crude lysates of interest is incubated with the library.
2. After appropriate incubation, the sample is loaded on an ultra-free 30 kDa molecular weight cutoff filter (Millipore, MA) and washed with 1X PBS buffer (5 μL × 500 μL). The volume of the sample retained in the 30 kDa filter was then adjusted to 60 μL solution of 40% formamide and deionized water and serially diluted by taking 20 μL of this solution in 40 μL of hybridization buffer.
3. The samples are hybridized on a microarray for 4–12 h.

4. Notes

1. Higher loading than 0.2 mmol/g leads to lower efficiency in the PNA couplings. PEG-resin was also successfully used, but without notable improvements. Rink resin purchased with a

loading higher than 0.2 mmol/g was coupled to a substoichiometric amount of the first residue and capped with acetic anhydride.

2. Depending on the nature of the library, one or two Fmoc–Lys(Boc)–OH residue may be introduced before and after the PNA to enhance the water solubility of the final library.
3. The resin sinks in DMF and floats in CH_2Cl_2. The resin is most conveniently suspended in a minimum amount of DMF, and CH_2Cl_2 is added with intermittent agitations until an even suspension is obtained. The use of glass pipette should be avoided to prevent electrostatic adherence of the resin to the glass. The resin is most easily manipulated as wet using disposable Eppendorf tips.
4. Depending on the final use of the library, the resin may be recombined before step 10.
5. Caution must be applied not to touch the side of the slide containing the DNA as physical contact may be sufficient to damage the surface and invalidate further experiments.
6. Control experiments showed very little increase in signal intensity after 4 h of incubation.
7. Prologue washing with deionized water leads to a decrease in signal intensity.

Acknowledgments

The authors thank the granting agencies which have supported this work (Human Frontier Science Program, Agence National de la Recherche, la Région d'Alsace and the French Ministry of Science). The authors thank their collaborators, in particular Jennifer Harris.

References

1. Winssinger, N., Pianowski, Z. & Debaene, F. (2007) Probing Biology with Small Molecule Microarrays (SMM) *Top. Curr. Chem.* **278**, 311–342.
2. Harris, J. L. & Winssinger, N. (2005) PNA encoding (PNA=peptide nucleic acid): From solution-based libraries to organized microarrays *Chem. Eur. J.* **11**, 6792–6801.
3. Harris, J. et al. (2004) Activity Profile of Dust Mite Allergen Extract Using Substrate Libraries and Functional Proteomic Microarrays *Chem. Biol.* **11**, 1361–1372.
4. Urbina, H. D. et al. (2006) Self-assembled small-molecule microarrays for protease screening and profiling *Chem.Bio.Chem.* **7**, 1790–1797.

Chapter 16

A Novel Combinatorial Approach to High-Density Peptide Arrays

Mario Beyer, Ines Block, Kai König, Alexander Nesterov, Simon Fernandez, Thomas Felgenhauer, Christopher Schirwitz, Klaus Leibe, Ralf F. Bischoff, Frank Breitling, and Volker Stadler

Abstract

Combinatorial synthesis of peptides on solid supports *(1)*, either as spots on cellulose membranes *(2)* or with split-pool-libraries on polymer beads *(3)*, substantially forwarded research in the field of peptide–protein interactions. Admittedly, these concepts have specific limitations, on one hand the number of synthesizable peptide sequences per area, on the other hand elaborate decoding/encoding strategies, false-positive results and sequence limitations. We recently established a method to produce high-density peptide arrays on microelectronic chips *(4)*. Solid amino acid microparticles were charged by friction and transferred to defined pixel electrodes onto the chip's surface, where they couple to a functional polymer coating simply upon melting (**Fig. 16.1A–D,F**). By applying standard Fmoc chemistry according to Merrifield, peptide array densities of up to 40,000 spots per square centimetre were achieved (**Fig. 16.1G**). The term "Merrifield synthesis" describes the consecutive linear coupling and deprotecting of L-amino acids modified with base-labile fluorenylmethoxy (Fmoc) groups at the *N*-terminus and different acid-sensitive protecting groups at their side chains. Removing side chain protecting groups takes place only once at the very end of each synthesis and generates the natural peptide sequence thereby.

Key words: Solid-phase peptide synthesis, peptide array, combinatorial peptide libraries, atom transfer radical polymerization (ATRP), amino acid microparticles, peptide chips.

1. Introduction

Peptide libraries are important tools for the investigation and understanding of biological and immunological interactions at the protein level. Due to the high number of individual building blocks, the proteinogenic amino acids, the use of combinatorial approaches is advisable to flexibly generate arrays of truly high diversities.

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_16,

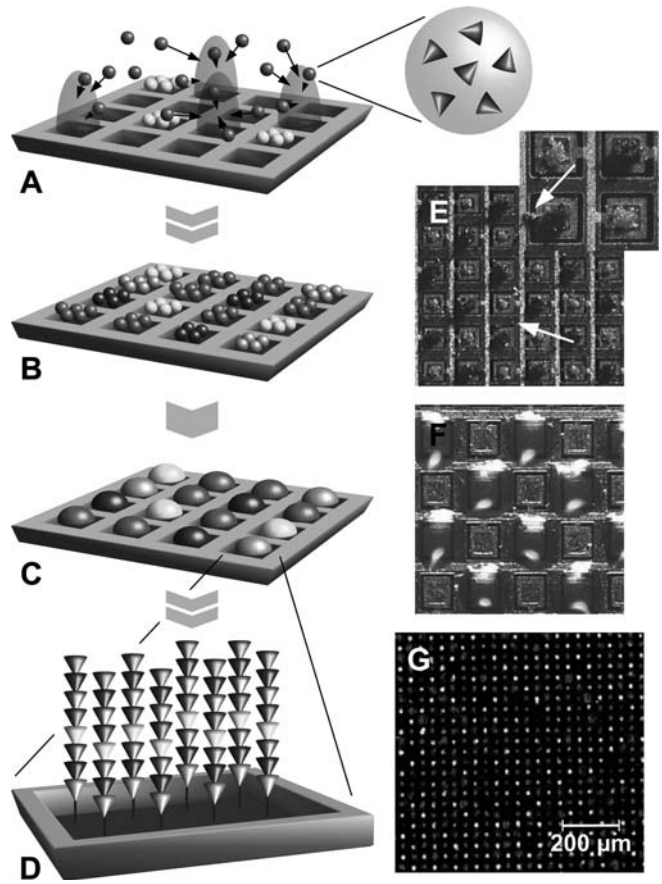


Fig. 16.1. Synthesis of peptide arrays on microchips. Activated amino acids are embedded within particles that are addressed onto a chip's surface by electrical fields generated by individual pixel electrodes (**A**). A whole layer of consecutively addressed amino acid particles (**B**,**E**) is melted at once to induce the coupling reaction (**C**). Repetitive coupling cycles finally generate a peptide array (**D**). Melted particles delimit individual coupling areas. For better visualization, pixel areas are overloaded (**F**). Particle-based in situ synthesis of chessboard arranged Flag® (*white spots*) and HA epitopes (*grey spots*) with a density of 40,000 cm^{-2}; peptides were stained with fluorescently labelled specific antibodies (**G**). Reproduced from (4) with permission from AAAS.

A fundamental change in providing and addressing those building blocks via small, chargeable microparticles with activated amino acids embedded within enables us to increase spot densities and thereby complexities of peptide arrays. By generating localized electric fields on very small pixel electrodes on a microelectronic chip, we can successively deposit these "so-called" amino acid microparticles with micron resolution. Peptide synthesis is induced simply by melting the particles at elevated temperature (T~90°C) for a certain time. Based on higher homologues of DMF, the particle matrix then liquefies and serves as solvent for

the embedded and activated amino acid derivatives during the coupling step. A specific and optimized chip surface modification ensures a functionalized, nanometre thin support coating, similar to conventional PEG-coated or PEG-based polymer beads for solid-phase peptide synthesis (SPPS). Additionally, this coating has protein-repelling properties that suppress interfering background signals due to non-specific protein adsorption. Therefore, we can obviate any additional blocking with BSA or milk powder, when we stain our peptide arrays with primary or secondary antibodies.

2. Materials

2.1. Chips

Microchips were designed in our research group and processed by the Institute for Microelectronics Stuttgart (IMS CHIPS, Stuttgart, Germany). They were glued onto a circuit board and bonded at the Kirchhoff Institute for Physics (University of Heidelberg, Germany).

2.2. Chip Surface Modification

1. Synthesized 2-bromo-2-methyl-*N*-propyltriethoxysilyl isobutyramide (5); methylene chloride p.a., ethanol p.a., methanol p.a., poly(ethylene glycol) methacrylate (PEGMA, $n \sim 5$), copper(I) bromide p.a., 2,2′-bipyridyl (bipy) 98%, *N*,*N*′-diisopropyl carbodiimide (DIC) purum and *N*-methylimidazole (NMI) p.a. were purchased from Sigma-Aldrich (Taufkirchen, Germany), Fmoc-ß-alanine >99% from Iris Biotech (Marktredwitz, Germany).
2. High-pressure mercury lamp and power supply (UV-Consulting Peschl, Mainz, Germany); 100 mL Schlenk flask with stirring bar, ultrasonic bath, vacuum desiccator, vacuum pump ($>10^{-2}$ mbar) and inert gas for silanization and polymerization under oxygen-free and dry gas phase conditions.

2.3. Amino Acid Microparticles

1. Side chain protected Fmoc amino acid pentafluorophenyl esters (OPfp) as combinatorial building blocks from Merck Biosciences GmbH (Bad Soden, Germany).
2. Formulating, milling, sieving and characterizing of amino acid microparticles were done according to literature (4).

2.4. Peptide Synthesis

1. Solvents and liquids: *N*,*N*-dimethylformamide (DMF) peptide grade (Biosolve BV, Valkenswaard, The Netherlands) was dried over molecular sieve; methanol p.a., ethanol p.a., chloroform p.a., methylene chloride p.a., piperidine 99%, acetic anhydride p.a., *N*,*N*-diisopropylethylamine (DIPEA)

p.a., triisobutylsilane (TIBS) puriss. were purchased from Sigma Aldrich (Taufkirchen, Germany) and were used without further purification. Trifluoroacetic acid (TFA) 99% (Acros Organics, Geel, Belgium).

2. Plexiglas aerosol generators for charging and deposition of amino acid microparticles onto the chip's surface; washing chamber made from PTFE (Teflon®) (both from Kirchhoff Institute for Physics, University of Heidelberg, Germany) with stirring bar; temperature-controlled, gas-proofed reaction container.

2.5. Immunostaining

1. Tris-buffered saline: prepare a 10X stock with 0.5 *M* (60.57 g) Trisma® base and 1.5 *M* (87.66 g) NaCl (adjust to pH 7.4 with concentrated HCl). Prepare working solution by dilution of one part with nine parts of water. Add 0.1 v% of Tween 20 to obtain TBS-T buffer for washing and antibody dilution (*see* **Note 1**).
2. Primary antibodies: mouse monoclonal anti-FLAG® M5 and rabbit anti-HA antibody (both from Sigma Aldrich, Taufkirchen, Germany).
3. Secondary antibodies: Alexa Fluor® 647 goat anti-rabbit IgG (H+L) and Alexa Fluor® 546 goat anti-mouse IgG (H+L) (both from Invitrogen GmbH, Karlsruhe, Germany).
4. GenePix 4000B fluorescence scanner and GenePix Pro 4.0 Microarray Image Analysis software (both from Molecular Devices, Sunnyvale, CA, USA).

3. Methods

3.1. Graft Polymerization of Poly(ethylene glycol) methacrylate (PEGMA) on the Chip Surface

1. Before silanization, chips were cleaned and pre-activated by irradiation with a 150 W mercury vapour lamp positioned at a distance of 4 cm for 1 h in air. Afterwards, they were immediately incubated in a 10 mM solution of 2-bromo-2-methyl-*N*-propyltrimethoxysilyl isobutyramide in anhydrous methylene chloride over night. The reaction was stopped by addition of ethanol (p.a.), followed by extensive rinsing with ethanol and distilled water. After that chips were carefully dried in a stream of dried air. To achieve full condensation of silanol groups, silanized chips were heated for 1 h at 100°C in an oven. Storage was done under nitrogen atmosphere at –20°C. Chips were solely handled with PP tweezers to avoid scratching of the surfaces (5).

2. Graft polymerization of PEGMA was typically carried out with 5 mL monomer (15.3 mmol) and 10 mL of a water/methanol mixture (1:1) in a 100 mL Schlenk flask. Then, 64 mg of copper (I) bromide (CuBr, 0.45 mmol) and 141 mg of 2,2′-bipyridyl (bipy, 0.90 mmol) were added as ATRP catalyst under nitrogen atmosphere.
3. The brown reaction mixture was immediately degassed and sonicated under nitrogen atmosphere to dissolve the CuBr before the silanized chips were immersed. Polymerization took place under nitrogen atmosphere in a desiccator for at least 24 h.
4. Afterwards, the chips were extensively rinsed with distilled water for 15 min and washed in dimethylformamide (DMF) for 15 min.
5. The coated microchips are rinsed with ethanol and distilled water. Finally, they are dried in a stream of nitrogen and stored under nitrogen atmosphere at –20°C.

3.2. Surface Modification for SPPS

Terminal amino groups for anchoring the first amino acid were introduced by coupling Fmoc-β-alanine to the PEGMA side chains: carboxyl groups were activated in situ with *N*,*N*′-diisopropyl carbodiimide (DIC) and *N*-methylimidazole (NMI), followed by Fmoc deprotection with piperidine/DMF. Due to a better accessibility of growing peptide chains, three β-alanine residues were consecutively coupled to generate a NH_2–(β-alanine)$_3$ sequence.

1. Pre-swell the coated microchip surface with dry DMF for 30 min.
2. Mix 62.26 mg (0.2 *M*) of Fmoc-β-alanine and 37.2 μL (0.24 *M*) of DIC in 1 mL of dry DMF for at least 5 min.
3. Add 31.8 μL (0.4 *M*) of NMI to the DMF solution; shake your vessel thoroughly. Remove remaining DMF for swelling from the chip and add the mixture to it.
4. Incubate the chip with gentle agitation over night under nitrogen for at least 16 h.
5. Wash three times for 1 min each with DMF and treat the chip with capping solution (acetic anhydride:DIPEA:DMF, 5:10:35; *see* **Note 2**) over night to block residual reactive hydroxyl groups.
6. Remove the mixture (dark red coloured) and wash five times for 2 min each with DMF.

7. Remove the DMF and add 1 mL of 20% piperidine in DMF to remove N-terminal Fmoc groups for 20 min. You can determine the amount D of coupled β-alanine in (nmol/cm^2) by measuring the extinction E of the piperidine/DMF solution at 301 nm by using equation [1]:

$$D[\text{nmol/cm}^2] = \frac{E \cdot V[\text{L}] \cdot 10^9}{\varepsilon \cdot A[\text{cm}^2]} \quad [1]$$

where V is the volume (L) of piperidine solution; ε (=5129) is the molar extinction coefficient (6) and A is the area of derivatized chip surface (cm^2).

8. Wash three times for 2 min each with DMF. Repeat steps 1–8 until three ß-alanine units are introduced.
9. Wash two times for 2 min each with methanol and dry the chip in a stream of nitrogen. Store the functionalized microchips under nitrogen in dark at 0°C or proceed with step 3.4.

3.3. Amino Acid Microparticles

Fabrication of amino acid microparticles was done according to the following procedure: in order to produce a toner mass, one of 20 different Fmoc-amino-acid-OPfp-esters (Merck Biosciences GmbH, Bad Soden, Germany, 10% w/w), a toner resin (SEKISUI CHEMICAL, Tokyo, Japan; 84.5% w/w), pyrazolone orange (ABCR GmbH, Karlsruhe, Germany; 4.5% w/w) and an iron (III)-bis (naphthoic acid) complex (self-synthesized; 1% w/w) were melted, mixed and finally solidified. The homogenized and crushed toner mass was then mixed with ~4% w/w silica particles, and slowly fed into an air jet mill where grinded particles were collected within a narrow lattice (4).

3.4. Solid-Phase Peptide Synthesis on Chip's Pixel Electrodes

1. Each sort of amino acid microparticles is stored in a particular aerosol generator. For selective particle deposition, the particles were dispersed by a stream of dry air and charged by friction due to collision with Plexiglas generator walls. Charged amino acid microparticles were addressed onto the microchip's surface by directing electric forces. After each deposition of one particle sort, the chip is shortly heated up (<3 min) to fix addressed particles onto the surface.
2. After all different amino acid particles are directed with micron resolution, the chip is transferred into a pre-warmed, nitrogen-flushed reaction chamber. Here, the coupling reaction is initiated upon melting at 90°C for 90 min under inert gas.
3. After cooling to room temperature, the chip is mounted into the Teflon washing chamber. All washing steps are recommended to do under a fuming hood. Wear gloves and protective safety glasses!

4. Wash 1 min and then 20 min with 1 mL of a mixture of acetic anhydride, DIPEA and DMF (5:10:35) to remove residual particle material and to block residual amino groups at the same time (*see* **Note 2**).
5. Wash five times for 5 min each with DMF and remove the *N*-terminal Fmoc protecting group by incubating the microchip with 1 mL of 20 v% piperidine in DMF. Remove the piperidine–DMF solution.
6. Wash five times for 5 min each with DMF and two times for 5 min each with methanol (MeOH).
7. Dry the chip in a stream of nitrogen, remove it from the washing chamber and repeat steps 1–7 until your peptide sequence is completed.
8. To finally deprotect the amino acid side chains, keep the chip in the washing chamber and wash two times for 5 min each with methylene chloride. Mix 51% trifluoroacetic acid (TFA; very hygroscopic and corrosive!), 44% methylene chloride, 3% triisobutylsilane (TIBS) and 2% water (*see* **Note 3**).
9. Remove the methylene chloride and treat the chip six times for 15 min each with the TFA solution under stirring. Keep the chamber closed carefully. Remove the TFA solution and dispose it together with other halogenated waste materials. Avoid mixing TFA and DMF.
10. Wash the chip four times for 2 min each with methylene chloride and DMF, followed by two times for 2 min each with ethanol. Rinse the chip with water and dry it in a stream of nitrogen. Proceed to 3.5 or store the chip under nitrogen at –20°C.

3.5. Immunostaining with Fluorescently Labelled Antibodies

1. Rehydrate the chip surface by incubating in 1 mL of TBS-T buffer for at least 30 min. Because of the protein-repelling properties of the PEGMA coating on the chip, no additional surface blocking is necessary.
2. Dilute specific antibodies (e.g. mouse anti-Flag® and rabbit anti-HA) to 1:1000 each in TBS-T. Incubate the chip with 1 mL antibody solution for 1 h. Remove the buffer solution.
3. Wash six times for 5 min each in TBS-T with gentle shaking. Avoid dehumidification of the chip's surface (*see* **Note 4**).
4. Dilute secondary antibodies [e.g. Alexa Fluor® 647 goat anti-rabbit IgG (H+L) and Alexa Fluor® 546 goat anti-mouse IgG (H+L)] to an appropriate ratio and incubate the chip for 1 h.
5. Wash six times for 5 min each in TBS-T under gentle shaking.

6. Dry the chip in a stream of nitrogen and remove it carefully from the circuit board.

The chip can then be analysed with a fluorescence scanner (e.g. GenePix 4000B, Axon Instruments) at corresponding wavelengths (532 nm for Alexa Fluor® 546 and 635 nm for Alexa Fluor® 647 dyes, respectively).

4. Notes

1. Unless stated otherwise, all aqueous solutions like buffers, e.g. should be prepared with Millipore® water (resistivity of about 18.2 MΩ.cm).
2. At all times, the capping solution mixture should be prepared freshly prior to incubating the chip.
3. Because of the high volatility of methylene chloride, it is recommended to add it to the TFA:TIBS:water mixture at last to avoid its evaporating.
4. Keeping the chip wet is crucial to avoid Langmuir–Blodgett-like protein or antibody transfer onto the surface upon dehydration. That would result in non-specific adsorption of proteins and thus reduced signal-to-noise ratios in immunostainings.

References

1. For a short historical summary: Kimmerlin, T., Seebach, D. (2005). "100 years of peptide synthesis": ligation methods for peptide and protein synthesis with applications to ß-peptide assemblies. *J. Pept. Res.* **65**(2), 229–260.
2. Frank, R. (1992). Spot-synthesis: an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. *Tetrahedron* **48**, 9217–9232.
3. Review: Lam, K. S., Lebl, M., Krchnak, V. (1997). The "One-Bead-One-Compound" Combinatorial Library Method. *Chem. Rev.* **97**, 411–448.
4. Beyer, M.,* Nesterov, A.,* Block, I., König, K., Felgenhauer, T., Fernandez, S., Leibe, K., Torralba, G., Hausmann, M., Trunk, U., Lindenstruth, V., Bischoff, F. R., Stadler, V., Breitling, F. (2007). Combinatorial Synthesis of Peptide Arrays onto a Microchip. *Science* **318**, 1888. *, : These authors contributed equally to this work.
5. Stadler, V., Beyer, M., König, K., Nesterov, A., Torralba, G., Lindenstruth, V., Hausmann, M., Bischoff, F. R., Breitling, F. (2007). Multifunctional CMOS Microchip Coatings for Protein and Peptide Arrays. *J. Proteome Res.* **6**, 3197–3202.
6. Beyer, M., Felgenhauer, T., Bischoff, R. F., Breitling, F., Stadler, V. (2006). A novel glass slide-based peptide array support with high functionality resisting non-specific protein adsorption. *Biomaterials* **27**(18), 3505–3514.

Chapter 17

Polypyrrole–Peptide Microarray for Biomolecular Interaction Analysis by SPR Imaging

Marie-Bernadette Villiers, Sandra Cortès, Carine Brakha, Patrice Marche, André Roget, and Thierry Livache

Abstract

Nowadays, high-throughput analysis of biological events is a great challenge which could take benefit of the recent development of microarray devices. The great potential of such technology is related to the availability of a chip bearing a large set of probes, stable and easy to obtain, and suitable for ligand-binding detection. Here, we describe a new method based on polypyrrole chemistry, allowing the covalent immobilization of peptides in a microarray format and on a gold surface compatible with the use of surface plasmon resonance. This technique is then illustrated by the detection and characterization of antibodies induced by hepatitis C virus and present in patients' serums.

Key words: Peptide chips, pyrrole, SPR imaging, antibody, hepatitis C virus.

1. Introduction

Among the miniaturized devices used for biological analysis, chips bearing an array of peptides are of special interest. Indeed, peptide arrays open many applications such as enzyme–substrate interaction measurement (1) and epitope mapping (2, 3) (for reviews, *see* (4, 5)). Among them, the screening of antibodies in biological samples is an important topic related to the detection of infectious diseases and the monitoring of the evolution of the disease (6–8). Beyond this application side, peptide arrays are very attractive from a chemical point of view because peptides are much shorter and more stable than proteins, suitable for the production of stable microarrays. Nevertheless, the stability of the device is dependent on a strong linkage between the peptides and the surface of the microarray, allowing the chip to be easily regenerated to perform high-throughput assays. Screening analyses require a detection

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_17,

method compatible with microarray format. Surface plasmon resonance imaging (SPRi) is a good alternative as it is an accurate one-step method for real-time measurement of ligand binding without labelling or previous purification (9). This optical detection allows the monitoring of biological recognitions occurring on the surface of a biochip bearing spots containing different biomolecules. This is performed by following the change in the reflected light intensity induced by ligand binding (**Fig. 17.1**). Moreover, the SPRi signal is correlated with the amount of ligand present in the injected solution (10).

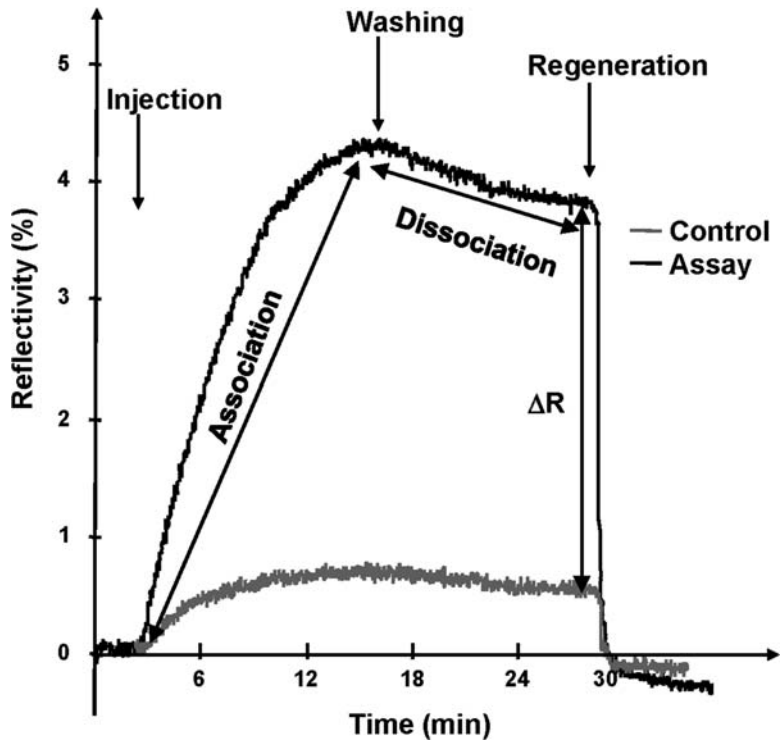


Fig. 17.1. Detection of ligand binding by SPRi: sensorgram obtained on two spots (control and assay) upon ligand injection.

We describe in this chapter an approach combining the use of peptide chips and the SPRi. In order to be compatible with the gold layer used because of its optical properties, we have developed an electrochemical-based process allowing the fast grafting of biomolecules into a matrix of polypyrrole: the biomolecule is first modified with a pyrrole moiety and then copolymerized with free pyrrole (11). In the case of peptide chip, the pyrrole moiety is added on the N-extremity of each peptide to be arrayed before its copolymerization. The peptide chip is constructed on a glass prism covered with a thin gold

layer and an application example is given, dealing with the screening of serums from patients infected with hepatitis C virus (HCV) for their content in anti-HCV antibodies.

2. Materials

2.1. Synthesis of N-(2-Mercaptoethyl)-6-(1H-pyrrole-1-yl) hexanamide (=Pyrrole-SH)

1. 2,5-Dimethoxytetrahydrofuran, mixture of *cis* and *trans* isomers 98% (Sigma-Aldrich).
2. 6-Aminocaproic acid 99+ % (Acros Organics).
3. Cysteamine 98% (Sigma-Aldrich).
4. 1,4-Dioxane anhydrous 99.8% (Sigma-Aldrich).
5. Acetic acid puriss (Fluka).
6. *N*-Hydroxysuccinimide 98% (Sigma-Aldrich).
7. *N*,*N*'-Dicyclohexylcarbodiimide (Fluka).
8. *N*,*N*-Dimethylformamide (DMF) for analysis (Carlo Erba).
9. Dichloromethane (CH_2Cl_2) puran (SdS).
10. Ethyl alcohol anhydrous (EtOH)), for analysis (Carlo Erba).
11. Chloroform-d, 99.9% D, stabilized with 0.5 wt silver foil (Sigma-Aldrich).
12. Silicagel PF_{254} containing $CaSO_4$for preparative layer chromatography.

2.2. Preparation of Pyrrolylated Peptides

1. Dimethylsulphoxide (DMSO) analytical reagent (Prolabo).
2. Peptides modified with a maleimide group at their NH_2 terminus (Neosystem, Strasbourg, France).
3. Reacting buffer: 50 mM NaH_2PO_4, 50 mM NaCl, 10% (p/v) glycerol, final pH adjusted at 6.8 with a NaOH solution. Store at 4°C (−20°C for a conservation longer than 2 weeks).

2.3. Immobilization of the Peptides on the Chip

1. 1 M Pyrrole in acetonitrile (Tokyo Kasei, Japan). Store at −20°C in a dark bottle to minimize pyrrole oxidation.
2. 2 mM Pyrrole–peptide conjugate (prepared as described in **Section 3.2**).
3. Reacting buffer: 50 mM NaH_2PO_4, 50 mM NaCl, 10% (p/v) glycerol, final pH adjusted at 6.8 with a NaOH solution. Store at 4°C (–20°C for a conservation longer than 2 weeks).

4. The chips consist in glass prisms coated with a 50 nm-thick layer of gold (Genoptics, Orsay, France).
5. The microarrayer required for immobilizing the peptides on the chip is available commercially (Genoptics, Orsay, France). It mainly consists in an inox needle (260 nm internal diameter) which is filled with the solution (6 nL) containing the peptide to be grafted and which can move to a precise location on the chip. An electrical pulse between the needle (counter electrode) and the gold surface of the chip (working electrode) induces the synthesis of the polypyrrole film and its deposition on the gold surface.

2.4. SPRi Interaction Monitoring

1. Injection and washing buffer (PBS): 8 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 2.7 mM KCl, 138 mM NaCl, pH 7.2. Store at 4°C.
2. Saturation buffer: bovine serum albumin (BSA, Sigma-Aldrich) 1% (w/v) in PBS.
3. Regeneration buffers:
   (a) 0.1 M Glycine, final pH adjusted at 2.3 with a HCl solution. Store at 4°C.
   (b) 1% (w/v) Sodium dodecyl sulphate (SDS) in water. Store at room temperature.
4. The SPRi apparatus is a SPRiPlex (Genoptics, Orsay, France). The prism is inserted in the apparatus and a flow cell (6 μL) inside the instrument is hermetically sealed over the prism. A light source (660 nm) illuminates the prism and a CCD camera monitors the ligand–probe interactions. Data are analysed with dedicated software (SPRi-view).

3. Methods

3.1. Synthesis of N-(2-Mercaptoethyl)-6-(1H-pyrrole-1-yl) hexanamide (=Pyrrole-SH)

1. A mixture of 2, 5-dimethoxytetrahydrofuran (0.49 mol, 65 mL), 6-aminocaproic acid (0.43 mol, 56 g), acetic acid (450 mL), and dioxane (600 mL) is heated under reflux for 4 h and stirred at room temperature overnight (12).
2. Remove the volatiles under reduced pressure. The residue is co-evaporated in ethanol (2 mL × 100 mL) to eliminate acetic acid.
3. Dissolve the residue in 500 mL of dichloromethane. This organic solution is washed with 2 mL × 250 mL of water and evaporated to dryness.

Product 1 (**Fig. 17.2**) is obtained with a yield of 82% after purification by flash chromatography on a silicagel (600 g) column with a discontinuous gradient of ethanol in CH_2Cl_2. After deposition in the minimum amount of CH_2Cl_2, the elution starts with CH_2Cl_2 (500 mL), then CH_2Cl_2/EtOH: 98/2 (300 mL), CH_2Cl_2/EtOH: 95/5 (400 mL), and finally CH_2Cl_2/EtOH: 90/10 to elute the product. Product 1 is obtained after evaporation of the solvents as a brownish oil.

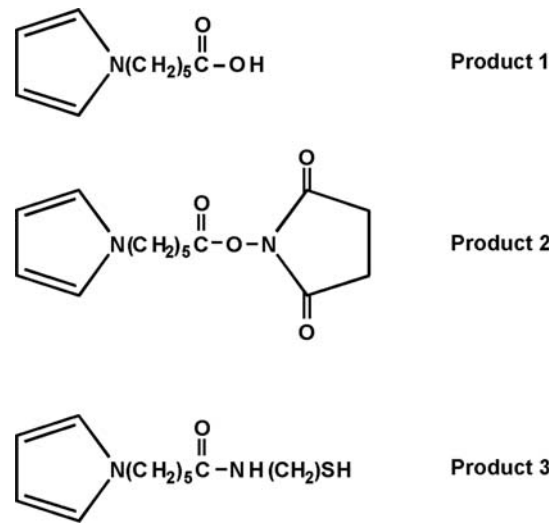


Fig. 17.2. Structures of the intermediates (products 1 and 2) and final product (3) for *N*-(2-mercaptoethyl)-6-(1H-pyrrole-1-yl) hexanamide synthesis (= pyrrole–SH).

M. S. product 1 (m/z) = 182.1 (M+).

1H-NMR product 1 (200 MHz; $CDCl_3$) δ(ppm): 1.72 (m, 6H, -CH_2-(CH_2)$_3$-CH_2-); 2.34 (t, 2H, -CH_2-COOH); 3.87 (t, 2H, CH_2-N); 6.13 (dd, 2H, 3-**H** and 4-**H** pyrrole); 6.64 (dd, 2H, 2-**H** and 5-**H** pyrrole).

4. Product 1 (144 mmol, 26.05 g), *N*-hydroxysuccinimide (144 mmol, 16.56 g), and *N*, *N* ′-dicyclohexylcarbodiimide (159 mmol, 32.75 g) are dissolved in DMF (1 l) and stirred overnight at room temperature.
5. Eliminate the white precipitate of *N*,*N*′-dicyclohexylurea by filtration on a sinter glass.
6. Remove the solvent by evaporation on a rotavapour. Product 2 can be used without further purification. Alternatively it can be purified on a silicagel column in the same way than product 1 (elution CH_2Cl_2/EtOH 95/5).

   M. S. product 2 (m/z) = 279.3 (**Fig. 17.2**)

1H-NMR product 2 (200 MHz; $CDCl_3$) δ(ppm):1.72 (m, 6H, -CH_2-(C**H**$_2$)3-CH_2-); 2.33 (t, 2H, -C**H**$_2$-CO); 2.88 (tt, 4H, C**H**$_2$ of NHS); 3.87 (t, 2H, C**H**$_2$-N); 6.13 (dd, 2H, 3-**H** and 4-**H** pyrrole); 6.64 (dd, 2H, 2-**H** and 5-**H** pyrrole).

7. Product 2 (5 mmol, 1.89 g) and cysteamine (6 mmol, 463 mg) are dissolved in DMF and stirred overnight at room temperature.
8. Evaporate the solvent and dissolve the residue in 50 ml of dichloromethane. This organic solution is washed by 2 mL × 50 mL of water and evaporated to dryness. The oily residue is taken off in 5 mL of dichloromethane and filtered on a small column of silicagel. After elution and evaporation of the solvent the product is obtained as pale yellowish oil with a yield of 76%. M. S. product 3 (m/z) = 266.1 (M+) (**Fig. 17.2**).

   1H-NMR product 3 (200 MHz; $CDCl_3$) δ(ppm): 1.72 (m, 6H, -CH_2-(C**H**$_2$)$_3$-CH_2-); 2.17 (t, 2H, -C**H**$_2$-CONH spacer); 2.68 (t,2H, C**H**$_2$-SH cysteamine); 3.53 (t,2H, C**H**$_2$-N cysteamine); 3.87 (t, 2H, C**H**$_2$-N spacer); 6.13 (dd, 2H, 3-**H** and 4-**H** pyrrole); 6.64 (dd, 2H, 2-**H** and 5-**H** pyrrole).

3.2. Preparation of Pyrrolylated Peptides

1. Each peptide-maleimide is solubilized in DMSO at 5 mM concentration **(Fig. 17.3A)** (*see* **Note 1**).
2. Prepare the reacting solution by mixing 40 µL of 5 mM peptide-maleimide with 10 µL of 0.2 M pyrrole-SH and complete to 100 µL final volume with reacting buffer. These conditions lead to a 2 mM final concentration of peptide-maleimide with a molar ratio peptide/pyrrole-SH = 10 (*see* **Note 2**).

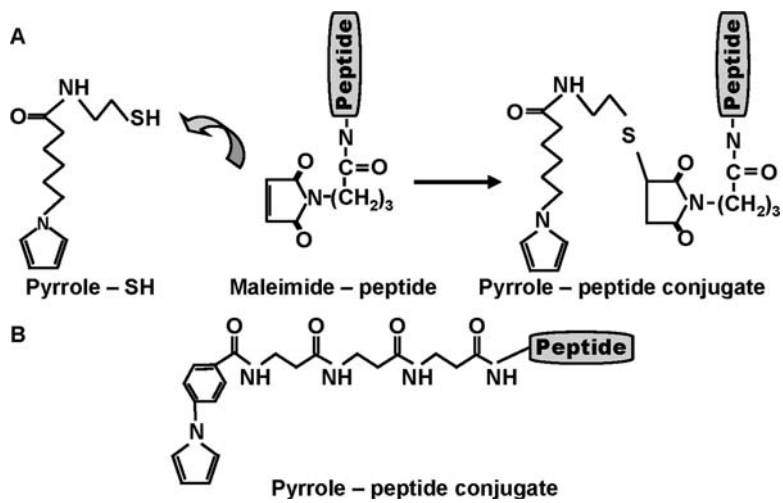


Fig. 17.3. Pyrrole–peptide conjugates. (**A**) Schematic representation of the coupling reaction between maleimide functional group at the N-Ter of the peptide and thiol group in the pyrrole to form a thioether linkage. (**B**) Structure of pyrrole–peptide conjugates provided by Altergen.

3. Incubate 2 h at room temperature under stirring and in the dark to minimize pyrrole oxidation. Store at –20°C.

3.3. Immobilization of the Peptides on the Chip ("Electrospotting")

All solutions are prepared extemporaneously.

1. Prepare a 40 mM pyrrole solution by diluting 20 µL of the 1 M stock solution with 480 µL of reacting buffer.
2. Prepare a 200 µM pyrrolylated peptide solution by diluting 5 µL of the 2 mM solution with 45 µL of reacting buffer.
3. Prepare the spotting solution in a 96-U-well plate by mixing 15 µL of the 200 µM pyrrolylated peptide solution with 15 µL of the 40 mM pyrrole solution. Final concentration for pyrrole–peptide and free pyrrole are 100 µM and 20 mM, respectively, (*see* **Note 3**), in a 30 µL final volume (*see* **Note 4**). Be careful to ensure a good homogenization of the spotting solution without introducing air bubbles (*see* **Note 5**).
4. Immediately put the plate containing the spotting solutions in the microarrayer and carry out the spotting on the chip. The spotting conditions consist of a 2 V electrical pulse for 100 ms (*see* **Note 6**) on defined areas of the gold prism (**Fig. 17.4**).

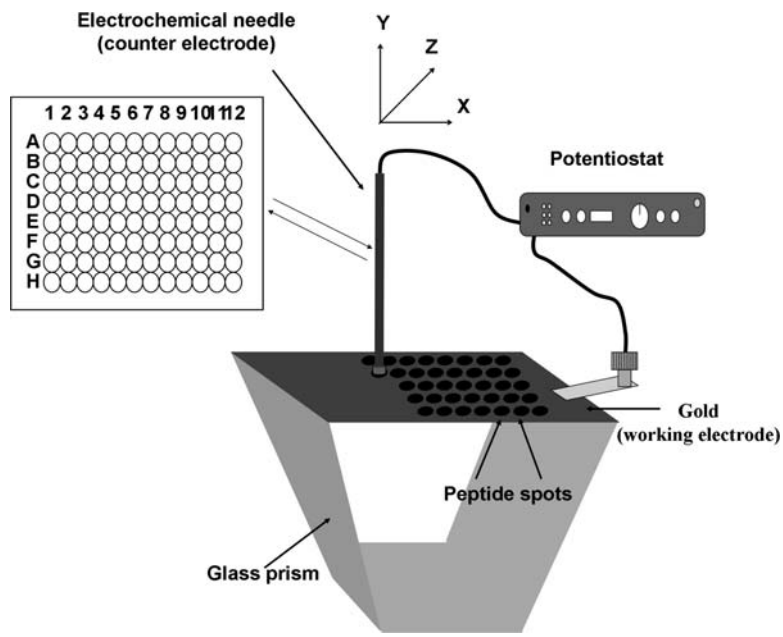


Fig. 17.4. General scheme of the peptide addressing on a glass prism coated with gold. Different pyrrole–peptide conjugates and pyrrole monomer solutions are in the 96-well plate. The "electrospotting" is carried out on the gold surface via the needle containing the solution to be copolymerized.

5. When all peptide spots have been realized (*see* **Notes 7** and **8**), the chip is removed from the microarrayer, rinsed with water and gently dried with an argon or air jet. Chip is kept dry at room temperature.

3.4. SPRi Interaction Monitoring

1. Insert the chip into the SPRiPlex and start the apparatus. The conditions are as follows: flow rate = 37.5 µL/min; room temperature, injection loop = 500 µL; working buffer = PBS. Be careful to avoid air bubbles which can be eliminated by flushing PBS through the circuit (*see* **Note 9**).
2. Saturate the chip by injecting PBS–BSA saturation buffer during 10 min (*see* **Note 10**).
3. Rinse the chip with PBS during 10 min.
4. Inject regeneration buffer during 5 min to stabilize and condition the chip according to the following tests (*see* **Note 11**).
5. Rinse the chip with PBS during 10 min. The chip is ready for sample injections. Each analysis includes the following steps:
   - injection of the sample diluted in PBS (*see* **Note 12**) during 10 min
   - washing with PBS during 10 min
   - regeneration with 0.1 M glycine pH 2.3 during 5 min
   - washing with PBS during 10 min.
6. After a large number of analyses, depending on the complexity of the samples (*see* **Note 13**), it is possible that the chip becomes less responsive. In this case, it could be useful to inject 1% SDS in water during 5 min, followed by PBS during 15 min.

3.5. Application to Antibody Detection

Detection and analysis (quantitative and qualitative) of antibodies in serum are crucial not only to diagnose a disease, but also to evaluate its developmental stage and its evolution. SPRi technology allows high-throughput analyses in real time using a label-free detection process suitable for screening a large number of analytes. This technology requires the construction of biochips. The approach described here is based on electropolymerization of pyrrole and presents several advantages: easy and rapid preparation, versatility, high reproducibility, and control of the density of immobilized probes. Moreover, the immobilization process used to prepare the peptide chips leads to a low background, making them suitable for the analysis of complex biological samples such as serums.

Recently, this technology has been used to analyse the humoral response induced by hepatitis C virus (HCV). A set of 20 aa-peptides (62) has been grafted in triplicates on two chips as described above, and 45 serums from patients with HCV infection were injected (1/50 dilution, in PBS). Representative sensorgrams are shown in **Fig. 17.5**. The histograms in **Fig. 17.6** show the corresponding final results, and represent the specific SPRi signal

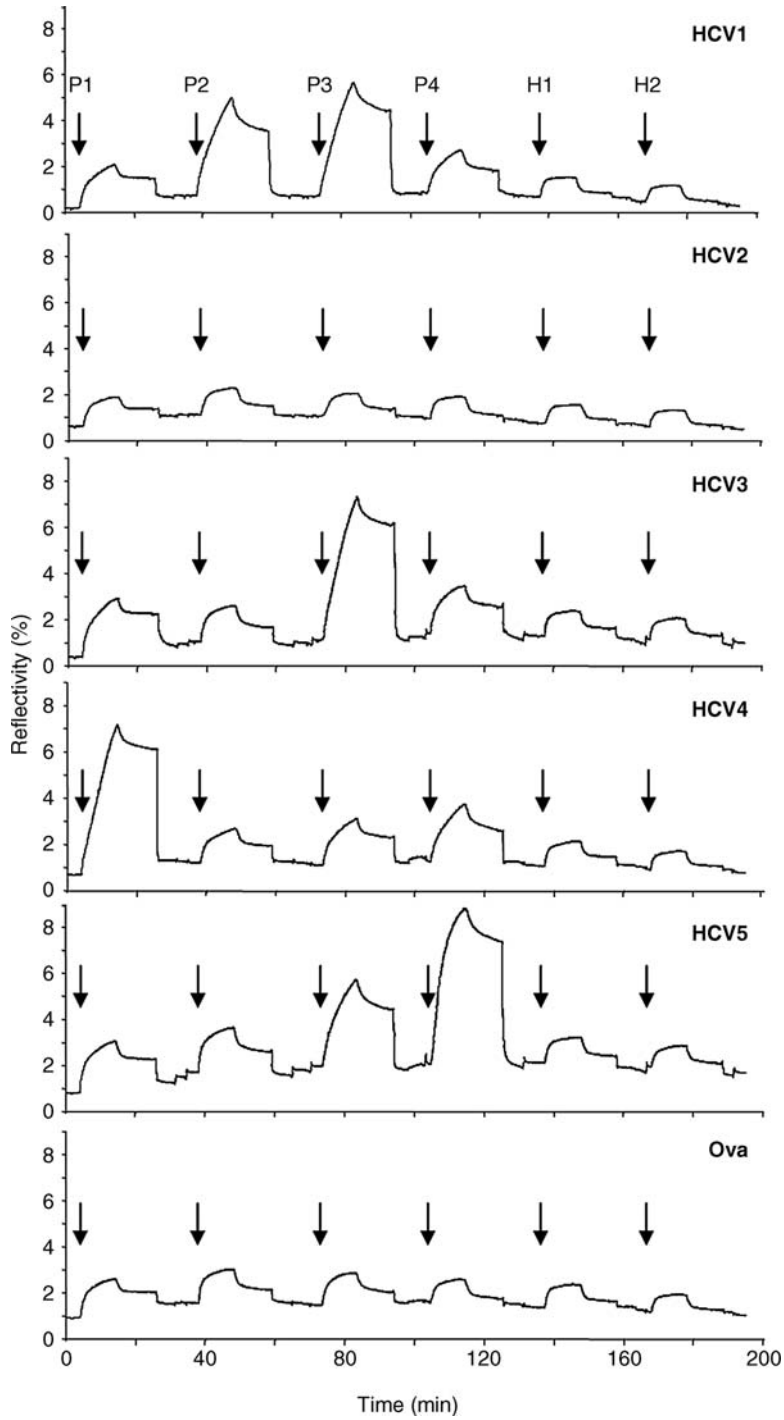


Fig. 17.5. Sensorgrams obtained after injection of serums from four patients (P1–P4) and two healthy donors (H1 and H2). The serums (1/50 dilution in PBS) were injected on a chip bearing 5 HCV peptides and Ova 273–288 as a control, with a regeneration step between each injections.

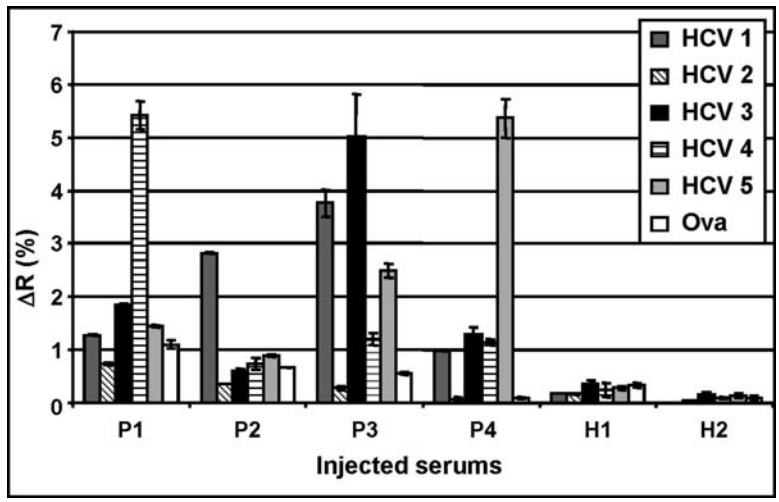


Fig. 17.6. SPRi signal obtained for each serum in the conditions described in Fig. 17.5. ΔR is determined as in Fig. 17.1).

obtained for each tested serum. Thus, this method enabled us to sketch an antibody profile for each of the patients, which is of great importance in the clinical field.

Other applications can take benefit from this technology, from epitope mapping to analysis of enzyme–substrate interaction, making it a powerful tool for both basic research and clinical applications such as diagnosis and pharmacological studies.

4. Notes

1. A more convenient but more expensive way to obtain peptide–pyrrole conjugates consists in getting it already synthesized by the manufacturer (Altergen, Bischheim, France) (**Fig. 17.3B**). The great advantage is that 100% of the peptide put in the spotting solution is pyrrolylated instead of 20–40% for homemade conjugates, depending on the yield of the reaction between peptide-maleimide and pyrrole-SH.
2. An initial molar ratio (IMR) peptide/pyrrole-SH of 10 leads to a good coupling reaction between the two molecules. Nevertheless, it is recommended to test several IMR (5, 10, 20, 40) to optimize the coupling conditions.
3. Different final peptide concentrations (50, 100, 200, 500 μM) have to be prepared to determine the optimum leading to the best reactivity with the tested ligands. This point is

important as the coupling yield (usually between 2 and 40%) could differ from one peptide to another. In all cases, a 20 mM final concentration of free pyrrole is required.

4. The final volume is not crucial as the volume required for the electrodeposition is in the nL range. Nevertheless, as the samples could remain for a long time in the plate before being electrodeposited, depending on the number of the spots to be realized, a significant evaporation must be feared if sample volume is too small.
5. The presence of air bubbles in the sampling needle introduces a defect in the electropolymerization (interruption of the electrical current).
6. These conditions lead to a 2–5 nm thick polypyrrole film. Different pulse durations can be tested (50, 100, 200 ms) to optimize the electrodeposition. It is worth to notice that a too long pulse and/or a too important voltage could lead to a thicker polypyrrole film. Moreover, it could induce a depletion of pyrrole molecules as well as hyperoxidation of the pyrrole and consequently, degradation of the film.
7. It is recommended to realize the spots in triplicates to improve the quality of the results. In this case, it is better to set the three identical spots perpendicular to the flow direction in the SPRiPlex, to avoid a potential ligand depletion between the first and the third spot, as it could occur if the three spots are parallel to the flux.
8. To quantify the background level, control spots including irrelevant peptides and/or polypyrrole alone have to be performed.
9. It could be useful to degas the PBS solution either before its use or by including a degasser in the fluid circuit.
10. The choice of the regeneration buffer depends on the interaction type between probes and ligands. Glycine is convenient for pH-dependent interactions such as antigen–antibody binding. Other solutions can be used, for example buffers containing chelator agent (EDTA, EGTA) for metallo-dependent interactions. Different regeneration solutions are proposed in (13).
11. If complex samples as serum or cell culture medium are injected, it is better as much as possible to saturate the chip using a control medium equivalent to the samples, but without the tested ligands (serum from healthy donors versus serum from patients, culture medium versus culture supernatant).

12. In the case of highly diluted ligand, it is better to use PBS–0.1% BSA as dilution buffer to avoid non-specific adsorption on the tube. In the same way, the use of silicone-treated tubes is recommended.
13. It is important to verify the stability of the chip upon the successive injection cycles and the SDS treatment. To this end, a standard solution containing the ligand for one of the grafted peptides is injected at the beginning of the experiment and then at regular intervals; the SPR signals obtained must be very similar from one control to the other.

References

1. Mah, A. S., Elia, A. E., Devgan, G., Ptacek, J., Schutkowski, M., Snyder, M., Yaffe, M. B., Deshaies, R. J. (2005) Substrate specificity analysis of protein kinase complex Dbf2-Mob1 by peptide library and proteome array screening. *BMC Biochem* **6**, 22.
2. Cherif, B., Villiers, C. L., Paranhos-Baccalà, G., Roget, A., Marche, P. N., Livache, T., Villiers, M.-B. (2006) Design and application of a microarray for peptide–antibody interactions analysis by fluorescence or SPR imaging. *J Biomed Nanotech* **2**, 29–35.
3. Reineke, U., Ivascu, C., Schlief, M., Landgraf, C., Gericke, S., Zahn, G., Herzel, H., Volkmer-Engert, R., Schneider-Mergener, J. (2002) Identification of distinct antibody epitopes and mimotopes from a peptide array of 5520 randomly generated sequences. *J Immunol Methods* **267**, 37–51.
4. Cretich, M., Damin, F., Pirri, G., Chiari, M. (2006) Protein and peptide arrays: Recent trends and new directions. *Biomol Eng* **23**, 77–88.
5. Lee, Y. S., Mrksich, M. (2002) Protein chips: from concept to practice. *Trends Biotechnol* **20**, S14–8.
6. Desai, A., Ravi, V., Satishchandra, P., Khanna, N., Rao, S., Gourie-Devi, M., Chandramuki, A. (2000) HIV antibody profiles in serum and CSF of patients with neurological disease can serve as predictors of outcome. *J Neurol Neurosurg Psychiatry* **68**, 86–8.
7. Sanchez-Rodriguez, C., Estrada-Chavez, C., Garcia-Vigil, J., Laredo-Sanchez, F., Halabe-Cherem, J., Pereira-Suarez, A., Mancilla, R. (2002) An IgG antibody response to the antigen 85 complex is associated with good outcome in Mexican Totonaca Indians with pulmonary tuberculosis. *Int J Tuberc Lung Dis* **6**, 706–12.
8. Zhang, L., Zhang, F., Yu, W., He, T., Yu, J., Yi, C. E., Ba, L., Li, W., Farzan, M., Chen, Z., Yuen, K. Y., Ho, D. (2006) Antibody responses against SARS coronavirus are correlated with disease outcome of infected individuals. *J Med Virol* **78**, 1–8.
9. Yu, X., Xu, D., Cheng, Q. (2006) Label-free detection methods for protein microarrays. *Proteomics* **6**, 5493–503.
10. Cherif, B., Roget, A., Villiers, C. L., Calemczuk, R., Leroy, V., Marche, P. N., Livache, T., Villiers, M. B. (2006) Clinically related protein–peptide interactions monitored in real time on novel peptide chips by surface plasmon resonance imaging. *Clin Chem* **52**, p255–62.
11. Livache, T., Roget, A., Dejean, E., Barthet, C., Bidan, G., Teoule, R. (1994) Preparation of a DNA matrix via an electrochemically directed copolymerization of pyrrole and oligonucleotides bearing a pyrrole group. *Nucleic Acids Res* **22**, 2915–21.
12. Jirkowsky, I., Baudy, R. (1981) A facile large scale preparation of 1H-pyrrol-1 ethanamine and syntheses of substituted pyrrolo[1,2-a] pyrazines and hydro derivatives thereof. *Synthesis* **6**, 481–3.
13. Andersson, K., Hamalainen, M., Malmqvist, M. (1999) Identification and optimization of regeneration conditions for affinity-based biosensor assays. A multivariate cocktail approach. *Anal Chem* **71**, 2475–81.

Chapter 18

The Peptide Microarray-Based Assay for Kinase Functionality and Inhibition Study

Zhenxin Wang

Abstract

We report a microarray format for the detection of kinase functionality/inhibition based on marking peptide phosphorylation/biotinylation events by the attachment of gold nanoparticles followed by silver deposition for signal enhancement. The detection principle is resonance light scattering (RLS) or surface-enhanced Raman spectroscopy (SERS). α-Catalytic subunit of cyclic adenosine 5′-monophosphate (cAMP) dependent protein kinase (PKA) and its well-known substrate, kemptide, are used for the purpose of monitoring phosphorylation and inhibition. As expected, highly selective inhibition of PKA is demonstrated with the four inhibitors: H89, HA1077, mallotoxin and KN62. Furthermore, inhibition assay with inhibitors demonstrate the ability to detect kinase inhibition as well as derive IC_{50} (half maximal inhibitory concentration) plots.

Key words: Kinase, inhibition, peptide array, resonance light scattering (RLS), surface-enhanced Raman scattering (SERS).

1. Introduction

The phosphorylation of proteins by kinases play various vital regulatory roles in metabolic pathways and cell communication (1, 2). The family of mammalian protein kinases consists of over 500 members of which only a fraction has been characterized. Therefore, the identification of kinases, their substrates and, in particular, potential inhibitors is necessary for the understanding of many fundamental biochemical processes. This knowledge is also necessary in drug discovery (3, 4). Radiolabelling of the substrate using γ-$^{32/33}$P-ATP as co-substrate is the standard method to monitor kinase activity (1–4). High-throughput, non-radioactive alternative methods based on microarray or other

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_18,

techniques, e.g. fluorescent/luminescent labels and mass spectrometry have been used to monitor kinase activity on the surface of peptide chips (5, 6). These techniques have enormous benefits to biomedical and kinase-based research in particular.

Here, we report a highly sensitive, selective and simple microarray method, which can be employed for kinase inhibition study. As in our previous report, this method is based on labelling peptide phosphorylation events on a microarray with gold nanoparticles using avidin–biotin chemistry followed by silver enhancement and RLS/SERS detection (7–9, *see* **Notes 1** and **2**). We demonstrate that it is possible to screen kinases with single or multiple inhibitors simultaneously on the same microarray format, thereby increasing assay throughput in this proof-of-concept study.

2. Materials

1. Tetrachloroaurate ($HAuCl_4$), bovine serum albumin (BSA), biotinylated monoclonal anti-phosphoserine antibody (anti-phosphoserine-biotin produced in mouse) and succinylated avidin were purchased from Sigma-Aldrich Co. (USA).
2. Succinylated avidin solution, 1 mg/ml succinylated avidin in 50 mM PBS, 0.15 M NaCl, pH 7.5.
3. Avidin, fluorescein conjugate was purchased from Invitrogen Co. (USA).
4. α-Catalytic subunit cyclic adenosine 5′-monophosphate (cAMP) dependent protein kinase (PKA) was purchased from New England Biolabs (Beverly, MA, USA).
5. Peptides (e.g. kemptide (LRRASLG), control peptide (LRRAGLG), CALNN and CALNNGK(biotin)G) were purchased from Shenggong Ltd. (Shanghai, China).
6. Biotin–ATP (adenosine 5'-triphosphate [γ]-biotinyl-3, 6, 9-trioxaundecanediamine (ATP [γ]biotin-LC-PEO-amine)) was purchased from Alt. Inc. (Lexington, KY, USA).
7. Inhibitors (H89, HA1077, mallotoxin and KN62) were purchased from Calbiochem-Novabiochem Corp. (La Jolla, CA. USA).
8. NHS–ester hydrogel-coated glass microscope slides (Nexterion® Slide MPX 16) were purchased from Schott AG (Germany).
9. The MPX 16 slide consists of a coated glass slide partitioned into 16 individual wells by an ultra-hydrophobic patterned layer.
10. Other chemicals were analytical grade and used as received. Milli-Q water (18.2 MΩ) was used in all experiments.

3. Methods

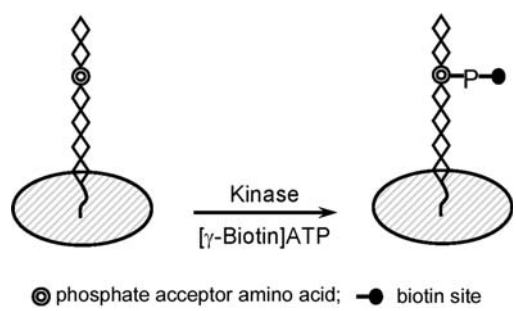
3.1. Preparation of Gold Nanoparticle Probes

The citrate-stabilized 13 nm gold nanoparticles were synthesized by Frens–Turkevich method (10, 11). Avidin-stabilized gold nanoparticles were prepared by stirring an aqueous mixture of 2 ml succinylated avidin solution and 20 ml of citrate-stabilized 13 nm gold nanoparticles (3.8×10^{-9}M) for 30 min at room temperature. Excess protein was removed by repeated centrifugation at 13,000 rpm (~16,100*g*, 3 ×) using an Eppendorf centrifuge (Eppendorf, Germany). The gold nanoparticles were re-suspended in PBS and stored at 4°C.

Peptide-stabilized nanoparticles were prepared by our previously peptide capping procedure (12, 13). Generally, an aqueous solution of peptide mixture (CALNN: CALNNGK(biotin)G) was added to the solution of 2.7 nM 13 nm gold nanoparticles to give a final concentration of total peptides of 1.5 mM. The ratio of CALNN and CALNNGK(biotin)G in the mixture is 9:1. Excess peptides were removed by repeated centrifugation at 13,000 rpm (~16,100*g*, 3 ×) using an Eppendorf centrifuge. After centrifugation, the purified gold nanoparticles were re-suspended in PBS and stored at 4°C.

3.2. Peptide Microarray Fabrication and Phosphorylation

Multiple copies of two different peptides, kemptide (LRRASLG, substrate of PKA) and control peptide (LRRAGLG which is not a substrate of PKA) were spotted and immobilized on commercial NHS–ester MPX 16 glass microscope slides by a standard robotic procedure (*see* **Note 3**). The peptide, kemptide (LRRASLG), is phosphorylated on the microarray spot by the kinase, PKA, using biotinylated ATP (Biotin-ATP) as a co-substrate (as shown in **Scheme 18.1**).Through this reaction, together with the phosphate, a biotin site is transferred to the kemptide. In a subsequent step, the biotinylated peptide is specifically labelled by attachment of avidin-



Scheme 18.1. Schematic representation of phosphorylation/biotinylation of the spotted peptide.

stabilized or peptide-stabilized gold nanoparticles followed by the conventional silver enhancement step and finally, detection by RLS or SERS.

3.2.1. Preparation of Assays for RLS Detection

1. The substrate peptide (kemptide) is phosphorylated on the microarray spot by the kinase, PKA, using biotinylated ATP (Biotin-ATP) as a co-substrate (*see* **Note 4**).
2. The biotinylated peptide is specifically labelled by attachment of avidin-stabilized gold nanoparticles (*see* **Note 5**).
3. A silver enhancement step was applied to the microarrays for signal amplification since the light scattering properties of gold nanoparticles by themselves are relatively poor, if the particles are smaller than ca. 40 nm. It was found that the detection sensitivity and dynamic range was critically related to the amount of silver deposited (*see* **Note 6**).

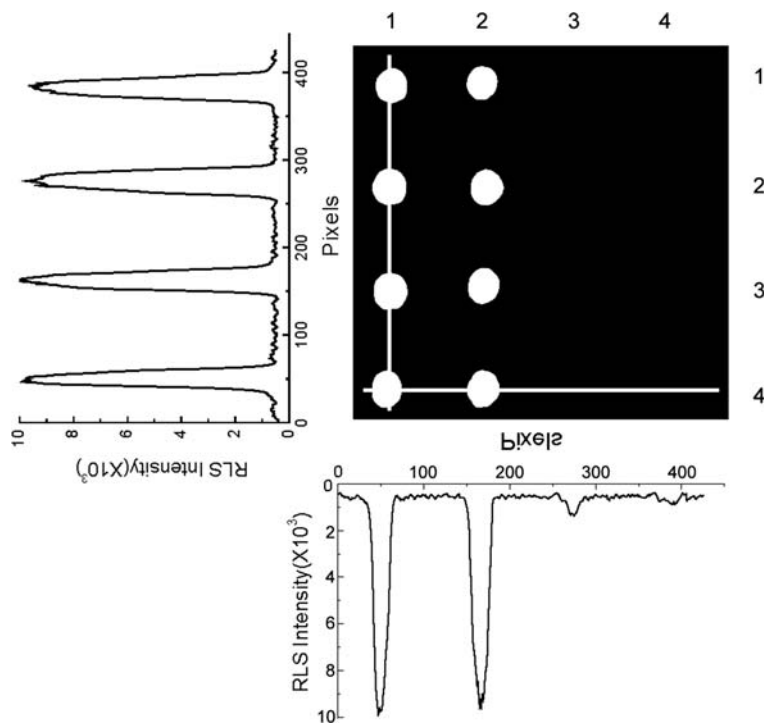


Fig. 18.1. Light scattering image of peptide microarray after phosphorylation, labelling positives with gold nanoparticles, enhancement by silver deposition and corresponding RLS intensity curves of column 1 and row 4 of the array (marked by *white lines*). Kemptide was spotted on columns 1 and 2, while control peptide was spotted at columns 3 and 4. The concentrations of spotted peptides were 1 μg/ml. Reproduced from (8) with permission from ACS.

4. Upon signal amplification by silver deposition, the slides were imaged with a highlight scanner using white light (Qiagen Instruments, Switzerland) and analysed with the Array Analyser software. Light scattering image of peptide microarray is shown in **Fig. 18.1** (*see* **Note 7**)

3.2.2. Preparation of Assays for SERS Detection

1. Here, kemptide is phosphorylated on the microarray spot by PKA as described in **Section 3.2.1**.
2. The phosphorylated peptide is specifically recognized by anti-phosphoserine-biotin and the recognition event was labelled by avidin-conjugated fluorescein (*see* **Note 8**).
3. The avidin-conjugated fluorescein was specifically labelled by attachment of peptide-stabilized gold nanoparticles followed by the conventional silver enhancement.
4. After signal amplification by silver deposition, the slides were detected with a Renishaw 2000 Raman spectrophotometer (Gloucestershire, UK) equipped by an Ar+ ion laser and a CCD detector. All spectra were collected with an excitation line at 514 nm with 10 s accumulation time, and 10 MW incident power. The detection signal is shown in **Fig. 18.2**.

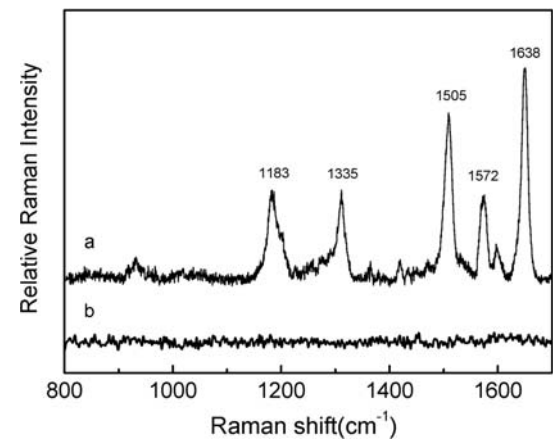


Fig. 18.2. SERS spectra of fluorescein acquired from one of spot with (**A**) kemptide or (**B**) control peptide in the microarray. The concentration anti-phosphoserine-biotin in the probing solution was 20 μg/ml and the concentration of peptides in spotted solution is 10 μg/ml. Reproduced from (9) with permission from Elsevier Science.

3.3. PKA Inhibition and IC_{50} Assays

1. The inhibition efficiency test was performed using the inhibition solution (kinase solution with each inhibitor at 10 μM) on different subarrays.
2. To measure the IC_{50} values of inhibitors (H89, HA1077, mallotoxin and KN62), different concentrations of the inhibitor (0.0001–50 μM) in the kinase solution were

applied to individual subarray, following incubated, washed spun-dry and probed by gold nanoparticles as described previously.

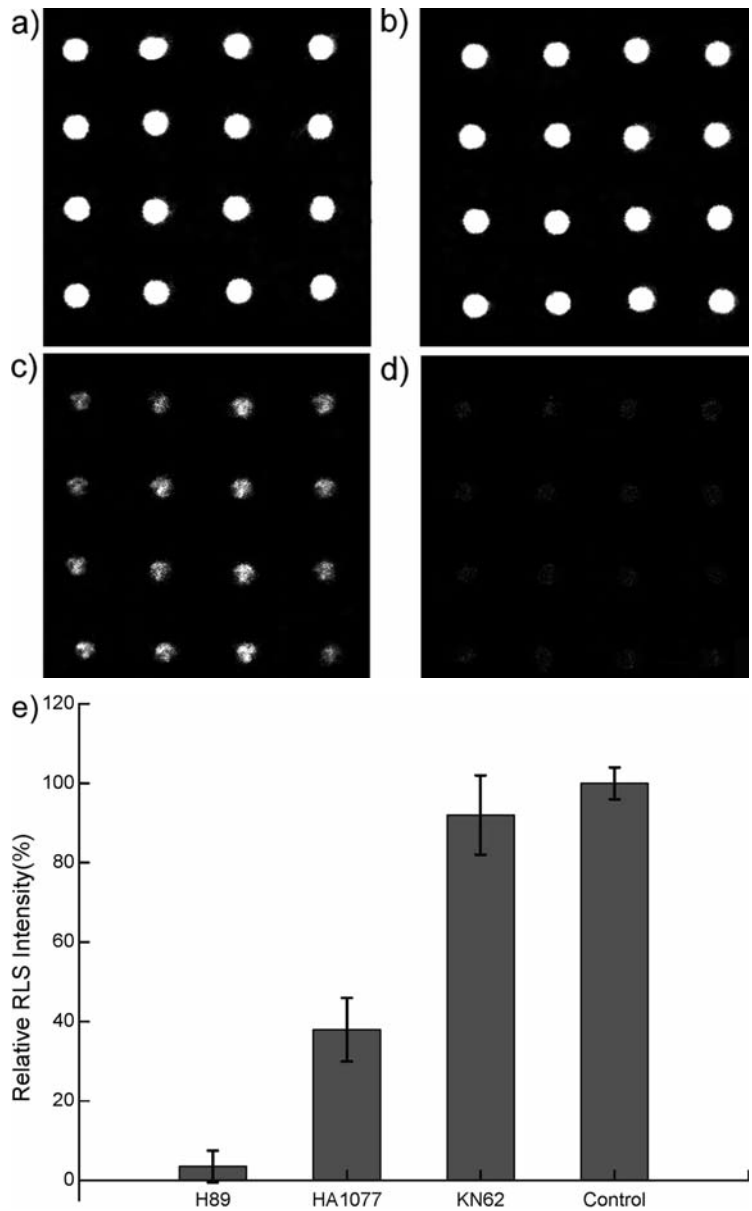


Fig. 18.3. Light scattering images of peptide microarrays with (**A**) in the absence of inhibitor (control experiment) and in the presence of 10 μM inhibitors: KN62(**B**), HA1077(**C**) and H89(**D**). (**E**) shows the preliminary quantitative analysis of RLS signals from arrays (**A**) to (**D**). The signals have been corrected for background noise and normalized to the average RLS intensity obtained in the absence of inhibitors. The concentration of spotted kemptide was 1 μg/ml. Reproduced from (8) with permission from ACS.

3. The slides were imaged with a highlight scanner using white light and analysed with the Array Analyser software. Light scattering images and quantitative analysis of peptide microarrays are shown in **Fig. 18.3** (*see* **Note 9**).

4. Notes

1. Gold nanoparticles exhibit the ability to resonantly scatter visible and near-infrared light (generally called as resonance light scattering, RLS). This property is the result of the excitation of surface plasmon resonances (SPR) and is extremely sensitive to size, shape and aggregation state of the particles. The RLS of gold nanoparticles can be enhanced after silver deposition.
2. Surface-enhanced Raman scattering (SERS) is a useful technique resulting in strongly increased Raman signals from molecules which have been attached to nanometre-sized metallic structures. It is generally agreed that more than one effect contributes to the observed large effective SERS cross-section. The enhancement mechanisms are roughly divided into electromagnetic and chemical effects.
3. We found that the PKA has excellent reaction activity on the NHS–ester modified slides.
4. Adenosine 5'-triphosphate [γ]-biotinyl-3, 6, 9-trioxaundecanediamine (ATP [γ]biotin-LC-PEO-amine) is an excellent replacement of ATP.
5. Biotin (also known as vitamin H) is a small organic molecule found in every cell. Avidin (also called strepavidin) is a much larger protein that binds biotin with a very high affinity.
6. We have found that the detection sensitivity and dynamic range were critically related to the amount of silver deposited. No spots can be detected when exposure time is less than 1 min. For short exposure times (2–4 min), there was poor sensitivity while long exposure times (>15 min) had significant background noise and saturated signal output which also reduced sensitivity and dynamic range. With the optimum exposure time of ~8 min, good features of the spots and a relatively higher signal-to-background noise ratio can be achieved.
7. An optimal signal-to-noise ratio was obtained at concentrations above 1 ng/ml. At this detection limit the actual total amount of kemptide on each spot (1 nl spotting volume) is equal to or less than 1 fg (1.5 amol). This is by at least three orders of magnitude more sensitive than currently reported fluorescence-based kinase functionality arrays.

8. The Raman-activity of the compound (fluorescein) is relatively weaker than Raman-activity of other dyes (e.g. Rhodamine B). The different Raman dyes can be excited with a single wavelength laser, which would greatly reduce the cost of instrumentation. Furthermore, the half-peak width of a Raman band is much narrower than the half-peak width of a fluorescent emission band.
9. As expected, in the presence of inhibitors, the RLS intensity decreases with increasing efficiency of inhibitor. This suggests that our method not only has the potential to screen for inhibitor activity qualitatively, but can also be developed to yield quantitative data on the efficiency of different inhibitors. More efficient inhibitors lead to lower levels of biotinylation, which in turn gave a decreased RLS intensity.

Acknowledgments

The author would like to thank Professors David G. Fernig and Mathias Brust for their helpful discussions (The University of Liverpool) and the NSFC (Grant No. 20675080) for financial support.

References

1. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002) The Protein Kinase Complement of the Human Genome. *Science* **298**, 1912–1934.
2. Espanel, X., Walchli, S., Ruckle, T., Harrenga, A., Huguenin-Reggiani, M., and van Huijsduijnen, R. H. (2003) Mapping of Synergistic Components of Weakly Interacting Protein—Protein Motifs Using Arrays of Paired Peptides. *J. Biol. Chem.* **278**, 15162–15167.
3. Noble, M. E. M., Endicott, J. A., and Johnson, L. N. (2004) Protein Kinase Inhibitors: Insights into Drug Design from Structure. *Science* **303**, 1800–1805.
4. Cohen, P. (2002) Protein Kinases: The Major Drug Targets of the Twenty-First Century? *Nat. Rev. Drug Discovery* **1**, 309–315.
5. Sun, H., Lu, C. H. S., Uttamchandani, M., Xia, Y., Liou, Y.-C., and Yao, S. Q. (2008) Peptide Microarray for High-Throughput Determination of Phosphatase Specificity and Biology. *Angew. Chem. Int. Ed.* **47**, 1698–1702.
6. Lesaicherre, M.-L., Uttamchandani, M., Chen, G. Y. J., and Yao, S. Q. (2002) Antibody-Based Fluorescence Detection of Kinase Activity on a Peptide Array. *Bioorg. Med. Chem. Lett.* **12**, 2085–2088.
7. Wang, Z., Lee, J., Cossins, A. R. and Brust, M. (2005) Microarray-Based Detection of Protein Binding and Functionality by Gold Nanoparticle Probes. *Anal. Chem.* **77**, 5770–5774.
8. Sun, L., Liu, D., and Wang, Z. (2007) Microarray-Based Kinase Inhibition Assay by Gold Nanoparticle Probes. *Anal. Chem.* **79**, 773–777.
9. Li, T., Guo, L., and Wang, Z. (2008) Microarray based Raman spectroscopic detection with gold nanoparticle probes. *Biosens. Bioelectron.* **23**, 1125–1130.
10. Turkevich, J., Stevenson, P. C., and Hillier, J. (1951) A study of the nucleation and growth processes in the

synthesis of colloidal gold. *Discuss. Faraday Soc.* **11**, 55–75.

11. Frens, G. (1973) Controlled nucleation for the regulation of the particle size in monodisperse gold suspensions. *Nature Phys. Sci.* **241**, 20–22.
12. Wang, Z., Levy, R., Fernig, D. G., and Brust, M (2006) Kinase-Catalyzed Modification of Gold Nanoparticles: A New Approach to Colorimetric Kinase Activity Screening. *J. Am. Chem. Soc.* **128**, 2214–2215.
13. Levy, R., Thanh, N. T. K., Doty, R. C., Hussain, I., Nichols, R. J., Schiffrin, D. J., Brust, M., and Fernig, D. G. (2004) Rational and Combinatorial Design of Peptide Capping Ligands for Gold Nanoparticles. *J. Am. Chem. Soc.* **126**, 10076–10084.

Chapter 19

An Advanced Application of Protein Microarrays: Cell-Based Assays for Functional Genomics

Roberta Carbone

Abstract

Microarrays have become common tools for approaching different experimental questions: DNA, protein and peptide arrays offer the power of multiplexing the assay and by means of miniaturization technology, the possibility to reduce cost and amount of samples and reagents.

Recently, a novel technology for functional assays has been proposed. Sabatini and co-workers have shown a cell-based microarrays method (1) that relies on the deposition and immobilization of an array of cDNA plasmids on a slide where cells are subsequently plated; the cDNA is then internalized by "reverse transfection" and cells overexpress or downregulate in each single spot the genes of interest. This approach allows the screening of different phenotypes in living cells of many genes in parallel on a single slide.

To overcome some relevant limitations of this approach, we have implemented the technology by means of viral immobilization (2) on a novel surface of cluster-assembled nanostructured TiO_2 (3) previously functionalized with an array of a docking protein.

In this work, we present the detailed development of the "reverse infection cell-microarray based technology" in U2OS cells on a novel coated slide that represents an advanced application of protein arrays.

Key words: Protein arrays, streptavidin, nanostructured TiO_2 slide, cell-based microarray, retrovirus, functional genomic screenings, automated microscopy.

1. Introduction

Cell microarrays represent a technological challenge: in order to set up the protocol, four "elements" must be selected and optimized for the specific aim of the assay; the cellular model (primary, tumour, stem cells), the genes that must be spotted on the slide in array format, the method of gene transduction that must be highly efficient for the specific cell line and finally the glass slide that must be biocompatible for cellular growth

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_19,

at different conditions (time and treatment), allowing localized and stable binding of the gene on the spot in array format and its efficient release upon cell plating.

We have recently proposed a novel biocompatible substrate for cell culturing obtained by coating glass slides with cluster-assembled nanostructured TiO_2(4) produced by supersonic cluster beam deposition of titania clusters. This coating results from a random stacking of nanoparticles that creates a highly porous and transparent thin film suitable for cellular growth (even at long term) (3); moreover, thanks to the high surface area, the film can be efficiently functionalized with proteins and peptides (3, 5).

We have set up the cell-based microarray assay with the following features: (a) human cancer cells as cellular model; (b) GFP-based retroviral vectors as gene transduction system; (c) cluster-assembled nanostructured TiO_2-coated slide as substrate for viral functionalization and cell growth; and (d) streptavidin functionalization of the slide and biotinylation of Moloney Murine Leukemia Virus (6) as viral arraying and immobilization system (**Fig. 19.1**). The mechanism by which the Moloney Murine Leukemia Virus particles enter cells is mainly through viral membrane fusion with target cells (7); therefore, retroviral particles can be stably docked on a streptavidin-functionalized surface, while the genetic elements of the virus can enter cells. This approach of "reverse infection", where viruses are immobilized on a surface and cells are plated on top, increases the probability of cell–virus interaction and the subsequent gene transduction events, even in absence of toxic polycations, normally required to increase the infection efficiency (8); finally the biocompatible nanostructured TiO_2 surface provides correct conditions of cell growth.

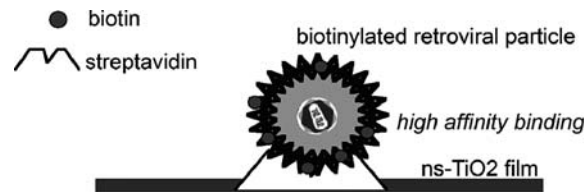


Fig. 19.1. Schematic representation of retroviral immobilization on nanostructured TiO_2 (Ns-TiO_2) coated slides: upon streptavidin functionalization of the slides in array format, biotinylated retroviral particles are spotted on *top* of the array; the high affinity binding between streptavidin and biotin ensures the precise localization of the viruses on the spot.

2. Materials

2.1. Slides

Nanostructured TiO_2 coating (average thickness, 50 nm) is prepared on glass slides (Nexterion glass D, cleanroom cleaned, SCHOTT AG, Mainz, Germany) by depositing under high vacuum a supersonic seeded beam of TiO_x clusters produced by a pulsed microplasma cluster source (PMCS). A detailed description of the PMCS and its principle of operation can be found in (4, 9). The slides are available through the company Tethis S.r.l., Italy.

2.2. Cell Transfection, Virus Biotinylation and Concentration

(1) All these procedures assume the use of a tissue culture room with sterile hood, incubators and other standard laboratory apparata.

(2) Phoenix Amphotropic cells for retroviral plasmid transfection are kindly provided by G. Nolan Lab (http://www.stanford.edu/group/nolan/retroviral_systems/phx.html).

(3) Dulbecco's Modified Eagle's Medium (DMEM) (Gibco/BRL, Bethesda, MD) supplemented with 10% fetal bovine serum (FBS, Invitrogen Corporation, Carlsbad CA), 1% (w/v) L-glutamine (SIGMA, St.Louis, MO) and 1% penicillin–streptomycin mixture (Lonza Group, Basel, Switzerland): when used in this formulation it is defined as complete medium.

(4) Retroviral plasmids: PINCO (10) and PINCO-NPM (kindly provided by Dr. Emanuela Colombo, European Institute of Oncology, Milan, Italy).

(5) 2 M $CaCl_2$, prepare filter-sterilized stocks and store at –20°C.

(6) 2XHBS Hank's Balanced Solution: 50 mM Hepes, 10 mM KCl, 12 mM dextrose, 280 mM NaCl, 1.5 mM Na_2HPO_4; filter, sterilize and store in aliquots at –20°C.

(7) Chloroquine (SIGMA): prepare as stock in sterile ddH_2O at 100 mM and store at –20°C.

(8) Biotinamidohexanoic acid *N*-hydroxysuccinimide ester (biotin, SIGMA): prepare at 333.5 mM in sterile dimethyl sulphoxide (DMSO) and store in single-use aliquots at –20°C; do not re-use after thawing.

(9) Hank's Modified Medium: 1.66 mM $CaCl_2$, 1 mM glucose, 0.81 mM $MgSO_4$, 5.37 mM KCl, 0.44 mM KH_2PO_4, 4.12 mM $NaHCO_3$, 0.13 M NaCl, 0.32 mM Na_2HPO_4; filter, sterilize and store at 4°C.

(10) 0.45 μm filter MILLEX®HA for viral clarification (Millipore Corporation, Billerica, MA).

(11) PEG8000: prepare a stock solution of 30% (w/v) PEG8000 (SIGMA) plus 0.9% (w/v) NaCl in sterile ddH_2O and store in aliquots at 4°C.

(12) Dulbecco's phosphate buffered saline (DPBS) (Lonza Group).

(13) Bovine serum albumin (BSA) (Albumin FractionV BDH Biochemicals, VWR International Ltd., Lutterworth, UK): prepare a 2% (w/v) stock solution in 1 × DPBS and store in aliquots at –20°C.

(14) D+trehalose (SIGMA): dilute in sterile ddH_2O at 40% (w/v) and store at +4°C.

(15) 70% Ethanol and dry ice for fast freezing of viral aliquots.

2.3. Slide Functionalization

(1) Spotter: Biodot® BioJet Plus spotter with Axsys software, (BioDot Inc., Irvine CA).

(2) Streptavidin from *Streptomyces avidinii* (SIGMA) dissolved in sterile ddH_2O at 2 mg/ml aliquoted and stored at –20°C.

(3) 96-well plates (BD Bioscience, San Jose, CA) as source plate for spotting.

(4) Spotting buffer: 10 mM Hepes and 150 mM NaCl pH 7.4 prepared by diluting stocks of 1 M Hepes at pH 7.4 and 1 M NaCl; filter, sterilize.

(5) Washing buffer: 1 × DPBS/0.5% (w/v) Tween 20 (SIGMA).

(6) Washing buffer: 1 × DPBS/0.05% (w/v) Tween 20.

(7) Washing buffer: 1 × DPBS.

(8) Blocking buffer: complete medium.

(9) Vivadish™ plates (Vivascience AG, Sartorius Stedim Biotech S.A. Aubagne Cedex, France).

2.4. Target Cells

(1) Human osteosarcoma U2OS (American Type Culture Collection – ATCC), cultured in DMEM supplemented with 10% FBS (Invitrogen), 1% (w/v) L-glutamine (SIGMA) and 1% penicillin–streptomycin mixture (Lonza).

(2) Trypsin (0.25%) and ethylenediamine tetraacetic acid (EDTA) (1 mM) (Gibco/BRL).

(3) 100 mm tissue culture plates (Falcon, BD Bioscience).

2.5. Immunofluorescence Experiment, Acquisition and Analysis

(1) Prestige rectangular coverslips (22 × 40 × 0.15 mm^3) from Zeuss Super (Zeuss Padova, Italy).

(2) Paraformaldehyde 4% in Pipes buffer: prepare a 4% (w/v) solution in Pipes; Pipes buffer: 80 mM Pipes, 5 mM EGTA, 2 mM $MgCl_2$, prepared by stocks of 400 mM Pipes pH 6.8, 500 mM EGTA pH 8, 1 M $MgCl_2$: prepare fresh for each experiment. The solution may need to be carefully heated (use a stirring hotplate in a fume hood) to dissolve, and then cooled to room temperature for use.

(3) Washing buffer: 1 × DPBS.

(4) Nuclear counterstaining: DAPI (4,6-diamidino-2-phenylindole, SIGMA): prepare a 1 mg/ml stock in ddH_2O, store in aliquot at –20°C, in the dark. For nuclear counterstaining, use 1:5000 dilution in 1 × DPBS.

(5) Mounting media: Mowiol (Mowiol 4–88 CALBIOCHEM, Merck KGaA, Darmstadt, Germany): prepare the following reagents: glycerol, ddH_2O, 0.2 M Tris pH 8.5; place 6 g of glycerol in 50 ml disposable conical Falcon centrifuge tube; add 2.4 g Mowiol and mix thoroughly; add 6 ml of ddH_2O and leave for 2 h at room temperature; add 12 ml of 0.2 M Tris-HCl pH 8.5 and incubate at approximately 53°C until the Mowiol has dissolved. Vortex occasionally; clarify by centrifugation at 4000 rpm (3345*g*) for 20 min; transfer the supernatant into glass vials with screw caps (about 1 ml in each). Store in aliquots at –20°C (*see* **Note 1**).

(6) Scan^R, a microscope-based platform for imaging phenotype screening (Olympus Europa, Germany) equipped with an Orca ER C4742-80 Hamamatsu B/W Camera (Hamamatsu Italy, Italy). High numerical aperture planapochromatic dry lenses are employed to maximize resolution and light collection under screening acquisition conditions. Single 10 × and 20 × fields per spot are acquired.

(7) Image J (freely available at www.rsb.info.nih.gov/ij/) for image analysis and array visualization.

3. Methods

In this protocol we report the development of an optimized retroviral microarray on a streptavidin-functionalized nanostructured TiO_2 slide for phenotype screenings.

This method provides up to 70% of transduction efficiency of GFP-based retroviral vectors in osteoblastoma cells, in absence of polybrene, in an arrayed format on a chip. To

achieve this result, it is important to obtain a stable and efficient functionalization with streptavidin of the nanostructured TiO_2 slide; and a high titre retroviral supernatant and efficient viral biotinylation. Furthermore target cells must be actively dividing to allow gene transduction.

The use of GFP-based retroviral plasmids helps in setting up of the protocol, allowing a precise and quick evaluation of the array in terms of localization of spots; two GFP-based retroviral vectors, localizing in two different cellular compartments (cytoplasm and nucleoli), are spotted in alternate positions on the streptavidin-functionalized nanostructured TiO_2 slide (**Fig. 19.2**). From image acquisition and analysis of reverse infected cells, it is possible to confirm the specific localization of the GFP overexpressing cells on the spot, the localization of the GFP gene in single cells, the percentage of the infection per spot and the absence of inter-spots background.

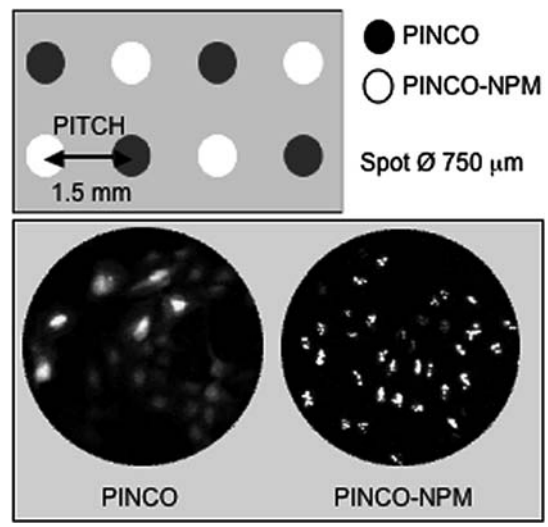


Fig. 19.2. Features of the retroviral array spotting scheme: PINCO vector localizes in the cytoplasm while PINCO-NPM localizes in the nucleolus; by spotting in alternate positions, with a pitch of 1.5 mm and a spot diameter of 750 μm, a precise pattern of over-expressed genes/spot in different cellular compartments can be obtained.

The successful setup of this protocol provides the basic tools for further assay development with complex libraries of genes.

3.1. Viral Preparation

3.1.1. Day 1: Plating of Cells for Transfection of Retroviral Vector DNA

(1) Phoenix Amphotropic cells are cultured in complete medium and passaged when 80% confluency is reached: it is recommended to avoid long periods of culturing (less that one month) to prevent the loss of the episomal viral genes and growing capability. Cells are grown in 100 mm plastic dishes

at 37°C in a humidified incubator with 5% CO_2 and kept in logarithmic growing condition until plating for transfection experiments.

(2) The day before the transfection, plate 3.5×10^6 Phoenix Amphotropic cells in complete medium on 100 mm plates and place in 37°C humidified incubator at 5% CO_2: use one plate for each retroviral plasmid to be transfected (*see* **Note 2**). Each plate will finally give 450 µl of concentrated retroviral supernatant.

3.1.2. Day 2: Transfection of Retroviral Plasmid PINCO and PINCO-NPM

1) All these procedures must be performed under sterile hood; culture medium must be warmed at 37°C in a water bath. For each retroviral plasmid to be transfected prepare the following solutions in a 15 ml Falcon tube: solution A composed by a mix of 10 µg of the retroviral plasmid (PINCO or PINCO-NPM), 72 µl of 2M $CaCl_2$, in a final volume of 500 µl of H_2O; solution B composed by 500 µl of 2 × HBS solution. Add the solution A to the solution B: to enhance DNA complexation with $CaCl_2$ bubbling with a pipette aid in the solution A is required during the adding of solution B (*see* **Note 3**).
2) Incubate the A plus B mix for 10 min at room temperature under the hood.
3) 6 min after mixing add 4 µl of chloroquine 100 mM (working concentration 40 µM) to the cells, then gently shake the plate (*see* **Note 4**).
4) 4 min later add the A plus B solution to the cells, gently mix and put back in the incubator (*see* **Note 5**).
5) After 6 h replace the medium with 10 ml of complete medium and incubate overnight in the 37°C incubator.

3.1.3. Day 3: Retroviral Biotinylation

1) Prepare the biotinylation solution by adding 10 µl of biotinamidocaproate *N*-hydroxysuccinimide ester to 10 ml of Hank's Modified Medium per plate.
2) Remove completely and carefully the cell medium from the plate (*see* **Note 6**), then gently add the biotinylation solution.
3) Leave cells for 30 min in the dark under the sterile hood at room temperature; at the end of the incubation replace completely the biotinylation medium with 10 ml of complete medium (*see* **Note 7**).
3) Put back cells in the 37°C incubator. After 4 h replace the culture medium with 5 ml of complete medium.
4) Incubate plate overnight on 37°C incubator; during the incubation time cells release the biotinylated viruses in the cell culture medium. To consider the viral preparation

suitable for the retroviral arraying assay, it is recommended to prepare a plate to be used for testing the percentage of GFP fluorescence in U2OS cells in a classical infection procedure (*see 10*): at least 90% of infection of U2OS cells must be achieved; otherwise the transfection procedure must be repeated (*see* **Note 9**).

3.1.4. Day 4: Retroviral Collection

1) Collect the retroviral supernatant (approximately 4.5 ml considering that the medium volume is decreased during the incubation time at 37°C) with a syringe and transfer it into a 50 ml Falcon tube by applying a 0.45 μm filter to the syringe before dispensing to eliminate cell debris from the retroviral supernatant.
2) Add 1.6 ml of a solution of 30% (w/v) PEG8000 and 0.9% (w/v) NaCl (in ddH_2O) to the filtered retroviral supernatant: the final concentration of PEG8000 will be 8% (w/v).
3) Vortex the Falcon tube and place in a refrigerator at +4°C overnight to allow PEG complexation with viral particles (*see* **Note 10**).

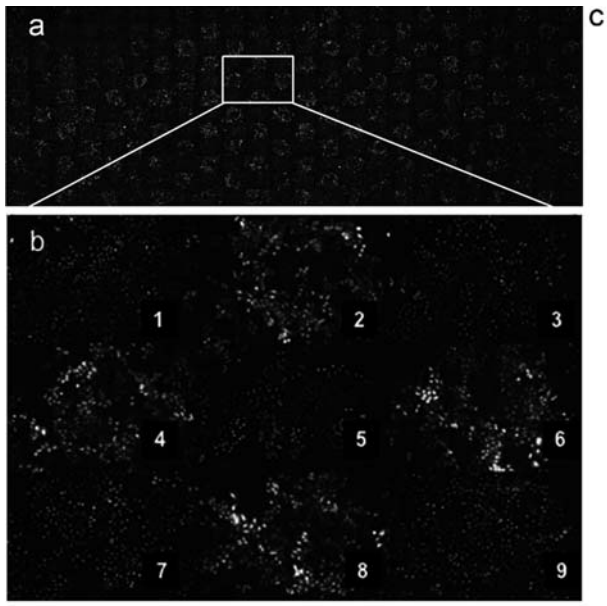
3.1.5. Day 5: Retroviral Precipitation, Concentration and Storage

1) Pellet the virus–PEG complexes by centrifuging tubes for 45 min at 3000 rpm in a refrigerated centrifuge (+4°C). Prepare a container with ice where to keep all the solutions and the virus–PEG complexes after centrifugation.
2) Discard the supernatant and carefully resuspend the pellet (*see* **Note 11**) in 450 μl of 0.1% (w/v) BSA and 5% (w/v) Trehalose in 1 × DPBS; this procedure allow to concentrate 10-fold the retroviral preparation.
3) Prepare 1 ml tubes with appropriate labels on the cap and aliquot the resuspended viral–PEG complexes in 50 μl/tube by snap freezing in ethanol 70% and dry ice.
4) Store the aliquots at –80°C until use (*see* **Note 12**). The procedure is resulting in 70–80% efficiency of retroviral biotinylation.

3.2. Retroviral Array Preparation

1) All these steps require to be performed in a cell culture room: the spotter must be located close to the sterile hood and the procedures need to be accurately followed using gloves, to keep sterility and cleanliness. The slide must be manipulated using clean forceps and possibly under the sterile hood until cell plating; the presence of antibiotics in cell culture medium is necessary to avoid bacteria contamination during cell growing.

2) The production of an array of proteins on slide requires the optimization of the operation conditions of the spotter: before the experiment the spotter must be thoroughly cleaned according the instrument specifications, and the program for the array must be created.
3) Prepare a solution of 5 μg/ml of streptavidin in 10 mM Hepes, 150 mM NaCl and place 30 μl of the solution in a 96-well plate.
4) The array of streptavidin molecules on nanostructured TiO_2-coated slides is prepared by spotting 40 nl of the streptavidin by a non-contact spotter (*see* **Note 13**) (Biodot® BioJet Plus spotter with Axsys software, BioDot Inc.). For the experiment (shown in **Fig. 19.3**), set the array at 12 spots per lane for 24 lanes with 1.5 mm pitch. The spotting procedure is performed at 70% constant humidity; after spotting the tip is extensively washed with degassed ddH_2O.
5) After spotting, place the slide in Vivadish™ plate and incubate overnight at room temperature in the presence of 75% humidity in a closed box with a NaCl saturated solution.
6) The day after, block the slide with complete medium for 1 h in a shaking platform at low speed at room temperature.
7) After blocking, wash the slides twice with 1 × DPBS/0.5% (w/v) Tween 20 and ddH_2O for 5 min: slides must be carefully dried under the sterile hood for 5 min (*see* **Note 14**); before the retroviral spotting it is recommended to store the slide in Petri dish at 4°C for short period (not more than 1 day).
8) Thaw an aliquot of the virus–PEG solution (PINCO and PINCO-NPM): place 40 μl of the solution into a 96-well plate and place the plate in the spotter's plate holder: use the spotting coordinates prepared for the streptavidin array; set the spotting scheme to obtain an alternate spotting of PINCO and PINCO-NPM (as shown in **Fig. 19.2**).
9) Carefully place the streptavidin-functionalized TiO_2 slide in the spotter's slide holder in the same position used with the streptavidin arraying procedure (*see* **Note 15**).
10) Dispense 40 nl of virus–PEG solution per spot in 65% humidity. Between the spotting of the two different retroviral preparations wash extensively the tip with sterile 1 × DPBS.
11) At the end of the spotting procedure, incubate the slide for 4 h at 4°C in a box with 75% humidity.
12) After the incubation, wash the slide twice with 1 × DPBS/0.05% (w/v) Tween 20 for 10 min and then twice with 1 × DPBS for 5 min each.



c

SPOT	CELL NUMBER	% INFECTION
1	507	74
2	478	62
3	586	73
4	545	69
5	636	77,8
6	611	65
7	607	75,4
8	529	60
9	671	67,8

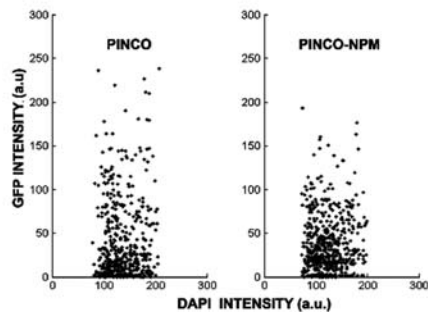


Fig. 19.3. (**A**) Panel of 12 × 24 retroviral array of U2OS cells acquired by ScanôR microscope at 10 × magnification: one image/spot has been collected and, by ImageJ software, montages of all the images have been produced: spots of infected cells are clearly identified by the fluorescent signal showing different localization of the two GFP-based retroviral vectors. (**B**) Magnification of a 3 × 3 array of PINCO and PINCO-NPM overexpressing cells: PINCO is represented by the following spots: 2, 4, 6, 8; PINCO-NPM by the 1, 3, 5, 7, 9 spots. (**C**) Image analysis using Image J software of the 3 × 3 array showing the number of cells/spot and the percentage of infection/spot; the graphs are a representative example of the scatter plot of DAPI intensity versus GFP intensity of one PINCO and one PINCO-NPM spot: the analysis provide a qualitative and quantitative evaluation of the retroviral array.

3.3. Cell Plating

1) Grow U2OS cells in 100 mm tissue culture plates at 37°C in a humidified incubator with 5% CO_2 until the day of the experiment.
2) Prepare U2OS cells for plating: detach cells from 100 mm plate with 1 ml of trypsin–EDTA, place the plate in the incubator and check when cells are completely detached from the plate: add 5 ml of complete medium and collect cells; centrifuge cells at 1200 rpm for 5 min, discard the supernatant and resuspend cells in 5 ml of complete medium. Count cells and dilute with medium appropriately to obtain 1.5×10^5 cell/ml.
3) Place the retroviral arrayed slide in a Vivadish™ plate and add 4 ml of the cell suspension to obtain 6×10^5 cells/ml (*see* **Note 16**): put the plate in the 37°C incubator.
4) After 72 h from plating, check the slide at the microscope to evaluate cell confluency and fluorescence signal: at 80–90% of cell confluency the slide can be processed for DAPI staining and image acquisition.

3.4. DAPI Staining and Image Acquisition

1) Completely remove the culture medium and add 4 ml 1 × DPBS to the Vivadish^{TM} plate: repeat the washing procedure for three times then fix cells with 4% paraformaldehyde for 10 min at room temperature.
2) Wash slides three times with 1 × DPBS and add 3 ml of DAPI solution: incubate for 5 min then wash with 1 × DPBS and with ddH_2O.
3) The slide is now ready for mounting with a coverslip: add one drop of the mounting media Mowiol on the coverslip and lay it down on the slide, carefully avoiding bubbles formation between the slide and the coverslip. The slide is ready for image acquisition and analysis.
4) Place the slide on the Scan^R automated microscopy station: prepare the acquisition program (based on the array configuration for the spot size, the pitch between spots and the fluorescence of the array) and after the identification of the first spot of the array, run the acquisition procedure. Usually 10 × or 20 × objectives are used to acquire the array. For GFP fluorescence use the FITC filter.

3.5. Image Analysis

After the acquisition, images are evaluated with ImageJ software and montages are prepared to obtain a 12 × 24 array (**Fig. 19.3**). The analysis of this image provides a qualitative evaluation of the experiment in term of cells distribution and localization of GFP overexpressing cells in precise spots.

The efficiency of infection can be measured on a statistically representative number of spots: as an example, for this analysis, select a 3 × 3 array of spots with four images of cells overexpressing PINCO and five images of cells overexpressing PINCO-NPM as shown in **Fig. 19.3**: by ImageJ analysis, after background subtraction, DAPI staining is employed to identify cell nuclei and to build a binary mask for successive calculations. The mean nuclear levels of DAPI and GFP for each image are measured by averaging fluorescence intensities in the areas identified by the nuclear mask. The threshold for infected and non-infected cells is identified as the mean GFP fluorescence of non-infected areas of the slides, plus one standard deviation.

4. Notes

1. The careful preparation of this reagent is very important since it affects the quality of the images that will be obtained by microscopy acquisition: check the viscosity of the final

solution that must be not too liquid, and if after freezing at –20°C the reagent remain completely liquid, discard it; use one aliquot for not more than 1 week after thawing.

2. The efficiency of retroviral transfection and the production of a high retroviral titre are strictly dependent on the quality of Phoenix Amphotropic cells and on the number of cells plated for the transfection: be sure to plate the exact number of cells and check if they are well adherent and not floating on the medium. Discard the plate if the day of transfection is too confluent (more than 60%).
3. A high retroviral titre is depending on DNA quality: the retroviral plasmid must be prepared from bacterial amplification by using a complete purification procedure (i.e. Maxi Prep by Quiagen or others supplies); check the concentration by spectrophotometer and verify that the 260/280 ratio is around 1.8. If this ratio is lower, repeat the purification procedure.
4. Add chloroquine by immerging the tip on a side of the plate after tilting the plate: then lay down the plate gently and move carefully up and down, left and right with the hands; the homogeneous distribution of chloroquine is necessary since this reagent prevents the eventual degradation of DNA inside cells, increasing the final percentage of transfection efficiency.
5. After adding on the Phoenix Amphotropic cells, the A plus B solution check by microscopy the formation and the quality of DNA–calcium precipitate: dark particles can be visible all over the cells; the dimension of this particles must be very small, like a fine powder; if aggregates are present, these can be toxic to cells and be responsible for the dropping down of the infection efficiency.
6. Since biotin reacts with proteins it is important to remove the medium completely to eliminate even small traces of proteins (derived from the serum): it is, however, *not* recommended to wash cells at this step since it can increase the probability of detaching them during the procedure of biotinylation.
7. This is a very delicate procedure: after 30 min of incubation in Hank's solution, cells can be loosely attached to the plate; it is crucial to remove and add the medium very gently to avoid perturbation of the cellular layer; however, a little percentage of detached cell do not represent an obstacle for the following steps of the procedure; only if all the cellular layer is detached it is recommended to start the procedure with another plate.
8. A good retroviral titre depends on the efficiency of tranfection: to evaluate this parameter it is important to inspect by microscopy the level of GFP fluorescence of Phoenix Amphotropic cells: by comparing the cell number by bright field inspection and the fluorescent cells it is possible to calculate

an approximate percentage of transfection: generally this procedure provides more than 90% of transfection efficiency on Phoenix cells: do not accept transfection efficiency lower than 70% since this will result in a low retroviral titre.

9. It is generally not necessary to verify the efficiency of retroviral biotinylation since the protocol is highly reproducible; however, in case of low infection efficiency in the retroviral array experiment it is possible to monitor the biotinylation efficiency as it was previously described (6).
10. The virus–PEG complexes can be left at 4°C for upto 3 days after preparation without losing infection capability.
11. After centrifugation the pellet must be white: be sure to remove completely the supernatant to avoid the remaining of free biotin and other low molecular weight contaminants in the virus–PEG complexes.
12. Aliquots can be left in the –80°C for upto 3 months without losing infection capability.
13. The streptavidin–virus spot size must be of 500–800 μm diameter: this is required to obtain around 300–600 cells/spot that represent a good cell number for statistical evaluation.
14. To perform a second round of spotting on top of streptavidin spots the slide must be completely dry and devoid of any moisture; eventually drain the excess of liquid and then leave the slide under the hood until completely dry.
15. It is recommended to exactly keep the order for slides positioning on the spotter slide holder for the two series of spotting (first streptavidin and then retrovirus); moreover, verify if the repositioning of the tip between two subsequently spotting procedures on the same spot is precise, to ensure good functionalization of the slides and low background.
16. Cell plating on the retroviral-arrayed slide must be homogeneous to have similar number of cells/spot: just after plating move carefully the plate to distribute the cell suspension and check by microscopy the absence of cell aggregation.

Acknowledgements

The author would like to thank Prof. Paolo Milani for providing nanostructured TiO_2-coated slides and Andrea Zanardi and Samuele Venturini for technical assistance. This work was supported by AIRC under OGCG grant "Development and integration of high-throughput technologies for the functional genomics of cancer".

References

1. Ziauddin, J., Sabatini, D. M. (2001) Microarrays of cells expressing defined cDNAs. *Nature* **411**, 107–110.
2. Carbone, R., Giorgetti, L., Zanardi, A., et al. (2007) Retroviral microarray-based platform on nanostructured TiO2 for functional genomics and drug discovery. *Biomaterials* **28**, 2244–2253.
3. Carbone, R., Marangi, I., Zanardi, A., et al. (2006) Biocompatibility of cluster-assembled nanostructured TiO2 with primary and cancer cells. *Biomaterials* **27**, 3221–3229.
4. Piseri, P., Vahedi-Tafreshi, H., Milani, P. (2004) Manipulation of nanoparticles in supersonic beams for the production of nanostructured materials. *Curr Op Solid State Mater Sci* **8**, 195–202.
5. Sousa, S. R., Lamghari, M., Sampaio, P., Moradas-Ferreira, P., Barbosa, M. A. (2008) Osteoblast adhesion and morphology on TiO2 depends on the competitive preadsorption of albumin and fibronectin. *J Biomed Mater Res A* **84**, 281–90.
6. Hughes, C., Galea-Lauri, J., Farzaneh, F., Darling, D. (2001) Streptavidin paramagnetic particles provide a choice of three affinity-based capture and magnetic concentration strategies for retroviral vectors *Mol Ther* **3**, 623–630.
7. Zavorotinskaya, T., Qian, Z., Franks, J., Albritton, L. M. (2004) A point mutation in the binding subunit of a retroviral envelope protein arrests virus entry at hemifusion. *J Virol* **78**, 473–481.
8. Themis, M., Forbes, S. J., Chan, L., et al. (1998) Enhanced in vitro and in vivo gene delivery using cationic agent complexed retrovirus vectors. *Gene Ther* **5**, 1180–1186.
9. Barborini, E., Piseri, P., Milani, P. (1999) A pulsed microplasma source of high intensity supersonic carbon cluster beams. *J Phys D: Appl Phys* **32**, L105–L9.
10. Grignani, F., Kinsella, T., Mencarelli, A., et al. (1998) High-efficiency gene transfer and selection of human hematopoietic progenitor cells with a hybrid EBV/retroviral vector expressing the green fluorescence protein. *Cancer Res* **58**, 14–19.

Chapter 20

Profiling the Autoantibody Repertoire by Screening Phage-Displayed Human cDNA Libraries

Roberto Di Niro, Sara D'Angelo, Paola Secco, Roberto Marzari, Claudio Santoro, and Daniele Sblattero

Abstract

The advent of the serological identification of antigens by procedures such as cDNA cloning and recombinant protein expression has allowed the direct molecular definition of immunogenic proteins. The phage-display technology provides several advantages over conventional immunoscreening procedures based on plasmid or lambda-phage cDNA libraries. So far, attempts to display open reading frames, such as those encoded by cDNA fragments, on filamentous phages have not been very successful. We managed to develop a strategy based on "folding reporters" which allows filtering out open reading frames from DNA and displaying them on filamentous phages in such a way that they are amenable to subsequent selection or screening.

Once the cDNA library of interest is created, phage-display technology is used for selection of novel putative antigens; these are then validated by printing isolated protein on microarray and screening with patients' sera.

Key words: Phage display, antigens, autoimmunity, protein microarray, recombinant fusion protein, open reading frames, Cre-recombinase, high throughput, serum profiling.

1. Introduction

Protein and peptide arrays represent a relatively novel tool with an outstanding potential for the dissection of the complex interactions between antigens and antibodies. Multiplex antigen arrays have been used to identify signature autoantibody profiles in human diseases (1, 2). Although these arrays have been usually constructed using known antigenic proteins, in autoimmunity the identification of novel self-antigens recognized by autoimmune sera has been long considered an important research goal, often leading to new diagnostic and therapeutic paradigms. Before the advent of molecular biology, such identification was difficult, and

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_20,

generally limited to either Western blotting or immunoprecipitation. In recent years novel proteomic strategies have been developed to identify the cohorts of antigens associated to autoimmune diseases. Phage display is a technique which involves the coupling of phenotype to genotype in a selectable format. It has been extensively used in molecular biology to study protein–protein interactions (5–7) and has a great potential for the study of complex molecular systems such as the antigen–antibody interaction networks. Phage-display expression libraries provided simple procedures for the identification of novel antigens, by allowing the screening of cDNA libraries created from appropriate tissues directly with autoimmune sera; this technology appears to be the most promising and affordable to set up high-throughput isolation of putative antigens. Phage-display cDNA libraries can be produced with a complexity that can easily exceed several millions different clones, each of them representing a protein or a protein fragment. Such libraries can be selected with antibodies (i.e. patients' sera), using settled protocols; this leads to a reduction of the library complexity and to the identification of a high number (thousands) of putative autoantigens. The antigenic properties of these proteins could be validated by means of protein array technology. Selected proteins are printed onto chips and analysed to obtain a detailed serum profiling of a given disease. Indeed, a large number of novel autoimmune antigens have been identified using the sera of patients affected by cancer (3) or autoimmune diseases (4).

The display technology based on filamentous phages offers several advantages over other procedures, such as lambda-phage screening: (a) as many as 10^{10}–10^{11} individual clones can be screened in a single experiment, thereby increasing the chance of isolating (auto)antigens; (b) the screening of the cDNA library is performed in a fluid-phase, so that denaturation of the proteins, which may occur on solid surfaces, is minimized; (c) low amounts of sera are needed to perform all the selection procedures; (d) the isolated phages carry the genes encoding for the binders of interest, and such genes can be easily identified and manipulated for further experiments.

In general, attempts to display random open reading frames (ORFs) on filamentous phage, such as those encoded by cDNA fragments, have not been very successful, despite the development of a number of different vectors (8–13). The main problems deal with the chance to clone a productive frame (one clone in three will start correctly, one clone in three will end correctly, and one clone in two will have the correct orientation), and with the ability of ORF fragments to fold properly. We managed to solve these problems by the use of a strategy based on "ORF reporters". We have designed a new vector (14, 15) which allows to filter out ORFs from

DNA and to display them on filamentous phage in such a way. that they are amenable to subsequent selection or screening. In this vector, cDNA fragments are inserted upstream a β-lactamase gene flanked by Cre-lox recombination sites. This gene fusion unit is upstream the gene coding for the phage capside protein g3p. The clones containing a DNA fragment encoding for an ORF will express a β-lactamase fusion protein and will grow in the presence of ampicillin. After this step, referred as ORF filtering, the β-lactamase gene is removed from DNA constructs by transforming *E. coli* cells expressing the Cre-recombinase. This enzyme excises the DNA within the lox sites and yields a standard phagemid vector that allows the display of an ORF-g3p fusion product on the phage surface. Filtered libraries are then available for selection and further analysis. After antigens isolation, protein production and purification, for array printing, could be carried on high-throughput systems. This will include microfermentation in 96-well format with novel medium formulation.

2. Materials

2.1. Construction of ORF-Enriched cDNA Libraries

1. cDNA is prepared from human fibroblasts mRNA, average size 350 bp, range 100–700 bp, ligated with adapters carrying BssHII (5′ end) and NheI (3′ end) restriction sites.
2. Bacterial strain used is *Escherichia coli* DH5αF′ (Gibco BRL), F′/*endA1 hsd17*(rK− mK+) *supE44 thi-1 recA1 gyrA*(Nalr) *relA1* _(*lacZYA-argF*) *U169 deoR*(F80*dlacD*-(*lacZ*)*M15*).
3. Plasmidic DNA (pPAO2 phagemid vector) is prepared using the Macherey–Nagel Nucleospin Plasmid kit, following the instructions of the manufacturer.
4. Stock solutions of antibiotics are prepared by dissolving chloramphenicol at 34 mg/mL in ethanol, kanamycin at 50 mg/mL in water and ampicillin at 100 mg/mL in water. Chloramphenicol stocks are kept at 4°C, kanamycin and ampicillin at –20°C. Repeated freeze and thaw of ampicillin is avoided, and aliquots are prepared for single use. All antibiotics are purchased from Sigma.
5. 2xTY liquid broth is prepared adding, for 1 L, 16 g bacto-tryptone, 10 g bacto-yeast and 5 g NaCl to ddH_2O. Final pH 7.0. Agar plates are prepared by adding 1.5% bacto-agar to 2xTY broth.

6. BssHII and NheI restriction endonucleases, T4 ligase and buffers are purchased from New England Biolabs. All cloning steps are performed according to the manufacturer suggestions and to standard molecular biology procedures.
7. The NucleoTrap kit and NucleoTraP CR kit (Macherey–Nagel) are used for purification of DNA from agarose gel and reaction mixtures, respectively, following the instructions of the manufacturer.
8. 100 μL aliquots of electrocompetent DH5αF′ bacteria are used for transformation; each aliquot is used for transformation of 1.5 μL of purified DNA, using 1 mm gap and 100 μL volume cuvette (Eppendorf) and an Electroporator 2510 (Eppendorf), set at 1800 V.

2.2. Cre Recombination

1. Bacterial strain used is *Escherichia coli* BS1365: BS591 F′ kan (BS591: recA1 endA1 gyrA96 thi-1 D lacU169 supE44 hsdR17 [lambda imm434 nin5 X1-cre]).

2.3. Phage Production and Titration

1. Solution for precipitation of phages: 20% PEG-8000 and 2.5 M NaCl (purchased from Fluka). The solution is filtered through a 0.22 μm filter before use.
2. PBS: 8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4 and 0.24 g KH_2PO_4 in 1000 mL H_2O, final pH 7.4.

2.4. Immunoprecipitation with Magnetic Beads

1. M450 Magnetic beads (Dynabeads) coupled to protein G, 4.5 μm size and small magnets designed for fitting of 1.5–2 mL tubes are purchased from Invitrogen.
2. Acetate buffer: dilute acetic acid to a final concentration of 0.5 M, and adjusting the pH to 5.
3. Serum samples from healthy donors and patients, suffering from infectious or autoimmune diseases, are stored in small aliquots at –70°C. When needed, a small aliquot is thawed and 5 μL are diluted into 95 μL of PBS.
4. MPBS and PBST: prepare from fresh PBS, by adding 4% non-fat powder milk or 0.1% Tween 20 (Sigma), respectively.

2.5. Subcloning Antigenic Fragments in Bacterial Expression Vector

1. pGEX vector plasmid DNA, engineered with unique BssHII and NheI restriction sites in the polylinker, is prepared using the Macherey–Nagel Nucleospin Plasmid kit, following the instructions of the manufacturer.
2. Restriction endonucleases, T4 ligase and buffers are purchased from New England Biolabs. All cloning steps are performed according to the manufacturer's suggestions and to standard molecular biology procedures.

3. 3% Agarose gel for DNA electrophoresis: dissolve 3 g of agarose (Sigma) in 100 mL TAE buffer (0.04 M Tris-acetate, 0.001 M EDTA). 10 μL of ethidium bromide (1 mg/ml, Sigma) are added.

2.6. High-Throughput Expression and Purification of Recombinant Polypeptides

1. Bacterial strain used is *Escherichia coli* BL21-Codon-Plus(DE3)-RIPL strain (Stratagene), *E.coli* B F– *ompT hsdS*(rB– mB–) *dcm*+ Tetr *gal*λ(DE3) *end A*Hte [*argU proL*-Camr] [*argU ileY leuW* Strep/Specr].
2. Standard stock solutions for auto-inducing medium are prepared according to (12). 100 mL of auto-inducing medium ZYM5052 are freshly prepared by adding each sterile stock solution in the following order: 80 mL of ZY; 2 mL of 50 × M; 200 μL of 1 M $MgSO_4$; 100 μL of metal mix; 2 mL of 50 × 5052; 100 μL of ampicillin. 15.6 mL of sterile water are added to adjust to the final volume of 100 mL.
3. Round bottom 2 mL 96-deep-well plates are used for bacterial cultures (Starlab).
4. Standard stock solutions for bacterial lysis are: lysis buffer (FAST Break, Promega); DNAseI (Sigma), at 5 mg/mL in TE buffer (10 mM Tris-HCl 1 mM EDTA pH 7.4); lysozyme (Sigma) at 20 mg/mL in TE buffer. Enzymes are stored at –20°C until use.
5. Wash/binding solution : 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% (v/v) Triton X-100 (Sigma).
6. Elution buffer is composed of 50 mM Tris-HCl pH 8.1, 50 mM GSH reduced (Sigma) pH 8, 100 mM NaCl .
7. GSH magnetic beads (MagneGST, Promega) are stored in 20% ethanol. Before use, beads are washed three times with 2.5 volumes of wash/binding solution using a magnetic stand for 1.5 mL tubes. Before removing supernatant it is necessary to allow particles to be captured by the magnet. Particles are then resuspended in 1 volume of wash/binding buffer.
8. 24 Rods magnetic stand and 96 tube aspiration manifold are purchased from V&P Scientific and round bottom 1 mL 96-deep-well plates from Costar, Corning.

2.7. SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

1. Upper buffer: 0.5 M Tris-HCl, 0.4% (v/v) SDS (from 10% (w/v) stock solution) and pH adjusted to 6.8.
2. Lower buffer: 1.5 M Tris-HCl, 0.4% (v/v) SDS and pH adjusted to 8.8. Upper and lower buffer can be stored at room temperature.
3. Acrylamide/bis-acrylamide (30% (w/v) acrylamide, 0.8% (w/v) bis-acrylamide) and ammonium persulphate (10% (w/v) in water) are stored at 4°C. Acrylamide is neurotoxic when unpolymerized and must be handled with care. *N,N,N,N′*-Tetramethyl-ethylenediamine (TEMED) is purchased from Sigma.

4. Running buffer: 25 mM Tris, 192 mM glycine, 0.1% (w/v) SDS and stored at room temperature.
5. 2x Sample buffer:125 mM Tris-HCl pH 6.8, 20% (v/v) glycerol, 4% (v/v) SDS (from 10% (w/v) stock solution), 10% (v/v) 2-mercaptoethanol, 0.04% (w/v) bromophenol blue and is stored at –20°C in 1 mL aliquots.
6. Coomassie staining solution: (0.1% (w/v) Coomassie Blue R-250, 45% (v/v) methanol, 10% (v/v) acetic acid.
7. Destaining solution: (10% (v/v) acetic acid, 10% (v/v) methanol); store at room temperature. Solutions are prepared and used under extractor fan due to methanol toxicity.
8. Prestained molecular weight markers are purchased from Fermentas and stored at −20°C.
9. BSA Calibrator stock solution (10 mg/mL) is freshly prepared by dissolving 10 mg of BSA (Sigma) in 1 mL of PBS.

3. Methods

3.1. Construction of ORF-Enriched cDNA Libraries

The description of the source and preparation of ORF-enriched libraries goes beyond the aim of this chapter and will not be discussed in detail (*see* **Note 1**). In the example described here, the starting material is human fibroblasts. The cDNA library will be prepared using random primer, the fragment length is controlled by mRNA boiling time and random primer concentration and the size of cDNA fragments ranges from 100 to 700 bp with an average of 350 bp (*see* **Note 2**); BssHII or NheI adapters have been added at 5′ or 3′ end, respectively (*see* **Note 3**). Briefly, a cDNA library is built as follows (*see* **Note 4**):

1. 5 μg of pPAO2 purified phagemid vector and adapters-ligated purified cDNA are sequentially digested, first with BssHII restriction endonuclease for 3 h at 50°C, then with NheI for 2 h at 37°C.
2. The digested phagemid and inserts are purified, resuspended in milliQ water, and ligated with T4 DNA ligase (phagemid:insert molar ratio of 1:5). After O/N ligation at 16°C, the DNA is purified by precipitation and resuspended in 15 μL of milliQ water.
3. The purified DNA is transformed by electroporation into DH5αF′ bacteria (*see* **Note 5**); after transformation, bacteria are added with 1 mL of liquid 2xTY medium, without any antibiotic pressure, and allowed to grow at 28°C, shaking at 220 rpm, for 60 min. Bacteria are then plated on 15 cm 2xTY agar plates (1 mL of bacterial suspension per plate) added

with chloramphenicol (pPAO2 resistance, 34 μg/mL) and ampicillin (selective marker for ORFs, 15 μg/mL) and allowed to grow for 24 h at 28°C. Dilutions are also plated on either chloramphenicol/ampicillin or chloramphenicol only 2xTY agar plates, for titration of the library size.

4. The size of the library is determined from the dilution plated on ampicillin/chloramphenicol (only ORFs will grow) or chloramphenicol only (all clones) (*see* **Note 6**). The general size of our libraries is above 10^6 ORF transformants. Bacteria are collected, thoroughly mixed, added with 20% sterile glycerol and kept at −80°C in small aliquots. One aliquot is immediately used for purification of plasmid DNA in order to undergo the recombination step.

3.2. Cre Recombination for Removal of the β-Lactamase Gene

In order to allow the phage display of ORFs, these must be fused to the sequence lying downstream the β-lactamase gene and encoding the phage g3p protein. To this end, the β-lactamase gene is removed by Cre-mediated recombination. The DNA of ORF-filtered constructs is transformed into a bacterial strain carrying the Cre-recombinase gene (BS1365). Cre-recombinase recognizes the Lox sites flanking the β-lactamase gene, excides the gene and yields a standard phagemid construct suitable for phage display. However, since BS1365 bacteria are not the best hosts for the production of high titre of phages, recombined phagemids are further transformed into DH5α strain. The detailed protocol is as follows:

1. A small aliquot of the plasmid preparation of the original library is electroporated into BS1365 bacteria, and plated onto 2xTY agar plates added with chloramphenicol. The bacteria are allowed to grow O/N at 30°C, for complete removal of the β-lactamase gene from the library (*see* **Note 7**).
2. Bacteria are collected, plasmid DNA is extracted and a small aliquot is electroporated into DH5α. Bacteria are plated onto 2xTY agar plates added with chloramphenicol and allowed to grow O/N at 30°C.
3. Bacteria are collected, added with 20% sterile glycerol, and split into several aliquots to be kept at −80°C until further use.

The library is then inspected for: (i) average size of the cDNA inserts; (ii) complete removal of the β-lactamase gene; and (iii) effective enrichment for ORFs. Points (i) and (ii) may be easily carried out by PCR on randomly picked colonies, before and after recombination, using primers external to either the polylinker (for the average size of fragments) or to the polylinker and the β-lactamase gene (to assess successful recombination). Point (iii) may be carried out by analysing the presence of the correct frame in the gene sequences obtained from a suitable number of randomly picked clones.

A summary of the procedure is reported in **Fig. 20.1**.

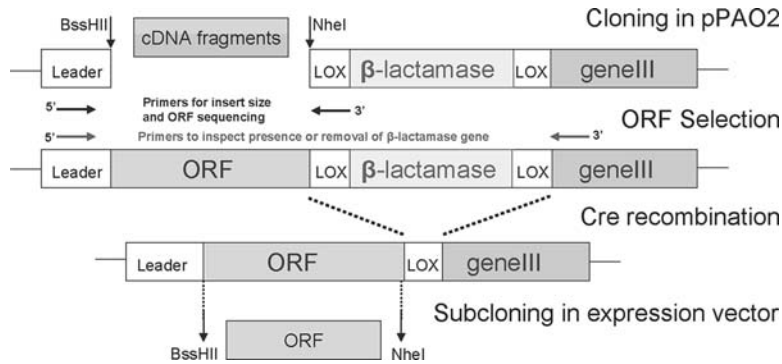


Fig. 20.1. Schematic summary of the procedure. cDNA fragments are cloned into the pPAO2 vector, allowing selection of ORFs. A recombination step removes the ß-lactamase gene, yielding a construct suitable for phage display. After selection with antibodies, ORFs of interest are subcloned into an expression vector for production of recombinant proteins and further characterization.

3.3. Phage Production and Titration

1. A bacterial stock aliquot of the library, stored at –80°C, is diluted in 10 mL of 2xTY liquid broth added with chloramphenicol (*see* **Note 8**) and grown in a 100 mL flask shaking at 220 rpm and at 37° C, until it reaches OD_{600} 0.5. Under these conditions, bacteria express the *pilus* necessary for phage infection (*see* **Note 9**).
2. Bacteria are infected with a wild-type helper phage (M13K.07 carrying all the genes needed for phage replication) at a MOI (multiplicity of infection, the ratio viral particles/bacterial cells) of 20 to 1. This step is performed at 37°C for 45 min, without shaking.
3. After infection, bacteria are centrifuged at 1500*g* for 10 min at 30° C, the supernatant is discarded and the pellet is resuspended in 40 mL of 2xTY broth added with chloramphenicol and kanamycin. Bacteria are then grown O/N, at 30°C and shaking at 220 rpm, to allow production of the recombinant phages in the culture supernatant.
4. The following day bacteria are centrifuged at 7000 rpm for 20 min at 4°C. The supernatant, containing phages, is collected.
5. Precipitation of phages is performed by adding 10 mL (one-fifth of total volume) of the PEG–NaCl solution to the phage-containing supernatant, followed by incubation on ice for 45 min, to allow precipitation (*see* **Note 10**). The solution is then centrifuged at 7000 rpm for 20 min at 4°C, to collect the phages. The supernatant is carefully and entirely discarded and the white pellet is resuspended in 1 mL of PBS. Since the phages are soluble in PBS, the solution is centrifuged for 5 min at 10,000 rpm, to pellet the remaining bacteria in

solution; the phage-containing supernatant is transferred into a fresh tube and kept in ice until needed (*see* **Note 11**). Titre of phages is around $10^{10}/\mu L$.

3.4. Phage Selection by Magnetic Beads Coated with Serum Antibodies

Two different preparations of magnetic beads are needed: the first (called here NEGATIVE) functionalized with serum antibodies from healthy donors, the second (called POSITIVE) with antibodies from sera of patients with immune–autoimmune disorders.

In the first step, in order to subtract the library from clones binding aspecifically to antibodies that are present in the normal repertoire of the immune system of healthy people, a selection with the negative antibody preparation is required. The positive preparation will then be used for the selection of the antigenic–autoantigenic clones.

The selection procedure is performed as follows (*see* **Note 12**):

1. 0.2 mL of phage preparation (corresponding to 10^{11}–10^{12} phages) is saturated by adding an equal volume of 4% MPBS and incubating for 1 h at RT (*see* **Note 13**).
2. *E. coli* DH5α are inoculated in 2 mL of 2xTY liquid broth and grown at 37°C in shaking platform. The culture must be checked carefully during growth, in order to reach an OD_{600} between 0.3 and 0.8 – best at 0.5 – when the elution step described later (step 8) is performed (*see* **Note 14**).
3. 10 μL of protein G-coated M450 dynabeads are transferred to a 1.5 mL Eppendorf tube. Two washes are performed with acetate buffer, pH 5. All the washing steps described in this protocol are performed in the same identical manner, independently from the buffer used for washing, as follows:
   - the tube is put on the magnet and beads allowed to be collected on the side of the tube facing the magnet for 30 s;
   - with a pipette, all the solution in the tube is gently removed, without touching the beads;
   - the tube is removed from the magnet, 0.75 mL of the washing solution are added and the tube is gently shaked until the beads are resuspended again in the liquid phase;
   - if needed, points 1–3 are repeated.
4. 5 μL of serum (or pool of sera) are diluted into 100 μL of PBS, added to the washed beads, and incubated in slow rotation at RT.
5. After 45 min, the beads from the NEGATIVE preparation are washed three times with acetate buffer, pH 5, to remove unbound antibodies and other serum proteins.
6. For the subtraction step, 200 μL of MPBS-saturated phages (about 10^{12}phages, called INPUT) are added to the NEGATIVE bead preparation, and incubated for 30 min in rotation

(*see* **Note 15**). Magnetic field is applied, and the beads with bound phages are collected on the side of the tube. The supernatant, containing the unbound phage population, is collected and used in the next step.

7. The beads from the POSITIVE preparation are washed three times with acetate buffer, pH 5. The phages obtained from the previous step are then incubated with the beads for 90 min in rotation.
8. Several washes, with PBST and PBS, are then performed to remove unbound phages (*see* **Note 16**).
9. Bound phages, representing the OUTPUT of the selection round, are then eluted by mixing the beads with 1 mL of *E. coli* DH5α, at OD_{600} 0.5, at 37°C, for 45 min, with occasional shaking.
10. Bacteria are plated on 15 cm 2xTY agar plates added with chloramphenicol and grown O/N at 30°C. The following day, the output is collected from the plates by adding 1 mL of fresh 2xTY broth and harvesting with a sterile scraper; most of the bacterial suspension is then added with 20% glycerol and stored at –70°C in 50 μL aliquots. An aliquot is immediately grown again (see the description of the phage production in previous paragraph) to undergo the next selection round.

3.5. Subcloning of cDNA Fragments in Bacterial Expression Vector

Although direct printing of lysates (13) or phages (14) onto protein arrays have been reported, we suggest to use purified recombinant proteins. The purification step is required since sera and biological fluids may contain immunoglobulins against bacterial antigens. We use a modified pGEX vector (GE Healthcare) that allows the insertion of the BssHII-NheI ORF fragments downstream the coding sequence of the GST gene. The cloning steps are as follows:

1. Plasmid preparation of the selected library is obtained from an aliquot of bacteria.
2. 5 μg of DNA are digested in a small volume using the restriction sites flanking the cDNA gene fragments (BssHII-NheI).
3. The digested sample is loaded on a 3% agarose gel in TAE buffer, and after short electrophoresis separation, the smear in the 100–700 bp range is cut out of the gel (*see* **Note 17**).
4. The fragments are purified and ligated, by using T4 ligase, into the expression vector cut with the same enzymes. The ligation reaction (vector:insert molar ratio of 1:5) is performed O/N at 16°C. The following day, the ligation is purified and transformed into BL21(DE3)RIPL *E. coli* strain. The resulting mini-library has a size around 10^5 clones.

3.6. High-Throughput Expression and Purification of Recombinant Polypeptides

Increased yields of protein production are a requirement for many applications, including protein microarrays. The high-throughput expression and purification of hundreds of different clones is here performed by coupling high-density growth with magnetic-beads affinity purification in a 96-well format.

By using the auto-induction system described in (12), 96 single bacterial cultures can be grown on deep-well plates, reaching high optical densities, and without the need for monitoring the bacterial growth to induce expression of the recombinant proteins. Growth is performed on a shaking platform at low temperature (28°C) to provide correct folding of polypeptides.

The bacterial strain used (BL21(DE3)RIPL) allows high-yield production of most polypeptides, including those with genes carrying rare codons for arginine, isoleucine, leucine and proline.

Production and purification of high number of proteins may be difficult to manage and time-consuming. The system described here provides easy steps for lysis and affinity purification (binding, washing and elution) that can be simultaneously carried out on 96 different cultures. Magnetic beads used in this protocol are functionalized with GSH to allow affinity purification of GST-recombinant proteins. The procedure is performed as follows:

1. A 96-well plate is prepared, each well with 100 µL of LB medium supplemented with 100 µg/mL ampicillin. Single colonies are picked from the Petri dishes, where the library has been plated, with sterile tips, and are inoculated in each well. The cultures are grown for 6 h at 37°C in a shaking platform, at 200 rpm.
2. 900 µL of auto-inducing medium are dispensed in each well of the 96-deep-well plate.
3. 5 µL of the growing cultures are transferred into the deep-well plates and grown overnight at 28°C, in a shaking platform at 200 rpm (*see* **Note 18**).
4. After O/N growth, 100 µL of FastBreak lysis buffer are added into each well; the plate is then sealed with adhesive aluminium film and incubated for additional 15 min with gentle shaking at 4°C.
5. 10 µL of DNAse are added in each well and the plate is incubated for 15 min with gentle shaking at 4°C. DNAse degrades genomic DNA released during lysis and reduces the viscosity of solution.
6. The lysozyme stock is diluted 1:10 in water just before use and 10 µL are added in each well. The plate is then incubated for 20 min with gentle shaking at 4°C. Clarification should be observed (*see* **Note 19**).

7. GSH magnetic beads are prepared as described in **Section2**; the bacterial lysates are transferred to a 1 mL 96-deep-well plate (Costar) and 25 μL of the magnetic beads suspension is added to each well. Aluminium sealing film is used to seal the plate.
8. The plate is incubated with gentle mixing on a rotating platform at 4°C for 30 min; particles are not allowed to settle for more than a few minutes, as this reduces binding efficiency.
9. The plate is set on the magnetic support, so that the GSH particles are captured by the magnets. A 96 tube aspiration manifold connected to the vacuum pump is then used to carefully remove supernatant. The manifold is then removed, the plate taken off from the magnetic stand, and 250 μL of binding/wash buffer are added to each well with a multichannel pipette. The plate is gently mixed for 5 min at 4°C. The plate is placed back on the magnetic stand and washing procedure is repeated for three times.
10. Elution of the bound protein is performed by adding 25 μL of elution buffer in each well (*see* **Note 20**); the plate is incubated with gentle mixing at 4°C for 15 min.
11. The plate is placed back on the magnetic stand, the supernatant is carefully removed with a multichannel pipette, and the eluted GST-fusion proteins are transferred into a new 96-well plate. A second elution step may be performed following the same procedure. The two aliquots of the eluted protein are pooled together and kept either at 4°C or, for longer storage, at –70°C (in this case, 10% glycerol is added).

3.7. Analysis of Recombinant Proteins by SDS-PAGE and Coomassie Staining

1. To check the amount and quality of recombinant proteins, SDS-PAGE analysis is performed. A mini-gel system (i.e. Bio-Rad Mini Protean III) is highly recommended due to the savings in material and time and also because it provides high-resolution protein separation. To resolve proteins with a MW ranging from 25 to 60 kDa, a 12.5% and 0.75 mm thick gel is recommended. Gel preparation and electrophoretic run are performed according to (15). Running gel solution, is prepared by mixing in the following order: 1.65 mL H_2O, 1.25 mL lower buffer, 2.1 mL acrylamide/bis-acrylamide, 25 μL ammonium persulphate, 7.5 μL TEMED. Stacking gel solution is prepared mixing reagents in the following order: 1.5 mL H_2O, 0.625 mL upper buffer, 0.35 mL acrylamide/bis-acrylamide, 15 μL ammonium persulphate, 5 μL TEMED. Assembly of gel sandwich into electrophoretic apparatus is performed according to manufacturer instructions.
2. Samples are prepared adding 2x sample buffer and purified protein in equal amounts in an Eppendorf tube. For the BSA calibrator sample, 1 μL of stock solution is diluted in 100 μL

final volume of PBS, then 100 µL of 2x sample buffer are added (final concentration: 50 ng/µL). All samples are heated at 100°C for 10 min, and then centrifuged for 1 min.

3. 5 µL of all samples are loaded into wells, avoiding cross-contamination of adjacent wells. Molecular weight standard and three dilutions of BSA Calibrator (i.e. 100 ng/µL, 300 ng/µL, 500 ng/µL, pipetting 2, 6 and 10 µL, respectively) are also loaded. The run is performed at constant voltage (100 V).
4. When prestained molecular weight marker bands (in particular those ranging from 25 to 70 kDa) are well resolved (about 1 h), the run is stopped.
5. When the run is over, the gel is removed from the assembly, transferred into a small box containing 10 mL of Coomassie staining solution, covered with a lid and incubated on a shaking platform for at least 30 min. The Coomassie solution is then discarded and replaced with 10 mL of destaining solution for 2 h (a piece of styrofoam can be added in order to completely absorb Coomassie stain that diffuse from the gel).
6. The height and the number of bands observed on the gel for each protein give an indication of MW, purity and degradation of the samples (*see* **Note 21**). Their concentration can be evaluated by comparing the intensity of bands with those of the calibrator, considering that the volume of purified protein loaded is 2.5 µL (total volume of sample loaded, including sample buffer, was 5 µL). If available, it is possible to use digital instrumentation (i.e. VersaDoc, Bio-Rad) to acquire the image of the gel and to perform digital densitometry of bands. **Figure 20.2** shows a Coomassie-stained SDS-PAGE of proteins produced and purified with the method described in this chapter.

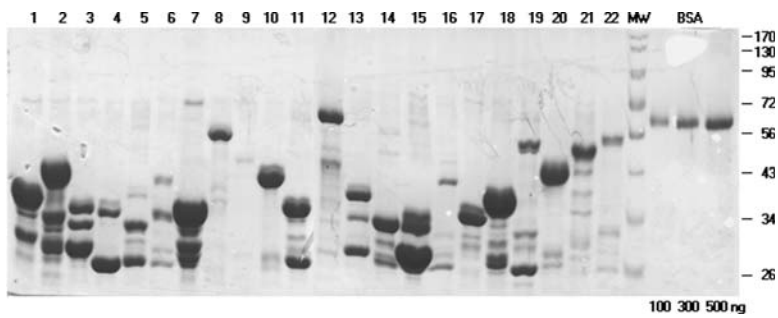


Fig. 20.2. SDS-PAGE followed by Coomassie staining of 22 different random GST-tagged proteins. Proteins are produced by bacterial cultures of BL21(DE3) strain grown in auto-inducing medium and purified with GSHmagnetic beads. On the right side, BSA calibrator: bands intensity correspond to 100, 300 and 500 ng of total protein loaded.

4. Notes

1. The source of genetic material, its collection, treatment and protocols should be carefully planned in accordance with the aims of the specific study. Any source of cDNA, genomic DNA, or other genetic material (i.e. synthetic) is potentially suitable for building of libraries which could then be handled with the methods described in this chapter.
2. Large fragments are likely to be more difficultly displayed on the phage surface whereas fragments up to 300 amino acids (i.e. scFvs) are known to be correctly displayed on phage surface. Considering that the average size of functional domains of proteins is assessed to be around 100–150 aa, we decided to select an average fragment size of the cDNA library of 350 bp.
3. The presence of linkers/adapters for sticky-end cloning is not strictly required: the pPAO2 vector polylinker carries a StuI blunt restriction site, allowing direct cloning of blunt-ended cDNA fragments. Although blunt-end cloning is a simpler strategy, the efficiency of ligation of the fragments with the vector is lower, resulting in libraries with a smaller size.
4. For the construction of large libraries, it is extremely important to set-up good digestion, ligation, and purification protocols. The digestion step requires that the DNA is entirely cut, there is no or little degradation, and that most of the vector can be re-ligated after cutting. The ligation should be done using a high-concentration T4 ligase enzyme. The purification should be carefully performed, especially in the final step before electroporation, where high concentration of DNA is required without presence of any salts in solution, since the presence of contaminants leads to a dramatic decrease in the electroporation transformation efficiency. Small decrease in the performance of protocols at any of these steps will easily lead to the production of libraries 10- to 100-fold smaller than expected.
5. Electrocompetent cells can be produced in-house or purchased from several manufacturers. The use of high-efficiency cells (above 10^9 transformants per μg of DNA) is required.
6. The theoretical expected ratio of transformants on chloramphenicol, compared to chloramphenicol/ampicillin, can be approximately set to 18. The ratio we observed in all experiments is slightly above 20.

7. A number of transformants at least 10-fold higher than the size of the original library is suggested when proceeding with the recombination step so that the diversity of the original library is fully maintained.
8. Ideal starting OD_{600} is 0.05, corresponding to around 5×10^7 bacteria/mL. This corresponds to a $100\times$ representation of the diversity of the library for a 10 mL culture.
9. The *pilus* expression is highest when the bacteria are in a log-phase growth, at 37°C. It is important to check carefully the temperature, since any drop in it would immediately determine lost of the *pilus*. It is, therefore, strongly suggested to prepare in advance all reagents before removing the bacteria from the shaker, and to perform all steps quickly, without allowing the temperature to decrease. At the time of infection, bacteria should be in log-phase growth, with an OD_{600} around 0.5; this means that bacteria grown to saturation (i.e. OD_{600} 2.0) cannot be diluted to OD_{600} 0.5 and then used for infection.
10. By gently inverting the tube, a smoky consistence should be observed a few seconds after adding the precipitation solution. This condition, representing phages clumping together, should increase during the incubation in ice.
11. For selection steps, it is, therefore, strongly suggested to use fresh preparation phages, and to store them at 4°C for no longer than 1 day.
12. The preparation of all reagents may be time-consuming. It is, therefore, needed to start at the same time with several different preparations, since the time needed for them will be different and their use will be required in different steps during the protocol (i.e. antibody functionalization, phage saturation, bacterial growth, etc.).
13. This blocking step is suggested to reduce background binding and should be performed in all selection rounds.
14. It is important that bacterial culture has the correct OD_{600} for the elution step at the end of the selection. It is, therefore, necessary to start bacterial growth early enough so that at the end of the experiment (about 3 h) bacteria are in log-phase growth with an OD_{600} around 0.5. It is suggested to grow bacteria in several tubes (only 1 mL is required for a single elution) inoculating different starting amounts of bacteria; choose the tube with the OD_{600} closest to 0.5 for the final elution step.
15. Incubation with slow and constant rotation is required, since the beads quickly form deposits on the bottom of the tube.
16. According to most experimental protocols in the phage-display techniques, the stringency of the washing steps should increase with the selection rounds. We use the following

procedure: beads are washed 10 times with PBST and 10 times with PBS, to remove unbound phages, for the first round of selection. Fifteen PBST washes and 15 PBS washes are performed for the second round; for the third round, 15 washes with PBST were followed by 10 min washing in rotation, followed by 15 more PBS washes.

17. It is suggested to perform a very short electrophoresis run (but enough to distinguish clearly the smear of the excised fragments from the vector backbone), so that the fragments ranging from 100 to 700 bp can be collected by cutting a small slice out of the gel.
18. After overnight growth, OD_{600} of some cultures is measured to check for growth rate. Final OD_{600} of cultures should range from 5 to 10 units. To ensure a correct reading in the linear range, cultures may be diluted in broth to obtain readings below 0.25 units.
19. If lysates are not clarified, incubation time may be increased, and the solution supplemented with DNAse and lysozyme (i.e. doubling the volumes of enzymes added, especially if OD_{600} of bacterial cultures is higher than 8).
20. The elution buffer must be freshly prepared and checking of final pH is required; GSH must be reduced to allow elution of GST-fusion proteins from GSH beads.
21. GST fusion polypeptides have an expected MW ranging from 27 kDa (GST alone) to 70 kDa: bands with higher molecular weight indicate impurity in the sample (mainly due to BL21(DE3)RIPL proteins co-purified with the GST recombinant protein). To avoid co-purification of contaminants, it is suggested to increase number and stringency of washings during purification with magnetic beads. If a sample shows two or more bands, degradation might have occurred during lysis and purification: a cocktail of protease inhibitors may be added (1 μM leupeptin, 0.1 mM PMSF, 1 mM EDTA; due to their toxicity, handle carefully) with lysozyme during bacterial culture lysis; the whole purification process should also be performed maintaining the plates at 4°C.

Acknowledgements

This work was supported in part by Fondazione Cariplo, Compagnia Sanpaolo, NIH RFA-DK-06-002, and Regione Piemonte Ricerca sanitaria Finalizzata 2006 and 2007.

References

1. Wang, X., et al., (2005) Autoantibody signatures in prostate cancer. N Engl J Med, **353**(12): 1224–35.
2. Chatterjee, M., et al., (2006) Diagnostic markers of ovarian cancer by high-throughput antigen cloning and detection on arrays. Cancer Res, **66**(2): 1181–90.
3. Kalnina, Z., et al., (2008) Evaluation of T7 and lambda phage display systems for survey of autoantibody profiles in cancer patients. J Immunol Methods, **334**(1–2): 37–50.
4. Somers, V., et al., (2008) Autoantibody profiling in multiple sclerosis reveals novel antigenic candidates. J Immunol, **180**(6): 3957–63.
5. Parmley, S.F. and G.P. Smith, (1988) Antibody-selectable filamentous fd phage vectors: affinity purification of target genes. Gene, **73**(2): 305–18.
6. Paschke, M., (2006) Phage display systems and their applications. Appl Microbiol Biotechnol, **70**(1): 2–11.
7. Kehoe, J.W. and B.K. Kay, (2005) Filamentous phage display in the new millennium. Chem Rev, **105**(11): 4056–72.
8. Fuh, G. and S.S. Sidhu, (2000) Efficient phage display of polypeptides fused to the carboxy-terminus of the M13 gene-3 minor coat protein. FEBS Lett, **480**(2–3): 231–4.
9. Fuh, G., et al., (2000) Analysis of PDZ domain–ligand interactions using carboxyl-terminal phage display. J Biol Chem, **275**(28): 21486–91.
10. Weiss, G.A. and S.S. Sidhu, (2000) Design and evolution of artificial M13 coat proteins. J Mol Biol, **300**(1): 213–9.
11. Jespers, L.S., et al., (1995) Surface expression and ligand-based selection of cDNAs fused to filamentous phage gene VI. Biotechnology (NY), **13**(4): 378–82.
12. Crameri, R., et al., (1994) Display of expression products of cDNA libraries on phage surfaces. A versatile screening system for selective isolation of genes by specific gene-product/ligand interaction. Eur J Biochem, **226**(1): 53–8.
13. Crameri, R. and K. Blaser, (1996) Cloning *Aspergillus fumigatus* allergens by the pJuFo filamentous phage display system. Int Arch Allergy Immunol, **110**(1): 41–5.
14. Di Niro, R., et al., (2005) Characterizing monoclonal antibody epitopes by filtered gene fragment phage display. Biochem J, **388**(3): 889–94.
15. Zacchi, P., et al., (2003) Selecting open reading frames from DNA. Genome Res, **13**(5): 980–90.

Section IV

Software/Web Tools and Data Analysis

Chapter 21

Visualisation and Pre-processing of Peptide Microarray Data

Marie Reilly and Davide Valentini

Abstract

The data files produced by digitising peptide microarray images contain detailed information on the location, feature, response parameters and quality of each spot on each array. In this chapter, we will describe how such peptide microarray data can be read into the R statistical package and pre-processed in preparation for subsequent comparative or predictive analysis. We illustrate how the information in the data can be visualised using images and graphical displays that highlight the main features, enabling the quality of the data to be assessed and invalid data points to be identified and excluded. The log-ratio of the foreground to background signal is used as a response index. Negative control responses serve as a reference against which "detectable" responses can be defined, and slides incubated with only buffer and secondary antibody help identify false-positive responses from peptides. For peptides that have a detectable response on at least one subarray, and no false-positive response, we use linear mixed models to remove artefacts due to the arrays and their architecture. The resulting normalized responses provide the input data for further analysis.

Key words: Data quality, response index, graphical display, detectable response, false positive, normalization, linear mixed models.

1. Introduction

After peptide arrays are scanned, the images are digitised by specialised software and the resulting data stored in data files, one for each array or subarray. While these files can be recognised and opened with a spreadsheet program such as Microsoft Excel, they also contain special features that facilitate manipulation with image-processing software. The R statistical computing package (1) provides a large collection of powerful functions for examining these data. In this chapter, we will describe how such peptide microarray data can be read into R, visualised using images and graphical displays, and normalized to remove artefacts in preparation for analysis.

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_21,

1.1. Data Example

We will illustrate our methods using peptide microarray data from an application to tuberculosis research. The experiment involved the comparison of a group of 34 TB patients with a group of 35 healthy people(2). A specimen from each of these individuals was prepared on a microarray slide consisting of two identical subarrays, and the digitised image from each subarray was saved as a data file. In addition, 13 slides were prepared using only buffer and secondary antibody (i.e. no patient sample); seven slides from the batch used for the TB patients and six slides from the batch used for the healthy individuals. Thus the data consist of four groups of arrays, and a total of 164 data files.

1.2. Slide Structure

The slides used in our illustrative examples are "second-generation" slides manufactured by JPT, Germany (JPT, 2008). These consist of two identical subarrays, each consisting of 24 blocks of spots arranged in an 18×18 matrix, for a total number of 7776 spots (*see* **Fig. 21.1**). Among these, 7503 are peptide spots, with most peptides (7466) appearing only once on the subarray. The 273 control spots consist of four spots each for IgA, IgE, IgG and IgM, 104 negative controls (i.e. empty spots) and 153 other miscellaneous controls. The methods we describe in this chapter apply with minimum modification to slides with more than two subarrays, or different block size and/or arrangement.

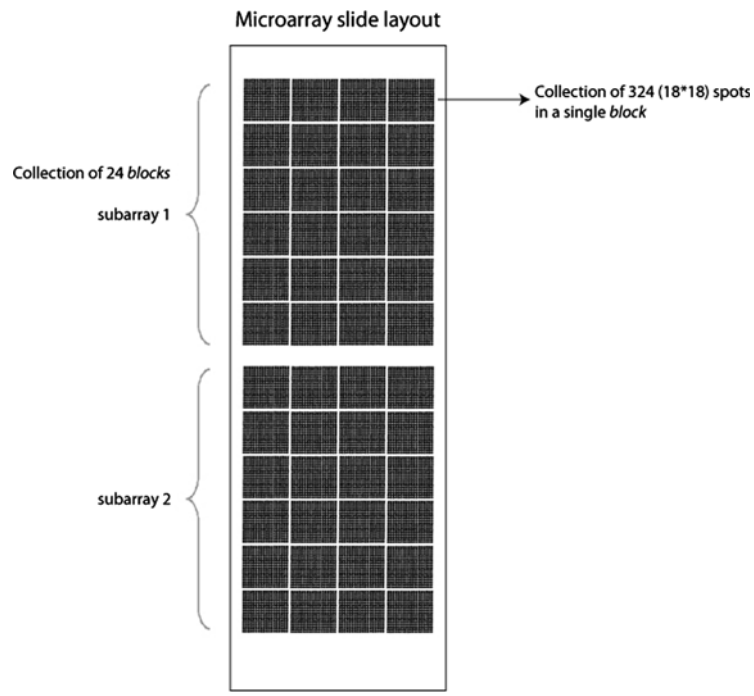


Fig. 21.1. Layout of a second-generation peptide microarray slide.

1.3. Digitised Peptide Microarray Data

The peptide microarray images are analysed using specialised software packages such as GenePix (3) or Agilent's Feature Extraction (4) that save the digitised images in formatted data files. In our applications, we used the GenePix package which produces data files in .GPR (GenePix Results) format. These are essentially readable text files, with a header of some lines (usually 31) used to store various internal information (for more details, see the Molecular Devices GenePix Web site (2)). This header is followed by the data spreadsheet with each spot on the array occupying a row and the characteristics of the spots on the columns.

Figure 21.2 shows an excerpt from a GPR file, showing the most important variables. The first three columns in the spreadsheet identify the location of the spot on the slide by specifying the

Microsoft Excel - 786-41.s2.gpr

File Edit View Insert Format Tools Data Window Help

A32 Block

	A	B	C	E	I	J	K	L	N	O	P	Q	AX
32	Block	Column	Row	ID	F635 Median	F635 Mean	F635 SD	F635 CV	B635 Median	B635 Mean	B635 SD	B635 CV	Flags
33	1	1	1	IgG-human	31	32	7	21	31	33	7	21	-50
34	1	2	1	LPPAANDVSVLTAAR	32	32	7	21	32	33	8	24	-50
35	1	3	1	HTKHYRVVSKPAALV	33	34	7	20	32	33	8	24	-50
36	1	4	1	FVALPAATADAYATT	34	34	7	20	33	33	8	24	-50
37	1	5	1	NYPAMMAHAGDMAGY	34	35	8	22	33	34	7	20	-50
38	1	6	1	RRALRQIGVLERPVG	34	34	7	20	33	35	8	22	-50
39	1	7	1	DCGTIRVGSFRGRWL	151	153	40	26	35	35	8	22	0
40	1	8	1	MAKLSTDELLDAFKE	35	35	8	22	35	36	9	25	-50
41	1	9	1	LELSDFVKKFEETFE	34	35	8	22	35	35	9	25	-50
42	1	10	1	VSRRAKVDVLIVHTT	35	36	8	22	35	36	11	30	-50
43	1	11	1	WGQLLLLGEGIPDPG	34	35	8	22	35	36	9	25	-50
44	1	12	1	NTGSSLFKQISDKMG	34	37	28	75	35	36	11	30	-50
45	1	13	1	PGTNWIGQAAEAYLN	36	36	8	22	35	36	8	22	-50
46	1	14	1	WTELAALPDFLGGFA	35	36	9	25	35	36	8	22	-50
47	1	15	1	SLGFGNLLSFASLPT	35	36	8	22	35	35	8	22	-50
48	1	16	1	VTATMGQLQQLVAAG	35	36	8	22	35	36	10	27	-50
49	1	17	1	SQLASMGSQQAQLIS	35	36	8	22	35	36	10	27	-50
50	1	18	1	IgG-human	35	36	8	22	35	36	9	25	-50
51	1	1	2	LTNVNSAAGLHAPMQ	32	33	7	21	32	33	9	27	-50
52	1	2	2	QEAGNFERISGDLKT	32	33	7	21	33	34	7	20	-50
53	1	3	2	QVESTAGSLQGQWRG	33	34	9	26	32	33	7	21	-50
54	1	4	2	TAAQAAVVRFQEAAN	32	33	7	21	33	33	10	30	-50
55	1	5	2	RPGLPVEYLQVPSPS	34	35	7	20	33	34	10	29	-50
56	1	6	2	DIKVQFQSGGNNSPA	35	36	8	22	34	35	9	25	-50
57	1	7	2	LDGLRAQDDYNGWDI	34	36	11	30	34	35	8	22	-50
58	1	8	2	AFEWYYQSGLSIVMP	36	37	8	21	35	36	9	25	-50
59	1	9	2	GLAALAVAVSPPAAA	34	35	8	22	34	35	8	22	-50
60	1	10	2	VGPGCAEYAAANPTG	35	36	11	30	34	35	8	22	-50
61	1	11	2	VQGMSQDPVAVAASN	34	36	16	44	34	35	8	22	-50
62	1	12	2	LTTLTAALSGQLNPQ	34	35	8	22	34	35	8	22	-50
63	1	13	2	EYTVFAPTNAAFDKL	35	36	8	22	35	35	8	22	-50
64	1	14	2	TIDQLKTDAKLLSSI	35	36	10	27	35	36	8	22	-50
65	1	15	2	HVIAGQASPSRIDGT	35	36	9	25	35	36	8	22	-50
66	1	16	2	LQGADLTVIGARDDL	35	36	8	22	35	36	8	22	-50
67	1	17	2	INISLPSYYPDQKSL	35	35	8	22	35	36	8	22	-50
68	1	18	2	IAQTRDKFLSAATSS	36	37	8	21	35	35	9	25	-50
69	1	1	3	EAPYELNITSATYQS	31	33	7	21	32	33	9	27	-50
70	1	2	3	PRGTQAVVLKVYQNA	32	34	13	38	33	33	8	24	-50
71	1	3	3	TPGYAQGQQQTYSQQ	33	33	7	21	32	33	8	24	-50
72	1	4	3	RYPPSPPPQPTQYRQ	33	34	7	20	33	34	10	29	-50
73	1	5	3	ALGGTRPGLIPGVIP	34	36	16	44	33	34	10	29	-50
74	1	6	3	PPPGMVRQRPRAGML	2553	2547	484	19	34	34	7	20	0
75	1	7	3	VNMNAQLVGVNSAIA	34	36	11	30	34	35	8	22	-50
76	1	8	3	ADSADAQSGSIGLGF	35	35	7	20	35	36	9	25	-50
77	1	9	3	VDQAKRIADELISTG	34	34	8	23	35	36	14	38	-50
78	1	10	3	HASLGVQVTNDKDTL	34	35	8	22	34	35	8	22	-50
79	1	11	3	GNSGGQGGTPRAVPG	35	35	8	22	34	35	10	28	-50

Fig. 21.2. Excerpt from a .GPR data file opened in Excel. Note that the first 31 "header lines" are not displayed.

printing block, followed by the row and column number within the block. Following the peptide or control name in column 4, various measures of intensity of the signals are given: for example, "F635 Median" reports the median foreground signal for the red channel (wavelength 635 nm) while the standard deviation and coefficient of variation are denoted as "F635 SD" and "F635 CV", respectively. The same measures are stored for the background signal (beginning with "B635 Median") and the whole series is repeated for the green channel (wavelength 523 nm). In our TB example, the red channel was used, so that the data in the F532 and B532 variables are not of interest and have been hidden in **Fig. 21.2** (between columns Q and AX).

The "Flags" variable in the last column reports values assigned to each spot by the Genepix software using criteria for the quality/reliability: in this example,"0" indicates a good spot, "–50" a spot "not-found" (a spot with very low-intensity values, or with some defect related to other parameters, for example the shape), "–75" an "empty" spot (a spot on which nothing is printed) and "–100" a "bad" spot. The GenePix user can modify the criteria that are used to flag spots as "bad". This is an important step for all peptide microarray analysis, as the default criteria are designed for the two-colour arrays originally used in genomics and are not suitable for peptide arrays. This will require some validation experiments for any given equipment or arrays: in our case we explored various criteria for the magnitude and variability (CV) of the foreground and background signals and found the following criteria to be satisfactory in identifying bad spots:

Non-uniform Foreground: ([F635 Mean] > (1.5*[F635 Median])) and [F635 Median] > 40
Non-uniform Background: ([B635 Mean] > (1.5*[B635 Median])) and [B635 Median] > 40)

2. Preparing the Computing Environment

All the graphical and statistical analyses presented here were carried out using R, a language and environment for statistical computing and graphics, available for free download (1). Some of the packages and tools used are part of Bioconductor (5), an open source and open development software project for the analysis of various kinds of high-throughput data. Bioconductor deals mostly with "biomics" data, and includes many packages designed for microarray data analysis. Detailed information about all the packages can be found at (6).

In brief, to install the basic Bioconductor packages, it is necessary to first have R installed. If the R package is launched with an active internet connection, the following command typed on the R console will automatically download and install a collection of the most frequently used Bioconductor packages for microarray data analysis:

```
source("http://bioconductor.org/biocLite.R")
biocLite()
```

As we will see later, it is useful to put all files from a specific study into a single folder. In our example we could store all the GPR files from the 82 arrays (164 subarrays) in the folder "C:/TBstudy/data".

2.1. Reading GPR Data Files

The *marray* package (in Bioconductor) is designed for the exploration of microarray data mainly for quality control purposes. There are functions available in this package for reading various types of data files: for example *read.GenePix()*, *read.Agilent()*, etc. Detailed descriptions of all these functions are included in the *marray* help files and tutorial. We will describe in detail below how to read in the GPR files produced by GenePix. Users of other platforms will simply need to substitute this step with the functions appropriate for their data format, and once the data have been read into R, one can save (and work with) the R data structure, so that the source of the data is no longer relevant.

The most direct way to read GPR data files into R is to set the working directory to be the folder containing the files of interest, and then use the *read.GenePix* function from the *marray* Bioconductor package. For example if the GPR data files are in the folder named "C:/TBstudy/data", the following R commands will read them into a special data structure (called an S4 variable) with the name "mydata":

```
setwd("C:/TBstudy/data")
mydata <- read.GenePix()
```

Note that the data structure created by these commands contains *all* the GPR files. This structure in R is said to be of the *marray* type, which means that it is recognised as having several useful slots: for example "maRf" and "maRb" are the slots in which the red channel foreground and the red channel background data are stored (if the green channel is used, "maGf" and "maGb" are the slots for foreground and background), "maW" (spot quality *W*eights) contains the flag values and the subslot "maLabels" in the slot "maGnames" contains all the peptide names. The data are stored as matrices with the spots on the rows and the subarrays on the columns.

3. Quality Control

3.1. Inspection of Flags

To examine the quality of the data, the first step is to check the distribution of the flags. We performed this quality control exercise for each group, first by combining the data from all the subarrays and inspecting all spots regardless of whether they were from control or peptide spots, and subsequently stratifying by the type of feature. We also looked at each individual slide to check whether there were any strange or aberrant slides that should not be included in the analysis.

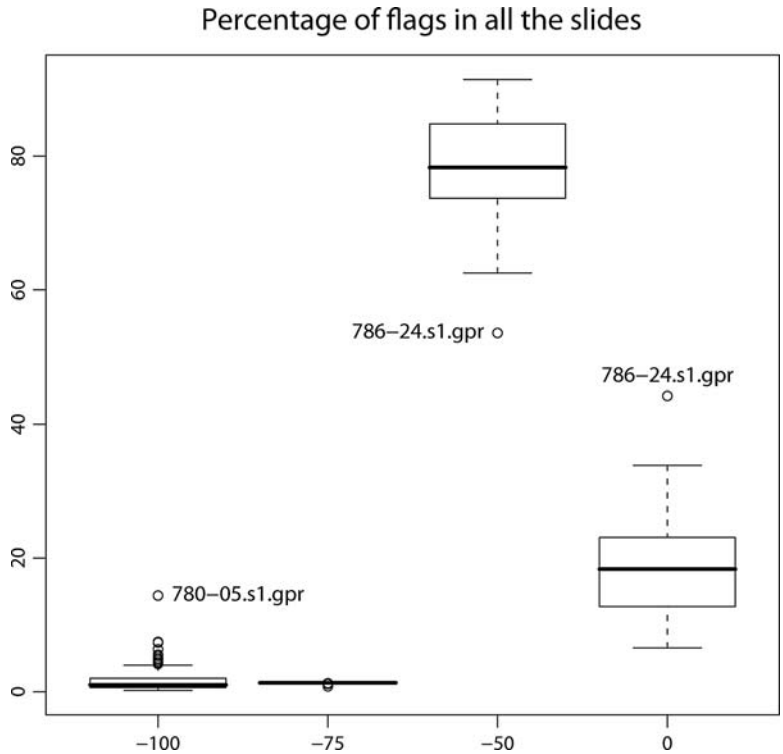


Fig. 21.3. Distribution of the flags across slides.

We present an example of one of our diagnostic boxplots of the flags in **Fig. 21.3**. This illustrates that the vast majority of flags represented spots that were "not found" (i.e. code = –50). There were relatively few "bad" flags (i.e. code = –100), less than 4% of the spots on most of the slides. In addition, subarrays with unusual numbers of flags (e.g. subarray 1 of slide 780-05 in **Fig. 21.3**) can be identified on the plots. We also inspected these proportions in each of the groups for analysis (for example, sick and healthy) to

see if there were relevant differences between groups. Such differences would indicate a potential bias and would be worthy of further investigation to determine if the slides can be validly compared. Finally, we checked that all "empty" spots were correctly flagged (code = –75).

3.2. Visual Inspection

Another important quality control step is the visual inspection of the data. We used the *image* function in Bioconductor to produce a graphical image of the spots on each slide to check if there were any slides that should be excluded from analysis. The images also highlight artefacts and one can verify if these are correctly flagged as "bad". In **Fig. 21.4**, we present images of two subarrays. The black grid divides the 24 printing blocks. **Figure 21.4a** represents the foreground of a "normal" subarray, with no obvious problems related to the quality of the data. In contrast, in **Fig. 21.4b** an artefact is clearly visible in the lower half of the subarray. This could be the result of a scratch on the subarray surface. In this case, the previous step had already highlighted this slide (number 780-5, subarray 1) as having a large number of bad flags, so we checked whether the dark area visible in **Fig. 21.4b** was flagged, as it should be. If the flags did not correspond correctly, then the subarray should be re-examined manually and the area correctly flagged. If this is not feasible, then the entire subarray could be removed from the analysis.

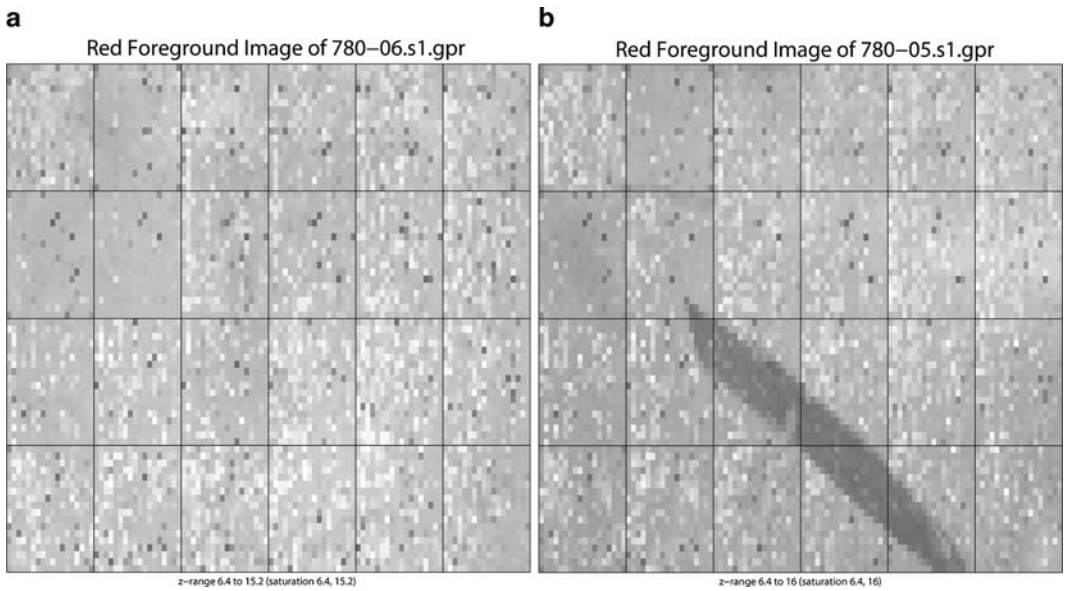


Fig. 21.4. Display of the red intesity of the 635nm foreground signal from two subarrays, obtained using the Bioconductor "image" function.

We performed similar plots for the background signals. These images are much more variable, but our interest is in identifying artefacts not due to the usual random background variability. For example, we found the same "scratch" as was apparent in the foreground signal for slide number 780-05 subarray 1 (data not shown). Thus images such as those in **Fig. 21.4** can provide a good indication of whether slides are spoiled or can still be included in analysis.

4. Choice of Response Measure and Construction of a Working Dataset

We chose the *index* value, the log-transformed ratio of foreground to background (6) as the measure of the strength of the response. This value was computed for all spots with background greater than zero (those few with zero background, and thus undefined index, were noted and excluded). Since the data are stored in two matrices, one for foreground and one for background, this computation is simply the ratio of these two matrices, taken element-by-element.

The index values can also be visualised using the image function as before. This will often present a much more homogenous image than the raw foreground, as correcting for variable background using the ratio eliminates many visual irregularities. For example, the darker zone in the bottom part of **Fig. 21.5a** disappears when the index is calculated (**Fig. 21.5b**).Variation from block to block in the raw data will also usually be considerably reduced in the index image, and any remaining block effects can be removed during normalization (*see* **Section 7**). In some

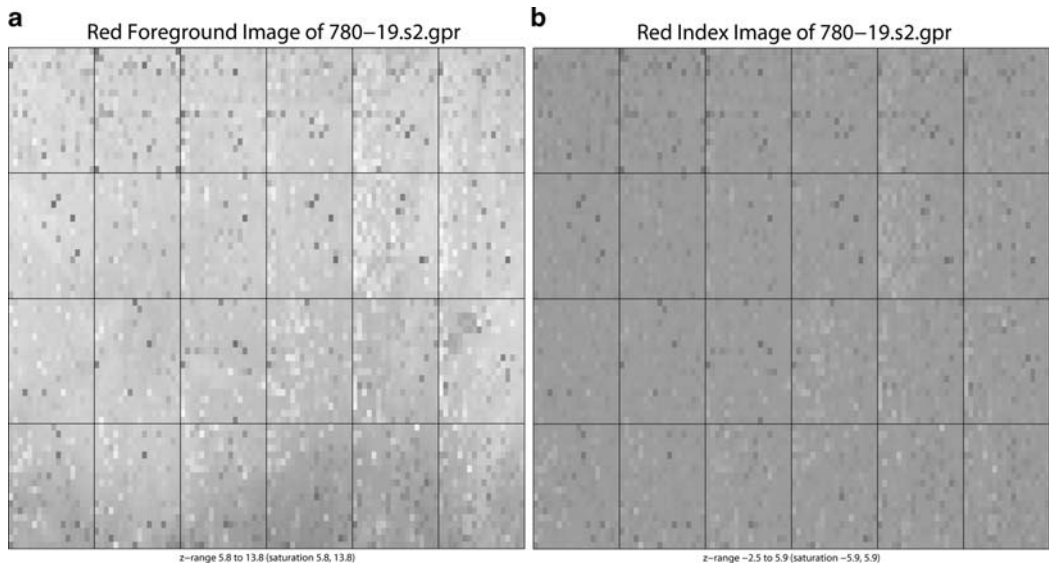


Fig. 21.5. Plot of (**a**) red intensity of foreground 635nm signal and (**b**) corresponding index of a single subarray.

applications, users may prefer other definitions of response, for example the difference (or log difference) of the foreground and background signal (7).

The chosen response measure is examined for each slide, and for all slides from a group combined, in order to identify and flag any extreme outliers that have not already been flagged. A plot of the index vs. the log-background can be very informative (*see* **Fig. 21.6**), as it shows the range of background values and the distribution of the index values for peptide and control spots. In this example, the negative control responses are clustered around zero as expected, with the exception of one spot with an extreme value. In contrast, the peptide spots have a wider "cloud" around zero, but also a number of spots with high responses. Since the true response of a peptide is unknown, further analysis is needed to determine if these are all valid responses or whether some are outliers that should be excluded.

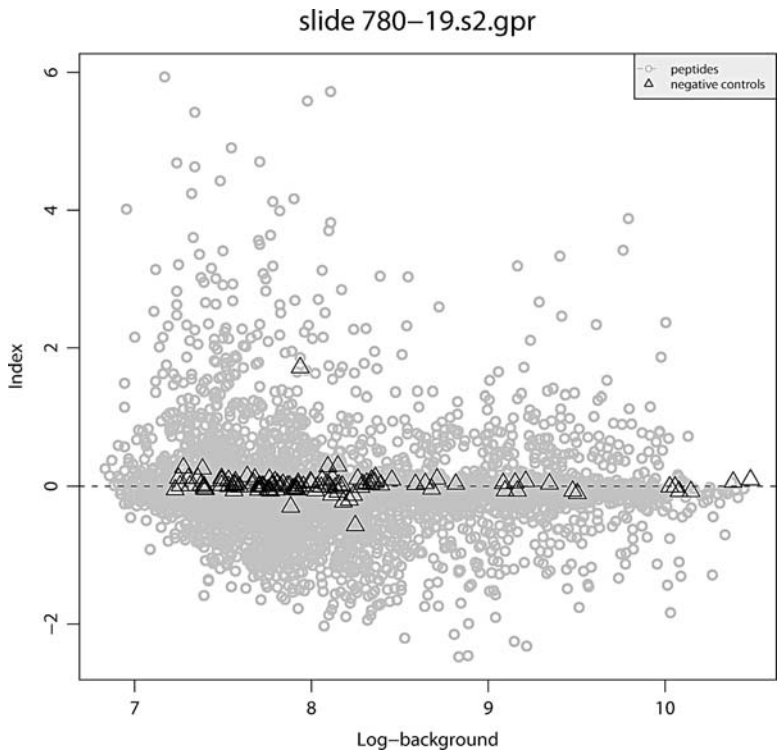


Fig. 21.6. Scatter plot of log-background vs. index for negative controls and peptides from a single subarray.

Lastly a working dataset is prepared containing the data from all subarrays that were deemed valid by the inspection described above, and with flags identifying spots that will not be analysed. The data from each group of slides are stored in a large matrix which will now contain an additional column with the chosen response measure for each spot.

5. Data Reduction

After the quality control steps described above have been carried out, and before any comparative analyses are performed, there are a number of ways in which the volume of data can be reduced while still maintaining the important information. As the technology becomes capable of measuring greater detail and consequently producing ever larger datasets, this reduction in the volume and dimension of the data is of growing importance for computational efficiency.

When thousands of peptides are printed, there are likely to be many whose intensity values are not substantially different from experimental noise. In some applications (e.g. our **Fig. 21.3** above), a large number of peptides may be flagged as "not found". This code usually indicates a very low-intensity signal from "non-responding" peptides, but can also be due to other problems, e.g. the shape of the spot. If removal of all the "not found" flags would result in a drastic reduction of the dataset, it may be preferable to remove only those "not found" spots that have a high (bright) intensity. These have presumably been flagged not for their low intensity but for other problems that are allocated this same flag code. By doing this, many low responding spots are retained in the data, but these can be informative if they represent peptides that have a high response on one or more of the other slides. In contrast, peptides that have a low response on all slides do not provide important information and these can be removed as explained in **Section 5.1** below.

Another important preprocessing step is the removal of false positives. We define as false positive those peptides that have a high response on the slides containing only buffer and secondary antibody. These responses are clearly not related to an immune response, but to the interaction between the peptides and the materials used for the measuring process, and can thus not be interpreted on patient slides. Removing these peptides prior to any comparative analysis is very important, as even if they are relatively few, they are at risk of being falsely identified as having differential response due to their high values in one group. We will illustrate below (**Section 5.2**) a method for identification of these false-positive peptides.

Before we can validly characterise the non-detectable and false-positive peptides, we need to remove any artefacts from the data. Some of the differences between peptide responses will be due to the slides, subarrays and blocks that they are printed on, i.e. due to technical variability and not the biological variability that is our primary interest. Thus, for any

meaningful analysis, the responses from all spots must be "normalized" to remove such artefacts and make them as equivalent as possible to the responses that would have been obtained under the exact same experimental conditions. To avoid interrupting the flow of the procedures in the following three sections, we will simply state when appropriate that the data were normalized and refer the reader to **Section 7** where we describe and illustrate a method of normalization.

5.1. Identification and Removal of Peptides with Non-detectable Response

It is usual in microarray analysis to remove the spots with very low signal (8). One method of defining a very low response is to use the distribution of the responses from the negative controls as a reference. The negative controls on our arrays are flagged as "empty" spots, as there is nothing printed on those locations. Thus the signal we obtain from the negative controls is random experimental noise, so it can be used to define a cut-off for the intensity that can considered a real response. We defined a "detectable response" as an intensity at least two standard deviations above the mean of the normalized negative controls for a given group of slides. Given the large number of spots, this definition will likely define many spurious spots as detectable, but if the definition helps to reduce the data volume this high sensitivity is desirable since it protects against losing any potentially interesting spots. In our analysis we had two groups of individuals, a group of patients and a group of healthy controls. Using our working dataset (from which we have eliminated extreme outliers), we normalized the negative control responses on all slides in a group to remove the effects of slide, subarray and block (*see* **Section 7** for details on normalization). From the mean and SD of the normalized values we define the threshold for a detectable response as $t = \text{mean} + 2\text{SD}$ (*see* **Fig. 21.7**).

Using the cut-off determined from the negative controls, we identified the spots with non-detectable responses on each slide. If a peptide had no detectable response on any slide it was excluded from further analysis. All other peptides had all their responses included: i.e. any peptide that had a detectable response on at least one slide had its responses from all slides included in the analysis, whether or not these were above or below the cut-off. If these peptides were all automatically eliminated from the data to be analysed, a large volume of data would be lost. Instead, we have maintained all potentially important peptides while significantly reducing the dimension of the data to allow more flexibility for subsequent analyses. In our example, the data records were reduced by almost 20% in both groups.

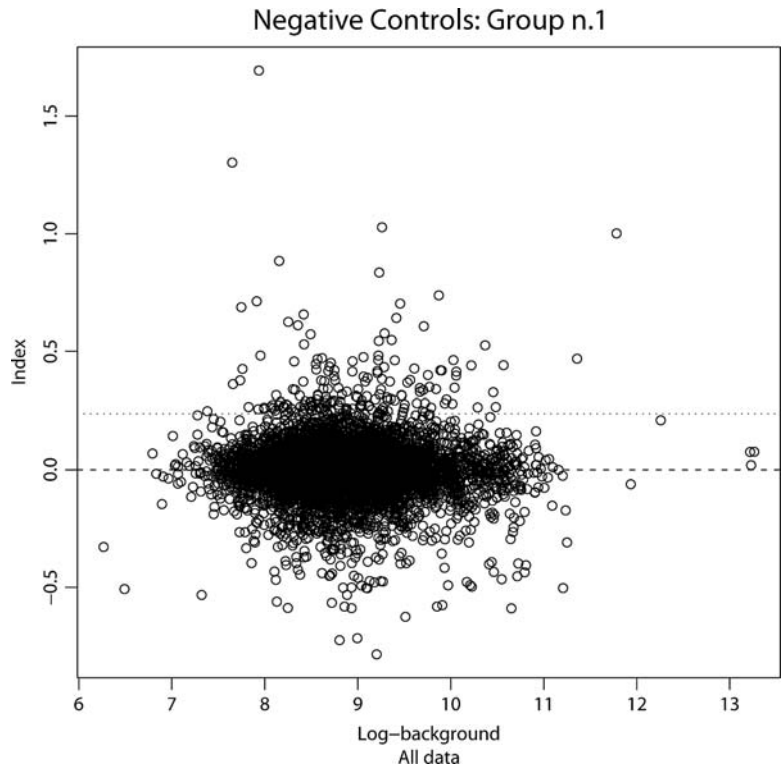


Fig. 21.7. Index vs. log-background plot for normalized negative controls in the group of TB patients.

5.2. Identification of False Positives

Some peptides will show a high index not because of a true immune response, but because of the technical limitations of the process of measurement. Such "false-positive" responses can occur for example if the peptides interact with the buffer and/or the secondary antibody used on the slide. For any batch of slides to be used with human specimens, data from slides that have only buffer and secondary antibody (i.e. no human specimen) can be used to identify false-positive responses. Since no peptides are expected to give a true response on a slide where only buffer and secondary antibody are printed, all the "responding" peptides are considered as false positives. In our application, we prepared 13 slides using only buffer and secondary antibody (six from the batch used for the TB patients and seven from the batch used for the healthy controls). We normalized all the valid (i.e. not flagged as –100) peptide responses on these slides using the method described in **Section 7** for each group in turn.

From a scatter plot of the normalized indices vs. normalized background, the responses that are likely to be false positives can be identified (**Fig. 21.8a,b**). QQ-plots of the normalized indices (**Fig. 21.8c,d**) show clearly the two populations of peptides on these slides, i.e. the non-responding

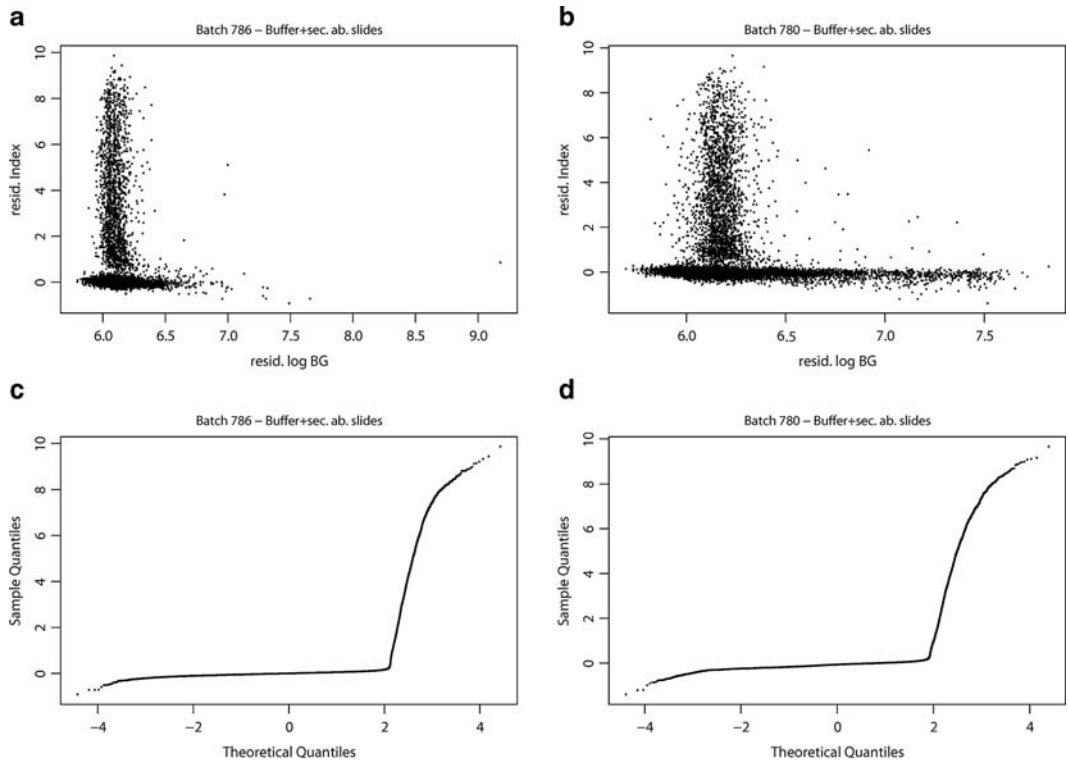


Fig. 21.8. Index vs. log-background for responses from slides with only buffer and secondary antibody in (**a**) batch 786, corresponding to the TB patient group, and (**b**) batch 780, corresponding to the healthy group. The corresponding QQ-plots of the response index for the two batches are shown in (**c**) and (**d**), respectively.

(as expected) and the responding (false positive) peptides. We declared all the peptides with a normalized value above 0.2 on the buffer and secondary antibody slides to be false positives in the two groups. We did not expect a priori to find the same cut-off value for false positives in both groups. However, since these slides are similar and use no patient specimen, it is not surprising that a similar cut-off value is appropriate. These peptides were excluded from any analyses comparing the immune responses in the two groups of study subjects.

6. Preparation of Data for Use in Comparative and Predictive Analyses

As in all microarray experiments, to be able to perform a differential response analysis (to compare the peptide responses between two groups) the data must be comparable. Data coming from different slides and, inside each slide, from

different subarrays and different locations in the subarray are not directly comparable. Some of the differences between the peptide responses from different groups of individuals may be due to the differences in the signal strength from different slides. For example, a spot (peptide) may be brighter simply because it is located in a brighter zone of the subarray and not because it has a higher response. Thus the final step prior to any comparative or predictive analysis is to remove these systematic effects from the responses so that the data can be analysed to reveal real biological differences. The removal of systematic effects is called "normalization".

The normalization procedure (**Section 7**) is applied to the index values in the clean working dataset (i.e. excluding peptides flagged as "bad", false positives and peptides with a non-detectable response on all slides). The only systematic effect that is not included in the normalization models is the peptide effect. Thus the differences between the observed responses and the responses predicted by the model (i.e. the residuals) provide the estimates of the peptide effects. Since the systematic effects of slide, subarray and block have been removed, we refer to these as the "normalized responses", and these provide the input data to software tools for comparative or predictive analysis (9–11). For those peptides that have replications, their normalized values are averaged, and an average is also computed for each pair of subarrays belonging to the same slide. The result of this averaging is a list of unique peptides with their normalized responses for each slide (*see* **Fig. 21.9**). This information can be stored in one data matrix for each group.

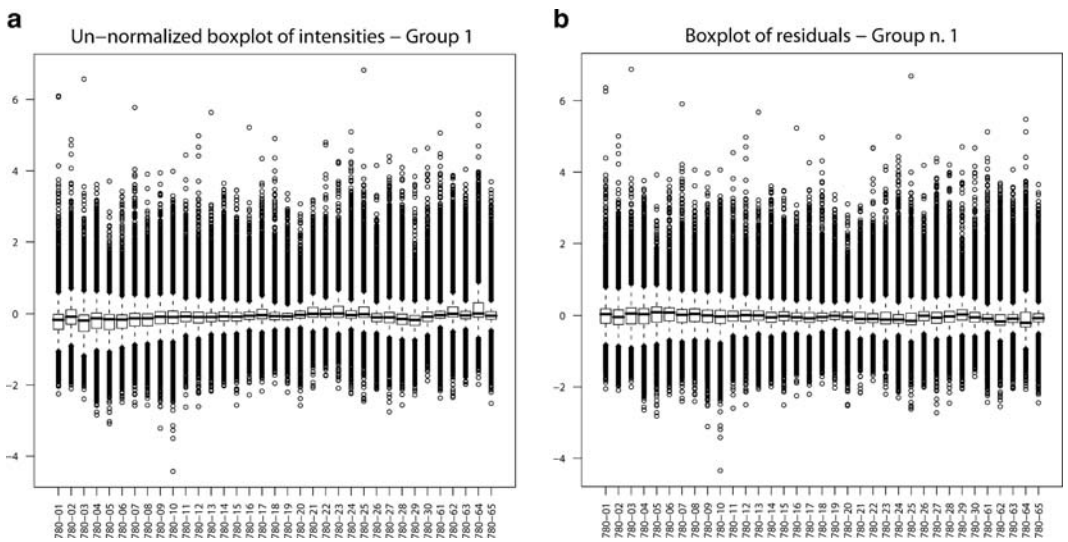


Fig. 21.9. Boxplot of index values for peptide responses from 34 TB patients (**a**) before and (**b**) after normalization.

7. Normalization

The purpose of normalization is to adjust the response from each spot on a slide for the systematic effects due to the variation between slides, subarrays and blocks. Numerous methods of normalization have been developed for genomic microarray data (12–15), and several of these have been used in the field of peptide microarrays (8, 16). Many of the available normalization methods from genomics make the assumption that only a very small proportion of genes will be differentially expressed (13, 17). In applications to peptide arrays, this would be equivalent to assuming that only a very small proportion of peptides will have a different response in the groups being compared. We were unwilling to make this assumption and chose instead to use linear regression models to estimate the slide, subarray and block effects in the responses and to correct for these measurement artefacts.

We denote the response from the spot on slide i, subarray j, block k as $Y_{ijk.}$ The normalization of a particular type of feature (e.g. negative controls or peptides) is carried out by fitting one of the following linear models, which is applied simultaneously to all spots (controls or peptides) on all slides in a group.

(i) $Y_{ijk} = \text{slide}_i + \text{subarr}_j + \text{block}_k + \epsilon_{ijk}$

(ii) $Y_{ijk} = \text{slide}_i + \text{subarr}_j + \text{block}_k + \text{slide}_i * \text{subarr}_j + \epsilon_{ijk}$

where slide_i is the "slide effect", subarr_j is the "subarray effect", block_k is the "block effect" and the interaction (subarray*slide) in model (ii) allows for the possibility that the subarray effect can be different from slide to slide. The term ϵ_{ijk} is the remaining signal (the "residual") when the systematic effects have been removed. For example, in **Fig. 21.10**, the subarray in (a) has a gradient across the slide so that a spot in the upper right corner block could have a different intensity than a spot situated in the lower left corner simply because of its position. Removal of the block effect will correct for this difference. Similarly, removal of the subarray effect corrects for situations where all the spots on one subarray are brighter or darker than on another, such as the subarrays in **Fig. 21.10b,c**.

Since the raw data from peptide arrays are highly skewed, it is usual to apply the normalization on the log scale. The response index is already on a log scale, but if one wishes to normalize the background or foreground signals, or their difference, then the data should first be log-transformed. We use whichever model provides the best fit to the data. For the response index, a simple normalization model such as (i) is often sufficient, as much of the variability has already been removed by taking the ratio of foreground and background signal as we saw earlier (Fig. **21.5**).

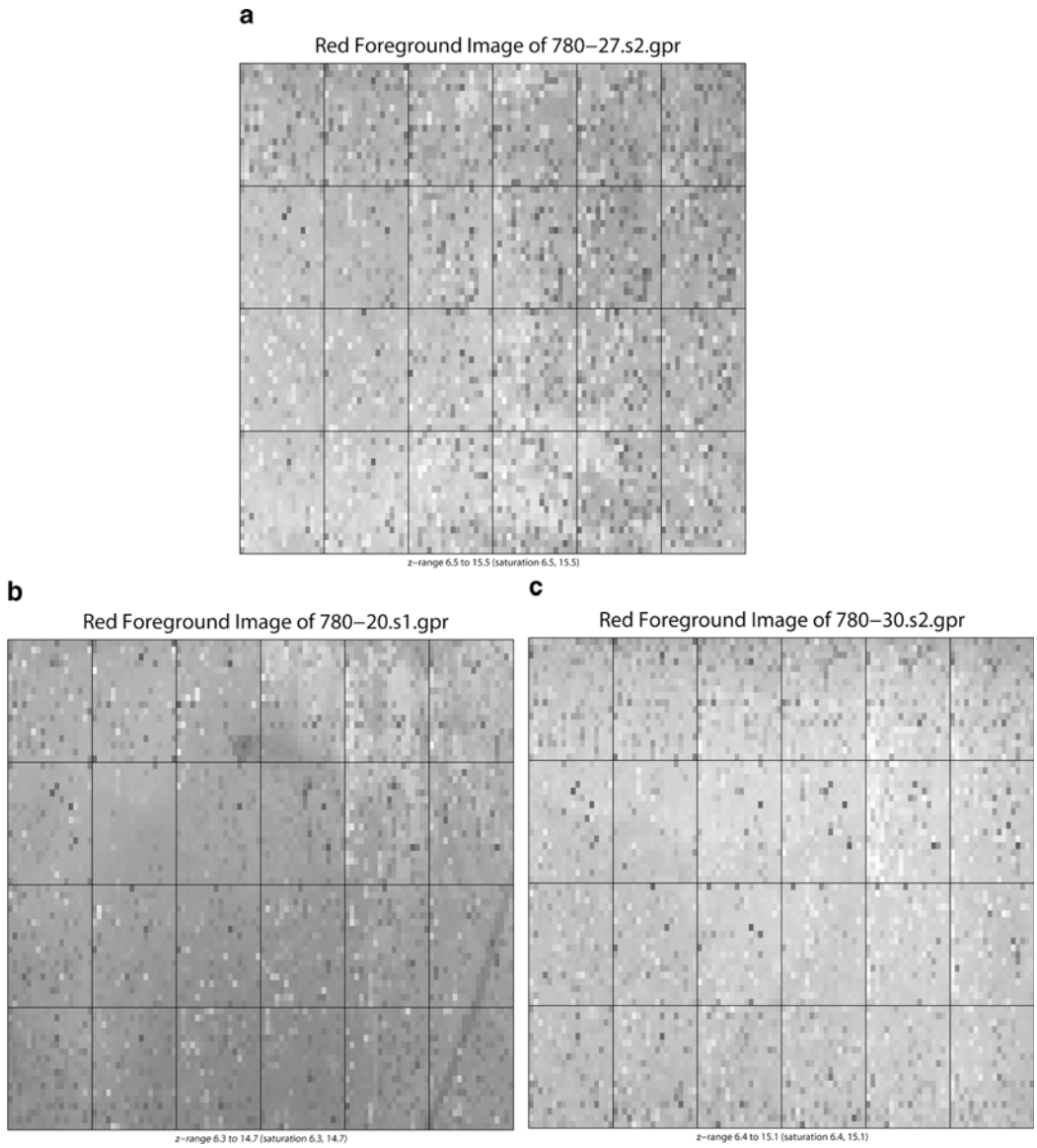


Fig. 21.10. Image plot illustrating block effects in (**a**) and subarray effects in (**b**) and (**c**).

The linear model can be run using the *lm* function in R, or if the number of spots to be normalized is too large, the *biglm* R-package is specifically designed to solve such computational problems.

Acknowledgements

The authors would like to thank Yen Ngo for many useful and insightful discussions about the management, display and analysis of peptide microarray data.

References

1. R Development Core Team (2007) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org
2. Gaseitsiwe S, Valentini D, Mahdavifar S, Magalhaes I, Hoft DF, Zerweck J, Schutkowski M, Andersson J, Reilly M, Maeurer MJ. (2008) Pattern recognition in pulmonary tuberculosis defined by high content peptide microarray chip analysis representing 61 proteins from M. tuberculosis. PLoS ONE 3(12).
3. Molecular Devices (2007) GenePix Pro, http://www.moleculardevices.com / pages/software/gn_genepix_pro.html
4. Agilent (2008) Feature Extraction Software http://www.chem.agilent.com/en-US/products/instruments/dnamicroarrays/featureextractionsoftware
5. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5(10):R80
6. Bioconductor developers (2008) URL: http://www.bioconductor.com
7. Nahtman T, Jernberg A, Mahdavifar S, Zerweck J, Schutkowski M, Maeurer M, Reilly M. (2007) Validation of peptide epitope microarray experiments and extraction of quality data. J Immunol Methods 328(1–2):1–13
8. Quintana FJ, Hagedorn PH, Elizur G, Merbl Y, Domany E, Cohen IR. (2004) Functional immunomics: microarray analysis of IgG autoantibody repertoires predicts the future response of mice to induced diabetes. Proc Natl Acad Sci USA 101(Suppl 2):14615–21
9. Shreffler WG, Lencer DA, Bardina L, Sampson HA. (2005) IgE and IgG4 epitope mapping by microarray immunoassay reveals the diversity of immune response to the peanut allergen, Ara h 2. J Allergy Clin Immunol 116(4):893–9
10. Tibshirani R, Hastie T, Narasimhan B, Chu G. (2002) Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci USA 99(10):6567–72
11. Tusher VG, Tibshirani R, Chu G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 98(9):5116–21
12. Ploner A, Calza S, Gusnanto A, Pawitan. (2006). Multidimensional local false discovery rate for microarray studies. Bioinformatics 22:556–565
13. Workman C, Jensen LJ, Jarmer H, Berka R, Gautier L, Nielser HB, Saxild HH, Nielsen C, Brunak S, Knudsen S. (2002) A new non-linear normalization method for reducing variability in DNA microarray experiments. Genome Biol 3(9):research 0048
14. Bolstad BM, Irizarry RA, Astrand M, Speed TP. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19(2):185–193
15. Affymetrix: Statistical Algorithms Description Document 2002
16. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP (2002), Normalisation for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 30(4):e15
17. Neuman de Vegvar HE, Amara RR, Steinman L, Utz PJ, Robinson HL, Robinson WH. (2003) Microarray profiling of antibody responses against simian–human immunodeficiency virus: postchallenge convergence of reactivities independent of host histocompatibility type and vaccine regimen. J Virol 77(20):11125–38
18. Quackenbush J. (2002) Microarray data normalization and transformation. Nat Genet 32(Suppl):496–501

Chapter 22

Web-Based Design of Peptide Microarrays Using μPepArray Pro

Tongbin Li, Zhixiang Zuo, Qi Zhu, Ailing Hong, Xiaochuan Zhou, and Xiaolian Gao

Abstract

Peptide microarrays (peptide arrays) have increasingly become an important research tool for studying protein detection, profiling, and protein–protein interactions, and they have the potential to foster high-throughput protein analysis as DNA arrays did for genomics research a decade ago. Recently, technologies have emerged that allow flexible synthesis of high-density peptide arrays based on specific application needs (e.g., phosphopeptide microarrays). To fully unleash the power of this promising research tool, significant efforts are required to develop computational and informatics resources that facilitate the experimental design and data analysis for a wide range of peptide array-based applications. The design of peptide arrays is inherently more complex than that of DNA arrays. We herein introduce μPepArray Pro, a Web-based general-purpose peptide array design program. μPepArray Pro features strong content design capabilities and maximized user control. The program suits the needs of a diversity of design tasks, works with a variety of peptide array configurations, and is highly expandable: new functionalities can be developed and added to μPepArray Pro with relative ease.

Key words: μPepArray Pro, peptide microarray, peptide array, array design.

1. Introduction

Recent technological advances have stimulated the development of a new generation of peptide array platforms that emphasize higher capacity, miniaturized scale, and more efficient synthesis (1–3). These new peptide array platforms promise to foster high-throughput research of protein functions much as maturing DNA array technologies did for genomics research a decade ago (4). In order to take full advantage of the power of these new array-based research platforms, significant efforts are required to develop computational and informatics tools assisting the experimental design

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_22,

and the data analysis for peptide array-based research. In this chapter, we take on the issue of computational design of peptide arrays.

The design of a peptide array is the process of determining a proper way to fill in – using a set of peptides – the spatially arranged spots (or features) in a pre-defined array configuration, in order to accomplish the objectives of a peptide array-based study. In most of the published peptide array studies, the array design was conducted in an "offhand" manner. This is in sharp contrast to the DNA array design practice, where a large number of computational tools – most of which are guided by well-established design principles (5, 6) – are available to help the researchers customize and optimize the array designs (e.g., (7–11)). In this work, we discuss the similarities and differences between the peptide array design and the DNA array design (in particular, the oligonucleotide array design), and describe the use of μPepArray Pro, the first general-purpose peptide array design tool. μPepArray Pro is a Web-based program capable of guiding the researcher through the entire peptide array design process – from content design (task-based seed peptide selection, derived peptide generation, and peptide sequence evaluation) through array layout writing. μPepArray Pro is developed to suit a diversity of design needs and work with a variety of peptide array configurations, and it is highly expandable – new functionalities can be developed and added into each element of the array design process with relative ease.

2. Overviews of Peptide Array Design Process and μPepArray Pro

2.1. Peptide Array Design vs. Oligonucleotide Array Design

The aim of an array design process is to design a set of *probes*, and devise a proper way to fill in the spatially arranged spots (or features) using these probes in a pre-defined array configuration, to accomplish the objectives of an array-based study. What differentiates the peptide array design from the oligonucleotide array design is that the probes used to fill in the spots are peptide molecules instead of oligonucleotides. There are issues and considerations that are shared between designing a peptide array and designing an oligonucleotide array, e.g., the use of spot replicates for accessing and improving the confidence of the measurements, and the preferred use of randomized layouts to offset possible spatial bias introduced during array production or array experiments. However, there are also considerable differences between a peptide array design process and an oligonucleotide array design process, due to the different purposes of the two types of array

experiments, and due to the different technicalities involving the synthesizing or spotting of the two different types of molecules on an array surface.

First, because of the different purposes of the two types of array experiments, the researcher is much more involved and plays a much more active role in the content design of a peptide array than in that of the oligonucleotide array. Research works involving oligonucleotide array experiments are often of a "surveying" or "measuring" type. In these studies, researchers are often concerned with the genes (for a gene expression study) or the chromosome locations (for a comparative genomic hybridization study or a ChIP-chip study), but not particularly with the probes that are designed to target the genes or chromosome locations. In fact, in a typical oligonucleotide array design process, after the researcher specifies a list of genes or chromosome locations, the rest of the content design is dominated by well-established probe design principles (about avoiding cross-hybridization, and about controlling ΔT_m variation, sequence complexity, and folding energy for selecting high-quality probes) and requires minimal input from the researcher. In contrast, studies involving peptide array experiments are often "hypothesis-driven" in nature. In these studies, the researchers are immensely interested in probe-level details – what is the peptide sequence, how is the sequence generated (is it part of a natural protein or is it an artificial peptide? – in the latter case, are there similar peptides occurring in any natural proteins?)? Therefore, in peptide array design, the researcher needs to sit in the "driver's seat" and be given much more control over the content design process than in oligonucleotide array design. A proficient peptide array design program must allow ample flexibility for the user to compose the lists of peptides freely as he or she desires; meanwhile, it should provide the user with adequate assistance in the functionalities he or she needs (e.g., for protein database look up, for derived peptide writing, i.e., generating sequences related to a given peptide as specified by the user) to carry out the array design efficiently.

Second, due to the different types of molecules synthesized or spotted on the array surface, the rules (or principles) governing the quality of the probes are different between oligonucleotide arrays and peptide arrays. The factors influencing the probe quality in oligonucleotide arrays have been investigated for many years (5), and they include ΔT_m variation, sequence complexity, and the formation of secondary structures. In contrast, rather limited work has been reported to investigate factors influencing the quality of the probes in peptide array design (12). These factors are likely to be different between in situ synthesized peptide arrays and spotted peptide arrays, and when different peptide synthesis techniques are compared. A useful feature of a proficient peptide array design program should have is to put in place an *undesired sequence pattern filter* to help the researcher remove peptides that are problematic during array synthesis.

2.2. μPepArray Pro Features and a Dissection of the Design Flow

μPepArray Pro is a fully functional peptide array design program developed to suit a diversity of design needs and work with a variety of peptide array configurations, although it was developed with μParaflo® microfluidic chip platform (13) as the primary target platform. Conceptually, a μPepArray Pro-based peptide array design process consists of two components – the *content design component* and the *array layout component*. The content design component further comprises two elements: (i) constructing *task-based peptide groups* (TBPGs) and (ii) assembling the *array-level peptide list* (ALPL) (**Fig. 22.1**).

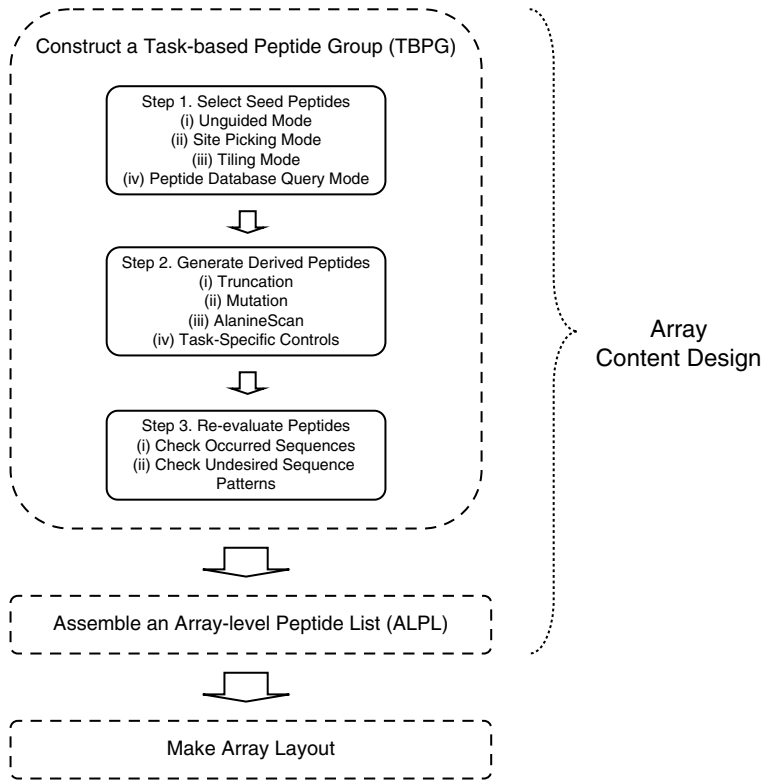


Fig. 22.1. A flow chart of peptide array design process using μPepArray Pro.

In a peptide array study, the researcher often desires to carry out multiple, relatively independent *tasks* on a single array experiment. For example, in a B-cell epitope screening study, the researcher might want to examine (with a tiling design paradigm) two different segments of the antigen protein. As another example, in a kinase substrate specificity study, the researcher might want to include the reported substrate peptides (and some peptide sequences derived from them) of several different kinases in a

single peptide array, and perform binding experiments using each kinase to check its interactions with its reported substrates as well as its cross-reactions with the reported substrates of other kinases. In μPepArray Pro, the term *task-based peptide group* (TBPG) is used to refer to the group of peptides designed for a single *task*. In the example of B-cell epitope screening, the tiling peptides for each of the two segments constitute a TBPG. Similarly, in the kinase substrate specificity study example, the known substrates for an individual kinase and the peptides derived from these known substrates constitute a TBPG.

With one or more TBPGs constructed, the user would assemble an *array-level peptide list* (ALPL) – the list of all peptides (with replicate information) to be included in a peptide array. The number of peptides included in an ALPL is constrained by the capacity of the selected *array configuration*, which defines the length and width of the array (in numbers of spots, e.g., 128 × 31) as well as the locations of spots reserved for quality control purposes.

Finally, the array layout component of μPepArray Pro is responsible for making the layout of the array, conforming to the specified array configuration according to specifications provided by the user.

μPepArray Pro is accessible at http://pepcyber.umn.edu/microPepArray using a recent Web browser (Internet Explorer V5.5 or higher, Firefox V 1.5 or higher) with JavaScript enabled. Users are required to register with a valid email address before being able to use the program.

2.3. TBPG Construction Explained

The TBPG is a basic unit of the content design component, and TBPG construction is the most involved element of the peptide array design process with μPepArray Pro. The construction of a TBPG consists of three steps: (1) seed peptide selection; (2) derived peptide generation; and (3) peptide assessment and filtering. Considering the diversity of content design needs researchers have, each of these steps is developed to be easily expandable, i.e., new functionalities can be developed and easily added into each of these steps to address new requirements raised by users.

At the first step, the user formulates a list of seed peptides that are most essential to the design task. These peptides are termed *seed peptides*. For example, in a B-cell epitope screening task, the user might choose to use a tiling design to cover a segment of the antigen protein using a number of tiling peptides. These tiling peptides would be considered as seed peptides for this design task. In a kinase substrate specificity study, the user might choose to include – in the list of seed peptides – all known substrate sequences of the kinase. Currently, four modes of creating seed peptide lists are

implemented in μPepArray Pro: (i) unguided mode, in which the user would prepare the seed peptide list manually; (ii) site picking mode, in which the user would pick a number of peptides truncated from a protein sequence; (iii) tiling mode, in which tiling peptides would be made along a protein sequence in a region specified by the user; and (iv) peptide database query mode, in which the user would obtain a list of peptides by querying an peptide database. At present, the function querying a specific peptide database, PepCyber:P∼Pep (14), has been implemented.

At the second step, the user could choose to include a number of *derived peptides* – which are peptide sequences that are related to selected seed peptides – into the TBPG. These derived peptides are added either for exploration purposes or as control peptides for the design task. Currently, four protocols of generating derived peptides are implemented in μPepArray Pro: (i) truncation, in which shorter peptide sequences are generated with either or both ends of an existing peptide sequence being truncated; (ii) mutation, in which new peptides are generated by mutating one or more amino acids in an existing peptide sequence to different amino acids; (iii) alanine-scan, in which each amino acid in an existing peptide sequence is substituted for by alanine, one at a time, to generate a list of new sequences; and (iv) task-specific controls – at present one task-specific control peptide generating function has been implemented for replacing one or more phosphorylatable amino acids (serine, threonine, and tyrosine) by a non-phosphorylatable amino acid as specified by the user.

At the third step, the peptide sequences generated during the first two steps are *re-evaluated*. Currently, two peptide re-evaluation functions are included in μPepArray Pro: (i) occurred sequence checking, where the peptide sequences are queried against the Swiss-Prot protein database, and the peptides that have occurred in any proteins in the protein database are highlighted for the user to examine; and (ii) undesired sequence pattern checking, where all peptides are checked for undesired sequence characteristics that may lead to synthesis efficiency problems. At present, the undesired sequence pattern checking function is tuned for the μParaflo® microfluidic on-chip peptide synthesis platform. Similar functions for other peptide array platforms can be easily added in the future.

3. Methods

3.1. Constructing a TBPG

After logging in, the user can choose the option *Make/View TBPGs* to arrive at a page where he or she can start to construct a new TBPG, or look up a previously constructed TBPG. The user

then chooses the *Construct a new TBPG* option, and he or she is guided through the three-step TBPG construction process: (1) seed peptide selection; (2) derived peptide generation; and (3) peptide assessment and filtering.

3.1.1. Selecting Seed Peptides

Four modes of seed peptide selection are provided in μPepArray Pro: (i) unguided mode; (ii) site picking mode; (iii) tiling mode; and (iv) peptide database query mode (for the PepCyber:P∼Pep database).

If the user chooses the *Unguided mode* for selecting seed peptides, he or she is prompted to provide a name and (optionally) a textual description of the seed peptide list. The list of seed peptides can either be uploaded as a comma separated value (CSV) file or be pasted to the text area provided. The data file (or pasted text) should contain no header row and include two columns: (a) sequence of the seed peptide; and (b) a brief description of the seed peptide.

If the user chooses the *Site picking mode* for selecting seed peptides, he or she is prompted to paste a protein sequence and (optionally) provide a description of the protein sequence pasted. Alternatively, the user can provide the Swiss-Prot accession of a protein, and the sequence and description of the protein is automatically retrieved from the Swiss-Prot database. Next, the user can select the number of peptides truncated from the protein sequence as seed peptides.

If the user chooses the *Tiling mode* for selecting seed peptides, he or she is prompted to provide a protein sequence, or provide the Swiss-Prot accession of a protein, as in the *Site picking mode*. Then he or she can specify the parameters for the tiling design. These parameters include the starting and ending positions, the peptide length, and the shift between neighboring peptides. A list of tiling peptides would be generated, and the user can choose to include any number of peptides in the list as seed peptides.

In the *Peptide database query mode* for seed peptide selection, the user could set query parameters and submit a query to the PepCyber:P∼Pep database. The user can examine the list of phosphopeptides returned, and choose to include any number of the phosphopeptides as seed peptides.

The detailed information about how each seed peptide is selected is documented automatically, and is displayed when the *Detailed info* button is clicked in the seed peptide table. During seed peptide selection, a summary of the number and type of the seed peptides is displayed in the "information box" in the right panel. After the seed peptide list is finalized, the user will enter the second step of TBPG construction: derived peptide generation.

3.1.2. Generating Derived Peptides

Four different protocols are available for generating derived peptides: (i) truncation; (ii) mutation; (iii) alanine-scan; and (iv) task-specific controls (non-phosphorylatable amino acid substitution). Starting from the seed peptide list, the user could choose a subset of the peptides and apply to these peptides any of these four protocols for multiple times in an iterative manner to generate a list of derived peptides. To achieve this, μPepArray Pro maintains a *current peptide list* (CPL) at any time during derived peptide generation. The CPL includes both the seed peptides and the derived peptides that have been generated thus far. The user chooses a subset of the peptides in the CPL, and chooses one of the four derived peptide generating protocols to apply to the selected peptides to generate a number of new derived peptides. The CPL is then updated to include these newly generated derived peptides. This process continues iteratively until the user decides that all derived peptides he wishes to include in the task have been generated.

When the *Truncation* protocol is selected, the user is prompted to select a subset of peptides from the CPL and choose from three options: (a) truncating from the N-terminal side; (b) truncating from the C-terminal side; and (c) truncating from both sides. After the truncated peptides are generated, the user can include some or all of these new peptides in the CPL.

The *Mutation* protocol can be applied to only individual peptides, i.e., the user can only choose a single peptide from the CPL when the *Mutation* protocol is specified. The user is prompted to specify which amino acids in the selected peptide he or she would like to be mutated, and which amino acids they should be mutated to. The user can choose to include some or all of the newly generated mutated peptides to the CPL.

When the *Alanine-scan* protocol is applied, a list of new peptides is generated by mutating each amino acid in the selected peptides to alanine, one at a time.

The *Task-specific controls* protocols allow the generation of control peptides in a task-specific manner. Currently, one task-specific controls protocol – non-phosphorylatable amino acid substitution – is implemented. When this protocol is selected, the user is prompted to choose a number of peptides containing phosphorylatable amino acids (serine, threonine, or tyrosine), and specify which amino acid(s) they should be substituted for.

The "history" about how each derived peptide is generated is automatically documented, and is displayed when the *Detailed info* button is clicked in the CPL table. During derived peptide generation, a summary of the peptides in the CPL is displayed in the "information box" on the right panel. After the derived peptide list is finalized, the user enters the third step of TBPG construction: peptide re-evaluating and filtering.

3.1.3. Re-evaluating Peptides

At this step, the peptide sequences generated at the previous steps are *re-evaluated*. Two peptide re-evaluation functions are available: (i) occurred sequence checking and (ii) undesired sequence pattern checking.

The *Occurred sequence checking* function sends a query to the Swiss-Prot database for each peptide sequence. All peptide sequences that have occurred in any proteins in Swiss-Prot are highlighted for the user to examine. The user can choose to remove any peptides that he considers problematic.

The *Undesired sequence pattern checking* function checks the peptides for sequence patterns that may lead to synthesis efficiency problems. Currently, a collection of sequence patterns known to cause difficulty in on-chip peptide synthesis on the μParaflo® microfluidic chip platform are checked. The peptides carrying these undesired sequence characteristics are tagged. The user can choose to remove some or all of the tagged peptides as he or she deems proper.

Following peptide re-evaluation, if the user is satisfied with the TBPG constructed, he or she can click the *Submit the TBPG* button. The user is prompted to provide a name and (optionally) a description of the TBPG. Following of a routine error-checking procedure, the TBPG is stored into the supporting database of μPepArray Pro.

3.2. Assembling ALPL

In μPepArray Pro, the user can choose the option *Make/View ALPLs*, then select *Assemble a new ALPL* to start assembling an ALPL. The user is first prompted to select an array configuration: he or she can either choose one from the list of existing array configurations, or follow the instruction to define a new array configuration. After the array configuration is selected, the user is guided to add one or more TBPGs to the ALPL. For each TBPG added, the user also needs to specify the number of replicates. The "information box" on the right panel displays the total number of spots in the array configuration chosen, the number of filled spots, and the number of spots remaining to be filled. After the user is satisfied with the ALPL assembled, he can click the *Submit the ALPL* button. After filling in the name and an optional description, the ALPL is stored into the supporting database of μPepArray Pro.

3.3. Making Array Layout

After assembling an ALPL, the user can choose the option *Make/View Layouts*, then select *Make a new layout*. The user is prompted to select an ALPL to use, and choose from three layout options: (1) randomized layout; (2) sequential layout (by rows); and (3) sequential layout (by columns).

If the *Randomized layout* option is chosen, the peptides in the chosen ALPL are arranged to the available spots in the chosen array configuration in a completely randomized manner. This option is suitable for most common array design purposes.

If the *Sequential layout (by rows)* or the *Sequential layout (by columns)* option is chosen, the peptides in the chosen ALPL are used to fill in the available spots sequentially in the row-by-row or column-by-column manner, respectively.

After the array layout is made, the user can submit it to the supporting database of μPepArray Pro. A stored array layout (with rich information about all peptides) can be retrieved from the database and output to a proper file format to direct the array production at a later time.

Acknowledgments

This work was supported in part by NIH/NCI (1R21CA126209 to XG and TL), Minnesota Medical Foundation (to TL) and the R. A. Welch Foundation (E-1270, to XG) and NIH/GM/AI (R43 GM076941 to XG and XZ). We thank Dr. K. Cassidy for critically reading the manuscript.

References

1. Gao, X., Pellois, J.P., Na, Y., Kim, Y., Gulari, E. and Zhou, X. (2004) High density peptide microarrays. In situ synthesis and applications. *Mol Divers*, **8**, 177–187.
2. Schutkowski, M., Reimer, U., Panse, S., Dong, L.Y., Lizcano, J.M., Alessi, D.R. and Schneider-Mergener, J. (2004) High-content peptide microarrays for deciphering kinase specificity and biology. *Angewandte Chemie-International Edition*, **43**, 2671–2674.
3. Tapia, V., Bongartz, J., Schutkowski, M., Bruni, N., Weiser, A., Ay, B., Volkmer, R. and Or-Guil, M. (2007) Affinity profiling using the peptide microarray technology: a case study. *Anal Biochem*, **363**, 108–118.
4. Brown, P.O. and Botstein, D. (1999) Exploring the new world of the genome with DNA microarrays. *Nat Genet*, **21**, 33–37.
5. Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. et al. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechmol*, **14**, 1675–1680.
6. Kaderali, L. and Schliep, A. (2002) Selecting signature oligonucleotides to identify organisms using DNA arrays. *Bioinformatics*, **18**, 1340–1349.
7. Rouillard, J.M., Herbert, C.J. and Zuker, M. (2002) OligoArray: genome-scale oligonucleotide design for microarrays. *Bioinformatics*, **18**, 486–487.
8. Rimour, S., Hill, D., Militon, C. and Peyret, P. (2005) GoArrays: highly dynamic and efficient microarray probe design. *Bioinformatics*, **21**, 1094–1103.
9. Wernersson, R., Juncker, A.S. and Nielsen, H.B. (2007) Probe selection for DNA microarrays using OligoWiz. *Nat Protoc*, **2**, 2677–2691.
10. Mrowka, R., Schuchhardt, J. and Gille, C. (2002) Oligodb – interactive design of oligo DNA for transcription profiling of human genes. *Bioinformatics*, **18**, 1686–1687.
11. Auburn, R.P., Russell, R.R., Fischer, B., Meadows, L.A., Sevillano Matilla, S. and Russell, S. (2006) SimArray: a user-friendly and user-configurable microarray design tool. *BMC Bioinformatics*, **7**, 102.

12. Hong, A., Zhu, Q., Na, Y., Qi, Y., Drabek, R., Yao, W., Srivannavit, O., Gulari, E., Zhang, X., Zhou, X. et al. (2005) In Gierasch, L. M. (ed.), *The 19th American Peptide Society Symposium*. Wiley Interscience, San Diego, CA, pp. 498.
13. Zhu, Q., Hong, A., Sheng, N., Zhang, X., Matejko, A., Jun, K.Y., Srivannavit, O., Gulari, E., Gao, X. and Zhou, X. (2007) Microparaflo biochip for nucleic acid and protein analysis. *Methods Mol Biol*, **382**, 287–312.
14. Gong, W., Zhou, D., Ren, Y., Wang, Y., Zuo, Z., Shen, Y., Xiao, F., Zhu, Q., Hong, A., Zhou, X. et al. (2008) PepCyber:P∼PEP: a database of human protein protein interactions mediated by phosphoprotein-binding domains. *Nucleic Acids Res*, **36**, D679–683.

Chapter 23

Qualitative and Quantitative Analysis of Peptide Microarray Binding Experiments Using SVM-PEPARRAY

Gang Chen, Zhixiang Zuo, Qi Zhu, Ailing Hong, Xiaochuan Zhou, Xiaolian Gao, and Tongbin Li

Abstract

A main objective of analyzing peptide array-based binding experiments is to uncover the relationship between a peptide sequence and the binding outcome. Limited by the peptide array technologies available for applications, few attempts have been made to construct qualitative or quantitative models that depict the peptide sequence:binding strength relationships in peptide microarray-based binding studies. There has been a long history of similar modeling efforts based on low-throughput binding data in the areas of T-cell epitope screening and kinase substrate mapping, however. The keen needs in peptide array applications and the success of the modeling efforts in related fields have prompted us to develop SVM-PEPARRAY, a Web-based program capable of constructing qualitative and quantitative models based on peptide microarray binding datasets using support vector machine (SVM) modeling methods. We expect that such modeling analysis will allow researchers to quickly extract sequence-based biological information from improved peptide array binding results and provide more precise and accurate information about the biological systems investigated.

Key words: SVM-PEPARRAY, on-chip binding experiment, SVM, SVR, modeling.

1. Introduction

Peptide microarrays (peptide arrays) allow simultaneous interrogation of a large number of peptides for their ability to bind to specific proteins or other macromolecules on a solid surface, and they have been used to investigate a wide range of important biological problems, including antigen–antibody interactions, MHC-peptide binding/T-cell epitope screening, enzyme–substrate interactions (e.g., kinase–substrate interactions and phosphotase–substrate interactions), and protein domain (e.g., WW domain or SH2 domain)-mediated protein–protein interactions

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_23,

(*see* reviews, e.g., 1–3). The central objective of analyzing peptide array-based binding data is to arrive at a *model* that depicts the relationship between a peptide sequence and the outcome of the binding. The model can be descriptive, qualitative, or quantitative. A descriptive model summarizes the common sequence features shared by strong binding peptides, often in the form of *consensus sequences* (4, 5). Descriptive models usually result from manual examination (or "eye-balling") of the binding data. A qualitative model (also termed a classification model) is constructed through training using a binding dataset. Once constructed, it can be used to make predictions about whether a new, untested peptide is a "binder" or a "non-binder". A quantitative model (also termed a regression model) is also constructed through training. Once a quantitative model is constructed, it can be used to predict the binding strength (a real-valued number) of a new, untested peptide.

The construction of a quantitative model requires training using quantitative binding data. Without high-quality quantitative binding data, reliable quantitative models cannot be achieved. Because most existing peptide microarray studies have relied on the SPOT™ peptide synthesis technique – the experiments were performed either directly on the cellulose-based surface where the SPOT™ synthesis took place (6, 7) or on a glass surface where pre-synthesized peptides (by the SPOT™ technique) were spotted to achieve higher density (4, 8) – the binding measurements made in these experiments have been qualitative (or at best, semi-quantitative) in nature, due to the limitations imposed by the peptide synthesis technique. Consequently, most models reported in peptide array studies have been descriptive models (4, 5, 9, 10), although some attempts at constructing PSSM-based qualitative models have been reported (11). Recent developments in light-directed photogenerated acid and in situ peptide synthesis on microchips have resulted in a high-density, quantitative, flexible peptide array platform (2, 12, 13), paving the road for quantitative on-chip peptide-binding studies. Careful modeling using the binding data obtained on the new generation of peptide microchips will reveal more precise and accurate information about the biological systems investigated.

Computational modeling of peptide sequence:binding activity relationships is not a new problem. Taking T-cell epitope screening as an example, the earliest efforts in modeling peptide–MHC binding specificity with low-throughput binding data were dated back to the late 1980s (14). Throughout the past two decades, new modeling methods have continued to be developed (*see* recent reviews, e.g., 15, 16). Two of the latest trends in the field of modeling peptide–MHC binding/T-cell epitopes are machine learning-based modeling methods (17–21) and quantitative

modeling (20, 22–24). Similarly, computational modeling of kinase substrate specificity has also received much attention in recent years, owing to the accumulation of phosphorylation site data curated from the literature (Phospho.ELM (25) and PhosphoSite (26)). As is true for the field of peptide–MHC binding modeling, machine learning methods are prevailing among the recently developed kinase substrate modeling methods, and many of these methods demonstrated impressive modeling performance (27–30).

Support vector machines (SVMs) are the most commonly applied machine learning techniques in these new modeling methods, owing to their solid theoretical foundation and excellent empirical performance. SVMs can be used to construct both qualitative models (classification models) (17, 29) and quantitative models (regression models) (20, 24) for peptide-binding systems. Here we describe SVM-PEPARRAY, a Web-based program developed for constructing SVM classification and regression models using peptide microarray datasets. SVM-PEPARRAY is a versatile, user-friendly program with various options for model constructing and testing, and it is designed to be used with ease by users with no programming experience.

2. SVM-PEPARRAY Overview

SVMs are a class of supervised learning methods motivated by Vapnik and Chervonenkis' Statistical Learning Theory (31). They are developed with several distinguished properties that make them very appealing modeling methods. These properties include the following: (1) by defining a regularization term inversely related to the geometric margin of the fitting in the loss function, SVMs attempt to minimize the fitting error and maximize the margin at the same time, making them a class of "maximum margin" methods, effective in curbing the overfitting problem; (2) by considering the Lagrange Dual function of the optimization problem and taking advantage of the quadratic programming (QP) techniques, SVMs efficiently solve the optimization problem and the resulting model involves only a subset of the training samples (the "support vector"); and (3) by applying the "kernel trick," SVMs implicitly map input attributes to a high-dimensional feature space, allowing accurate non-linear models to be constructed as efficiently just as linear models. SVMs were originally developed to solve classification (or qualitative modeling) problems. With the introduction of an ϵ-insensitive loss function, they have been extended to solve regression (or quantitative modeling) problems.

SVM-PEPARRAY is a Web-based program that constructs qualitative (SVM classification) and quantitative (SVM regression, or SVR) models with user-provided peptide microarray data. The dataset provided should be a list of *peptide sequence*:*binding result* pairs, where the *binding result* should be binary values (1 representing a binder and 0 representing a non-binder) for constructing an SVM classification model, and real-valued numbers (representing binding intensities obtained from an on-chip binding experiment) for constructing an SVR model. Three kernel options are provided – the linear kernel, the polynomial kernel, and the radial basis function (RBF) kernel (**Table 23.1**). Two different peptide encoding schemes are implemented. One of them is the commonly used "sparse encoding" scheme (17, 32), and the other is called the "10-factor encoding" scheme, where 10 orthogonal factors obtained from 188 physical properties of the 20 amino acids are used to encode the sequences (33). *N*-fold cross-validation (where *N* is specified by the user) and leave-one-out (LOO) cross-validation can be performed when the performance of the models constructed is assessed.

Table 23.1
The three kernel options implemented in SVM-PepChip

Kernel name	Kernel function
Linear kernel	$K(x, y) = x \times y$
Polynomial kernel	$K(x, y) = (x \times y + 1)^n$
RBF kernel	$K(x, y) = e^{-\gamma\|x-y\|^2}$

SVM-PEPARRAY is accessible at http://pepcyber.umn.edu/SVM-PEPARRAY with a Web browser (Internet Explorer V5.5 or higher, Firefox V 1.5 or higher). Registration is required (with a valid email address) before the user is able to use the program.

3. Methods

3.1. Establishing a Qualitative or Quantitative Model

3.1.1. Selecting or Uploading a Microarray Dataset

After logging in, the user should select the option *Construct a model*. A list of all peptide microarray datasets previously uploaded by the user is displayed. The user can select a dataset from the list. Alternatively, the user can choose to provide a new dataset. In the latter case, the user is prompted to provide a name and (optionally) a textual description of the dataset. The dataset can either be uploaded as a comma separated value (CSV) file or pasted to the

text area provided. The data file (or pasted text) should contain no header row and should include three columns: (a) *peptide ID* (numerical or textual); (b) *peptide sequence*; and (c) *binding result*. The *peptide sequence*s should be aligned across rows. The character '-' should be used as a space holder for leading empty positions. The *binding result*s can be either binary values (where 1 represents binder and 0 represents non-binder) or real-valued numbers (binding intensities measured from the on-chip experiments). After the user clicks the *Submit* button, the dataset is checked for errors and, if no errors are found, the dataset is accepted and a statistics summary of the dataset is displayed.

3.1.2. Constructing a Qualitative SVM Model

To establish a qualitative SVM model (or classification model), the user should click the link *Construct a SVM classification model*. If the dataset selected by the user is a quantitative dataset (i.e., the *binding result* column of the dataset contains real-valued binding intensities), the user would be prompted to choose a cut-off value which would be used to discretize the data: the *binding result* values greater than or equal to the cut-off value would be converted to 1 representing binders, and the values less than the cut-off value will be converted to 0 representing non-binders. After data discretization, the user would be prompted to save the dataset before proceeding to the next step.

In the *Configure SVM classification model construction* page that follows, the user is prompted to choose one of the three kernel options – the linear kernel, the polynomial kernel, and the RBF kernel, and specify how cross-validation should be done – either an N-fold cross-validation (where N is specified by the user) or an LOO cross-validation can be conducted. In addition, the user should choose one of the two peptide encoding schemes – "sparse encoding" or "10-factor encoding". After these selections are made, the user clicks the *Proceed* button to go to the next step.

In the *Choose SVM classification model parameters* page, the user chooses the non-kernel and kernel parameters for the model construction. There is only one non-kernel parameter – the regularization parameter C – that needs to be specified for constructing an SVM classification model. If the user has chosen the polynomial kernel option, then n – the degree in the polynomial function needs to be specified. If the user has chosen the RBF kernel, then the kernel parameter γ would need to be specified. SVM-PEPARRAY would examine the dataset and provide a list of recommended values for each non-kernel or kernel parameter. The user can choose to alter one or more parameter values, or, if deemed proper, specify a completely different list of parameter values. In SVM-PEPARRAY, a grid-search is performed to find the parameter combination giving rise to the model with the best performance (evaluated under cross-validation). For example, if the user chooses the RBF kernel option and specifies four different

values of C and five different values of γ, then 20 different SVM models – each constructed with one of the 20 ($=4 \times 5$) different parameter combinations – will be tested. The model that achieves the highest cross-validated accuracy would be selected and presented to the user.

After specifying the model parameter values, the user clicks the *Start constructing model* button to initiate the model construction process. The user will receive an automatically generated notification email after the model construction is completed.

3.1.3. Constructing a Quantitative SVM Model

To construct a quantitative SVM model (or a SVR model), a quantitative dataset (for which the *Binding result* column of the dataset contains real-valued binding intensities) must be selected. After the dataset is selected, the *Configure SVR model construction* page is displayed, where the user specifies the kernel selection, cross-validation option and peptide encoding scheme.

At the next step, the *Choose SVR model parameters* page, the user specifies non-kernel and kernel parameter values. SVR models involve two non-kernel parameters: the regularization parameter C, and the parameter ϵ in the ϵ-insensitive loss function. The kernel parameters involved in SVR model construction are the same as those involved in the construction of SVM classification models. As in SVM classification model construction, SVM-PEPARRAY examines the dataset and provides a list of recommended values for each parameter. These values can be altered by the user as deemed fit. After the model construction job is initiated, a grid-search would be conducted to find the parameter combination that gives rise to the model with the best performance (evaluated according to the cross-validated R^2 value). After the model construction is completed and the best model is determined, a notification e-mail is automatically generated and sent to the user.

3.2. Examining a Newly Constructed Qualitative or Quantitative Model

After receiving an e-mail notification that a model construction job is completed, the user can log into the SVM-PEPARRAY system, choose the option *View recently constructed models* and select the model from the list. A comprehensive summary of the model is then displayed, which includes the model construction configurations, non-kernel and kernel parameters chosen, the performance assessment of the model, the original dataset used in the training of the model and the predictions made for each peptide in the original dataset under cross-validation. If the user is satisfied with the model, he could choose to store the model in the SVM-PEPARRAY system. If the model is not satisfactory, the user could discard it and try to construct a new model with modified configurations.

3.3. Making Predictions Using an Established Qualitative or Quantitative Model

When the user chooses the option *Make predictions using a stored model*, a list of models previously stored by the user is displayed. After selecting the model to use, the user is prompted to provide a list of peptides, either by uploading a CSV file or by pasting the peptide list into the text area displayed. The CSV file or the pasted text should contain no header and should include two columns: (a) *peptide ID* (numerical or textual) and (b) *peptide sequence*. After the user clicks the *Make predictions* button, the predicted binding results will be displayed to the user with a brief delay. If the model selected by the user is an SVM classification model, qualitative predictions would be made. Otherwise, if the selected model is an SVR model, then quantitative predictions would be made.

4. Notes

1. The performance of an SVM model constructed using a peptide array dataset is dependent on the interplay of a variety of factors, including the quality of the binding data (which in turn is determined by the peptide synthesis and arraying techniques, the quality of the binding experiment and the intrinsic affinity or specificity of the peptide–protein binding system), the size of training dataset, and the similarity between the peptide sequences in the training samples. Although it is possible to construct an SVM model with a training dataset of size $\sim$30, a larger dataset (with $>$100 samples) is often required to achieve a model with acceptable performance. The level of sequence similarity between the peptides in the training dataset influences the constructed model in a less straightforward manner: a high level of sequence similarity in the training dataset will result in a model with good performance when tested under cross-validation setting, but the performance may not extend to untested peptides whose sequences are very different from the ones used during the training of the model.
2. Our experience is, for most datasets, the models constructed using the RBF kernel demonstrated most satisfactory performance. However, there are cases where models constructed with the polynomial kernel or the linear kernel achieved better performance. It is, therefore, advisable that the user tries out different kernel options when constructing models. Similarly, although the "10-factor encoding" scheme often yields models with better performance than the "sparse encoding" scheme, it is sensible for the user to try out both encoding schemes. In theory, the way cross-validation is conducted does not have direct influence on the performance of the

model, but it influences how accurately the model performance is assessed. A larger N (in N-fold cross-validation) would lead to more accurate assessment of the model performance. The downside of choosing a larger N is that it would result in higher running time cost for model construction. Generally, 5- or 10-fold cross-validation suffices for most model construction tasks.

Acknowledgments

This work was supported in part by NIH/NCI (1R21CA126209 to XG and TL), Minnesota Medical Foundation (to TL) and the R. A. Welch Foundation (E-1270, to XG) and NIH/GM/AI (R43 GM076941 to XG and XZ). We thank Dr. K. Cassidy for critically reading the manuscript.

References

1. Reimer, U., Reineke, U. and Schneider-Mergener, J. (2002) Peptide arrays: from macro to micro. *Curr Opin Biotechnol*, **13**, 315–320.
2. Gao, X., Pellois, J.P., Na, Y., Kim, Y., Gulari, E. and Zhou, X. (2004) High density peptide microarrays. In situ synthesis and applications. *Mol Divers*, **8**, 177–187.
3. Panicker, R.C., Huang, X. and Yao, S.Q. (2004) Recent advances in peptide-based microarray technologies. *Comb Chem High Throughput Screen*, **7**, 547–556.
4. Schutkowski, M., Reimer, U., Panse, S., Dong, L.Y., Lizcano, J.M., Alessi, D.R. and Schneider-Mergener, J. (2004) High-content peptide microarrays for deciphering kinase specificity and biology. *Angewandte Chemie-International Edition*, **43**, 2671–2674.
5. Diks, S.H., Kok, K., O'Toole, T., Hommes, D.W., van Dijken, P., Joore, J. and Peppelenbosch, M.P. (2004) Kinome profiling for studying lipopolysaccharide signal transduction in human peripheral blood mononuclear cells. *J Biol Chem*, **279**, 49206–49213.
6. Frank, R. (1992) Spot-Synthesis – an Easy Technique for the Positionally Addressable, Parallel Chemical Synthesis on a Membrane Support. *Tetrahedron*, **48**, 9217–9232.
7. Frank, R. (2002) The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports – principles and applications. *J Immunol Methods*, **267**, 13–26.
8. Nahtman, T., Jernberg, A., Mahdavifar, S., Zerweck, J., Schutkowski, M., Maeurer, M. and Reilly, M. (2007) Validation of peptide epitope microarray experiments and extraction of quality data. *J Immunol Methods*, **328**, 1–13.
9. Murray, E., McKenna, E.O., Burch, L.R., Dillon, J., Langridge-Smith, P., Kolch, W., Pitt, A. and Hupp, T.R. (2007) Microarray-formatted clinical biomarker assay development using peptide aptamers to anterior gradient-2. *Biochemistry*, **46**, 13742–13751.
10. Landgraf, C., Panni, S., Montecchi-Palazzi, L., Castagnoli, L., Schneider-Mergener, J., Volkmer-Engert, R. and Cesareni, G. (2004) Protein interaction networks by proteome peptide scanning. *PLoS Biol*, **2**, E14.
11. Rychlewski, L., Kschischo, M., Dong, L., Schutkowski, M. and Reimer, U. (2004) Target specificity analysis of the Abl kinase using peptide microarray data. *J Mol Biol*, **336**, 307–311.
12. Pellois, J.P., Zhou, X., Srivannavit, O., Zhou, T., Gulari, E. and Gao, X. (2002) Individually addressable parallel peptide synthesis on microchips. *Nat Biotechnol*, **20**, 922–926.
13. Hong, A., Zhu, Q., Na, Y., Qi, Y., Drabek, R., Yao, W., Srivannavit, O., Gulari, E., Zhang, X., Zhou, X. et al. (2005) In Gierasch, L. M. (ed.), *The 19th American Peptide Society Symposium*. Wiley Interscience, San Diego, CA, pp. 498.

14. Sette, A., Buus, S., Appella, E., Smith, J.A., Chesnut, R., Miles, C., Colon, S.M. and Grey, H.M. (1989) Prediction of major histocompatibility complex binding regions of protein antigens by sequence pattern analysis. *Proc Natl Acad Sci USA*, **86**, 3296–3300.
15. Flower, D.R., Doytchinova, I.A., Paine, K., P., T., Blythe, M.J., Lamponi, D., Zygouri, C., Guan, P., McSparron, H. and H., K. (2002) In Flower, D. R. (ed.), *Drug Design: Cutting Edge Approaches.* Royal Society of Chemisty, Cambridge, pp. 136–180.
16. Evans, M.C. (2008) Recent advances in immunoinformatics: Application of in silico tools to drug development. *Curr Opin Drug Discov Devel*, **11**, 233–241.
17. Donnes, P. and Elofsson, A. (2002) Prediction of MHC class I binding peptides, using SVMHC. *BMC Bioinformatics*, **3**, 25.
18. Bhasin, M. and Raghava, G.P. (2004) SVM based method for predicting HLA-DRB1*0401 binding peptides in an antigen sequence. *Bioinformatics*, **20**, 421–423.
19. Majeux, N., Udaka, K. and Mamitsuka, H. (2003) Prediction of MHC class I binding peptides using an ensemble learning approach. *Genome Informatics*, **14**, 687–688.
20. Liu, W., Meng, X., Xu, Q., Flower, D.R. and Li, T. (2006) Quantitative prediction of mouse class I MHC peptide binding affinity using support vector machine regression (SVR) models. *BMC Bioinformatics*, **7**, 182.
21. Yu, K., Petrovsky, N., Schonbach, C., Koh, J.Y. and Brusic, V. (2002) Methods for prediction of peptide binding to MHC molecules: a comparative study. *Mol Med*, **8**, 137–148.
22. Doytchinova, I.A., Guan, P. and Flower, D.R. (2004) Quantitative structure–activity relationships and the prediction of MHC supermotifs. *Methods*, **34**, 444–453.
23. Doytchinova, I.A. and Flower, D.R. (2002) Quantitative approaches to computational vaccinology. *Immunol Cell Biol*, **80**, 270–279.
24. Wan, J., Liu, W., Xu, Q., Ren, Y., Flower, D.R. and Li, T. (2006) SVRMHC prediction server for MHC-binding peptides. *BMC Bioinformatics*, **7**, 463.
25. Diella, F., Cameron, S., Gemund, C., Linding, R., Via, A., Kuster, B., Sicheritz-Ponten, T., Blom, N. and Gibson, T.J. (2004) Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinformatics*, **5**, 79.
26. Hornbeck, P.V., Chabra, I., Kornhauser, J.M., Skrzypek, E. and Zhang, B. (2004) PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics*, **4**, 1551–1561.
27. Berry, E.A., Dalby, A.R. and Yang, Z.R. (2004) Reduced bio basis function neural network for identification of protein phosphorylation sites: comparison with pattern recognition algorithms. *Comput Biol Chem*, **28**, 75–85.
28. Senawongse, P., Dalby, A.R. and Yang, Z.R. (2005) Predicting the phosphorylation sites using hidden Markov models and machine learning methods. *J Chem Inf Model*, **45**, 1147–1152.
29. Kim, J.H., Lee, J., Oh, B., Kimm, K. and Koh, I. (2004) Prediction of phosphorylation sites using SVMs. *Bioinformatics*, **20**, 3179–3184.
30. Huang, H.D., Lee, T.Y., Tzeng, S.W. and Horng, J.T. (2005) KinasePhos: a web tool for identifying protein kinase-specific phosphorylation sites. *Nucleic Acids Res*, **33**, W226–229.
31. Vapnik, V. (1998) *Statistical Learning Theory.* John Wiley & Sons, New York.
32. Baldi, P. and Brunak, S. (2001) *Bioinformatics: the machine learning approach.* The MIT Press, Cambridge, MA.
33. Zhao, Y., Pinilla, C., Valmori, D., Martin, R. and Simon, R. (2003) Application of support vector machines for T-cell epitopes prediction. *Bioinformatics*, **19**, 1978–1984.

Chapter 24

PASE: A Web-Based Platform for Peptide/Protein Microarray Experiments

Fabien Pamelard, Gael Even, Costin Apostol, Cristian Preda, Clarisse Dhaenens, Vronique Fafeur, Rémi Desmet, and Oleg Melnyk

Abstract

Peptide microarray technology requires bioinformatics and statistical tools to manage, store, and analyze the large amount of data produced. To address these needs, we developed a system called protein array software environment (PASE) that provides an integrated framework to manage and analyze microarray information from polypeptide chip technologies.

Key words: Proteomics, peptide, protein, molecular interaction studies, database management, Web platform and statistical analyses.

1. Introduction

Peptide microarrays are high-throughput screening systems which generate a huge amount of data in a short period of time. This technology thus requires bioinformatics and statistical tools to manage, store, and analyze the large amount of data produced. These data include a description of reporter molecules, of the physical array, and of the samples studied, experimental procedures, microarray images, and their extracted data matrices. In the field of DNA microarrays, an attempt has been made to organize such data through the definition of the standard MIAME (minimum information about a microarray experiment) that describes the minimum information required to ensure that microarray data can be easily interpreted and that results derived from its analysis can be independently verified (1–2). The ultimate goal of this proposition is to establish a standard for recording and reporting microarray-based gene expression data, which will in turn allow the establishment of

Marina Cretich, Marcella Chiari (eds.), *Peptide Microarrays,* Methods in Molecular Biology 570,
DOI 10.1007/978-1-60327-394-7_24,

databases and public repositories and enable the development of data analysis tools. Following this proposition several MIAME-compliant softwares have been proposed in order to store DNA microarray experiments, such as MIAMExpress, LIMaS, SAS microarray to cite only a few (*see* **Note 1**). BASE (BioArray Software Environment) is one of these softwares (3). This Web-based open source MIAME-supportive microarray database and analysis platform is now widely used in laboratories involved in DNA microarray projects. A similar story now happens with polypeptide chips.

Since few years, peptide or protein microarrays are increasingly used for high-throughput study of biomolecular interactions. Using protein chips, we can now investigate interactions between proteins on a large scale, even to the point of examining all proteins that are expressed in a given cell or tissue. Revealed information about protein–protein interactions could provide targets for designing new therapeutic strategies. In this technology, the analytical goal and the large data scale represent a new challenge.

To address these needs in comparison to the DNA chip software knowledge, we developed a system called protein array software environment (PASE) (4–5) that provides an integrated framework to manage and analyze microarray information from polypeptide chip technologies.

PASE is a free software that may be loaded and used by any laboratory. It is protected by the CeCILL [Ce(A)C(nrs)I(NRIA)-L(ogiciel)L(ibre)] license. CeCILL is the first license defining the principles of use and dissemination of free software in conformance with French laws, following the principles of the GNU GPL. Hence, PASE software was developed on the GNU/Linux operating system (OS) in the PHP language, with data being stored in a relational database (ProgreSQL/MySQL) and communicated to the user through the Apache Web server.

2. Materials: PASE Minimum Requirements

2.1. Web Server

Minimum configuration for the Web server:

1. Intel Pentium II 500 MHz or equivalent (AMD, etc.) CPU
2. 256 MB RAM
3. 4 GB of hard-disk
4. Network card 10/100 Mbits
5. Linux OS (any releases), Kernel 2.4.4 or later
6. MySQL, 3.23.4 × or later
7. PHP, 4.2.0 or later

8. Apache 1.3.24 or later*Recommended configuration for Web server:*
9. AMD Dual Athlon 1.2 GHz CPUs or equivalent
10. 1 GB RAM
11. Network card 10/100
12. Linux OS (Debian or Fedora release), Kernel 2.4.10 or later
13. MySQL 3.23.51 or later
14. PHP 4.2.1 or later
15. Apache 1.3.24 or later

2.2. Personal Computer

Recommended configuration for user:

- Windows XP/2000/NT/Me/98, Macintosh or Linux PC
- Network card 10/100
- Web Browser (Mozilla Firefox, Internet Explorer, Safari, etc.)
- Java Web Start and Java 2 Runtime Environment

3. Methods

The generation and the analysis of proteomics data are now widespread and high-throughput approaches are common place; the need for a dedicated platform to store and analyze data is crucial. It is important to note that DNA microarray platforms cannot be used for managing polypeptide microarray data. Indeed, these news chips bring new managing and analysis constraints which do not exist in the DNA chip area because the type of biomolecular interactions and parameters studied, the protocols as well as the type of information extracted from the microarray experiment are very different from standard DNA microarray experiments (6).

To this end we developed PASE (protein array software environment) (4–5) in an approach similar to MIAME. An effort has been made in order to assure that all information required to ensure that polypeptide microarray data can be easily interpreted and that results derived from its analysis can be independently verified will be stored. PASE is a Web-based core developed from BASE (BioArray Software Environment) 1.2.16 for keeping a database organization similar to the MIAME recommendations extended to the polypeptide case. PASE is a new database and analysis platform designed to be installed in any microarray laboratory which uses polypeptide chip technologies through its Web browser.

Let us note that this problem has been seen as a challenging one, since a new standard, dedicated to proteomics experiments, MIAPE (minimum information about a proteomics experiment) has been proposed by the international community in 2007 (7). This standard gives recommendations about the minimum reporting requirements for proteomics experiments. This is what is done in PASE.

The PASE platform such as the BASE system integrates an array LIMS (laboratory information management system), biomaterial information, raw images, and extracted data, and provides a plug-in architecture for polypeptide chip analysis, data viewing, and statistical modules. Because polypeptide or biopolymer chips could contain various probes type, we developed parts to manage various information fields for each probe type. The structure of this new platform was designed to follow the natural workflow of the polypeptide chip technology (from sample to scanning steps).

Data can be exported in a multitude of formats for local analysis and publication. PASE is a complete bioinformatics tool to manage, store, and analyze data according to the polypeptide chip environment.

3.1. Installation and First Use of PASE

PASE requires a Web server such as Apache, the interpreter for the language PHP, and an SQL relational database management system such as MySQL or PostgreSQL (*see* **Note 2**).

1. *Setup required softwar*e

   *MySQL:*you need to create two databases, typically called "pase" and "pasedynamic," and create tables in the pase database by executing the file sql/pase_mysql.sql from the mysql command line client (using the\. command). You must also execute sql/pase_mysql_data.sql to create the PASE super user and some other important data.

   *Web server Apache:*how to configure Apache is largely a matter of taste. We suggest that you create a Unix user called "base" and let Apache run as that user (*see* **Note 3**).

2. *Setup, compile, and install PASE*
   (a) Run the configure script on the PASE folder: ./configure –prefix=/path/to/pase, where the prefix determines where you wish PASE to be installed (*see* **Note 4**). PASE platform does not need compilation of plug-ins. So use, –disable-plug-ins. There are also options for turning off the compilation of the plotapplet, and for activating the jobRunner feature.

   Minimum parameters for PASE configuration are: –prefix=/path/to/pase –disable-plug-ins –with-base-user=pase (*see* **Note 5**).

(b) Compile the different elements with the command "*make*"and install everything into the prefix directory "*make install.*" *"Make install"* will create the PASE data directory and set user and permissions on it. If a data directory already exists the permissions might be changed, but the contents will not be touched. This will create two directories /path/to/pase/**plug-ins/** and /path/to/pase/**plug-ins/bin**.

3. *PASE configuration*
   (a) The PASE configuration file *config.inc.php* will be created when you do a "*make install*" (*see* **Note 6**). The new config file will have the right owner and permissions set. Some configuration options (data dir location, Unix user name, etc.) will be filled in from the information given to configure.
   (b) Edit config.inc.php to match your setup (*see* **Note 7**). Also take a look at the mime.types options, which should point to the file mime.types which can be found in Apache's conf directory. This is needed to get the right content-type when viewing uploaded files. If your MySQL was complied with – enable-local-infile you may set mysql.localinfile to 1 for better performance (when adding data to PASE).

The information sent to configure will be used to create www/getconfig.inc.php, which is needed for the Web part of PASE to find the rest of the files.

4. *Starting and stopping PASE*

   Start and stop PASE using the bash script bin/base.server. This script will have the path to your PASE installation set by make, so there is no need to edit it. The simplest way to use this script is to copy it to/etc/init.d (or/etc/rc.d), and make sure there are symlinks to it in rc*.d. (*see* **Note 8**). In any case, you should verify that PASE is really started after MySQL and Apache, and stopped before they are shut down.

   The base.server script uses two PHP scripts in bin, and you may call these directly: startBase.php and stopBase.php. These scripts can be run either by the PASE Unix user or by root. stopBase.php takes one optional argument; the message shown to users while PASE is down. If you are starting and stopping PASE from the command line you should use startBase.php and stopBase.php, since they will provide slightly more output than base.server, and you will be able to specify a shutdown message.

5. *Starting PASE for the first time*

   At this point PASE should be working. Make sure that MySQL and Apache are running, and then run bin/base.server start (or bin/startBase.php). Point your browser to your Web server, and you should see pretty much the same interface you see at http://www.genopole-lille.fr/gdlarray/pase-demo/www/index.phtml. You can now log in as the PASE user "root" with no password. The very first thing you should do is to go to the Users menu and set the root password to something safer.

6. *Logging in PASE*

   To log in PASE, you need to get from an administrator of the PASE server:

   a. The address (URL) to a PASE server

   b. A username to login with

   c. A password

   Logging in is simple; enter your login and password in the form on the front page and click the *Login button*.Once you have been logged in, the home page will be displayed. It displays some useful information.

7. *Additional information*

   Documentation provided: the PASE project description site is located at http://www.genopole-lille.fr/spip/spip.php?article30 (*see* **Note 9**).

   Demonstration server: to test PASE without installing it, you can use the demonstration server http://www.genopole-lille.fr/gdlarray/pase-demo/www/index.phtml (*see* **Note 10**).

3.2. Describing Reporters

1. Describe all the reporters (probes that sit on the array and that specifically bind your labeled target during the biomolecular interaction) used during experiments (*see* **Note 11**). Reporters must have a reporter ID and this ID has to be unique.
2. *Validation:* on the "Reporters" page (Reporters/Reporters) all reporters created in your PASE can be displayed (*see* **Note 12**).

3.3. Describing Additional Experiments Information (Array LIMS)

This part of the database is optional. The objective is to track probe information (*see* **Note 13**) in order to allow efficient analysis of the polypeptide chip data. LIMS features are accessed from submenu options under main menu option Array LIMS.

1. *Plate types*

   First define the types of plates that will be used (*see* **Note 14**).

2. *Plates*

   Once you have created plate types and reporters types you can upload plates to PASE from file (*see* **Note 15**). Once you have uploaded the first set of plates to PASE you can list the plates by selecting PLATES from the main menu (Array LIMS/Plates).

   In this plate section, you have the possibility to export your plate design into two formats. The first format is a summary of the information you entered when you upload the plate. The second format is compatible with Qarray spotters (Genetix). This export is possible by clicking on "Export marked plates" or "Export all matching plates."

3. *Array design*

   This part is specific to information about array design (i.e., physical position of the reporters on the array). An array design is defined with the following information: a name, some manipulation protocols, comments, and the plates name associated with the array. An array design needs an array batches (*see* **Note 16**).

4. *Array prints/batch*

   PASE offers the possibility to manage the spotting events. A spotting event has several data: a name, an owner, the quantity of chips produced, the number of chips used/deleted/available.

3.4. Describing Biomaterials

Biomaterials comprise samples, extracts, and labeled extracts in PASE (*see* **Fig. 24.1**). The workflow allows users to enter sample information in PASE; samples can then be aliquoted and the extracts can be labeled and used in incubations. All along, information used can be recorded.

1. *Sample origins*

   Sample origins are used to create ontologies of sample sources. This organization can be described in many ways: by species, by tissue, etc.

2. *Sample annotations*

   Sample annotations are used to create annotation types that subsequently can be used by users to annotate their entered samples. Give your new annotation type a name, type, size of the input field box, and default value (*see* **Note 17**).

3. *Protocols*

   Enter standard protocols for the aliquots preparation, target preparation, and revelation.

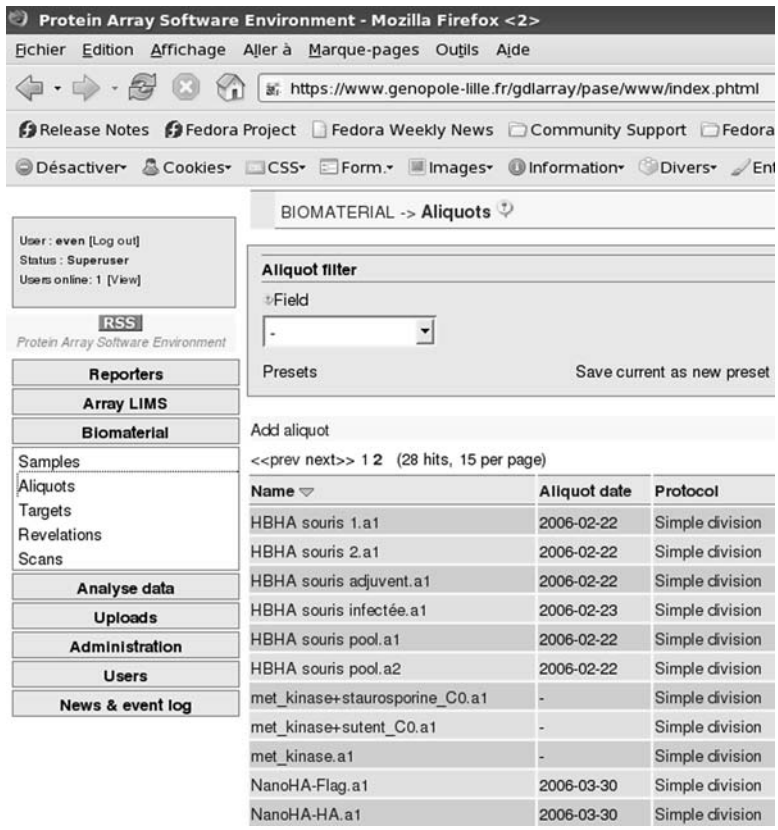


Fig. 24.1. PASE, Biomaterial: information for biological samples use in polypeptide chip experiments.

4. *Samples*

   Samples are the starting point of all data analysis in PASE. From the "Samples" page (Biomaterials/Samples), you may enter new samples, list all samples you have access to and annotate samples using available annotation types. Once you insert a new sample, if you choose "Accept and add aliquot," you will have access to the aliquot form page of your current sample. After filling the requested information for aliquot, you can click "Accept and target" to complete the information about target (*see* **Note 18**).

5. *Aliquots*

   From the "Aliquots" page (Biomaterials/Aliquots), you may enter, edit, and add new aliquots. The forms are similar to those in the sample section. You will have to choose a quantity of aliquot. This quantity allows you to check for the quantity of aliquot available during the targeting step. If the remaining quantity is null, this aliquot will not be proposed for targeting anymore.

6. *Targets*

   From the "Targets" page (Biomaterials/Targets), you may enter, edit, and add new targets. The forms are similar to those in the sample section. The quantity of a target is used in the next step revelations.

7. *Revelations*

   Create a new revelation, making the association of one or two targets to revelation and to an array (if the optional ARRAY LIMS feature is being used). Enter the revelation name, select the target(s) and the quantity used, the incubation protocol (these may be defined from the PROTOCOLS submenu), free text description or any notes, and the revelation date. Selecting 'YES' for 'Use array production end' will allow you to 'PICK' an array to associate with this revelation from all available arrays in the database. If you select 'NO,' this revelation will not be linked to any array production LIMS information (it is possible to change this association at a later time). Clicking on 'ACCEPT' will save this new revelation. Clicking on 'ACCEPT AND ADD IMAGE SET' will bring you to another Webform to create a new image set for this experiment.

3.5. Analyze Data

From this menu, you can manage your experiments in the submenu 'experiments.' If you have already inserted an experiment you can use 'Experiment Explorer,' a powerful tool of visualization of your experiment data and perform a data analysis with the plug-in architecture provided in PASE.

1. *Raw data sets*

   A Raw data set contains selected data from a results file. The results file is acquired when extracting numeric data from a polypeptide chip tiff image and is typically a tab-delimited text file. When this file is added to an image set a Result file format is used to select which data to import to PASE and how it should be imported. A Raw data set is automatically created each time you add a results file to the image set of revelation and once the Raw data set is created in PASE it will constitute a representation of the raw data from a revelation. Subsequently, when data is analyzed in PASE you start by selecting one or more raw data sets in the Analyze data section and then proceed to work with the selected data.

2. *Experiments*

   As in BASE, most of the data analysis in PASE is performed in the Experiments. An Experiment in PASE is merely a collection of Raw data sets and any associated analysis steps that has been performed on these raw data sets. Raw data sets can be sorted into any number of Experiments and a user can

create any number of Experiments. To enter experiments, go to the 'Experiments' page (Analyze data/Experiments) and select an available Experiment. The first step you have created an experiment is to associate Raw data set with this experiment. When analyzing data in PASE, you work with 'BioAssays' rather than 'Raw data sets' (*see* **Fig. 24.2**).

You will create a BioAssaySet which is an association of many BioAssays.

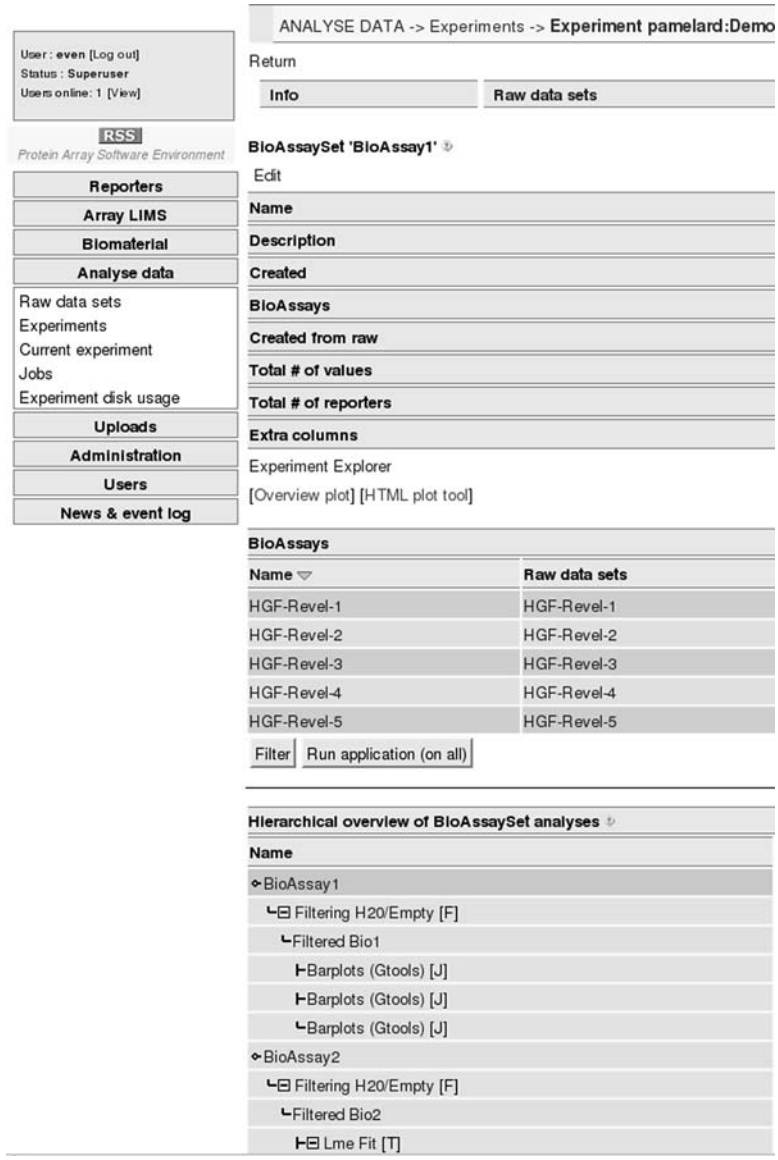


Fig. 24.2. Experiment view for analyzing data in PASE.

To create a BioAssaySet, go to the 'Experiment' in which you want to perform the operation and select the Raw data sets tab. Start by selecting one or more raw data sets (Fig. **24.2** yellow arrow). Enter a name for the BioAssaySet you are about to create and validate. When the BioAssaySet is created, BASE will process one Raw data set at the time and calculate the intensities according to your selected preferences. Each selected Raw data set will be transformed into a BioAssay with the same name as the corresponding Raw data set and the BioAssays will all be included in the same BioAssaySet. The BioAssaySet will appear automatically as an item in the experiment tab 'Analysis steps.' In this tab, you run jobs on BioAssaySets and uniformly apply transformations on all BioAssays in that BioAssaySet:

From the Hierarchical overview of BioAssaySet analyses you can filter data, run analysis applications (plug-ins) as well as visualize your data. PASE is created with a plug-in architecture to facilitate the addition of new analysis tools as they are developed. Available plug-ins can be accessed from the Hierarchical overview of BioAssaySet analyses (*see* **Fig. 24.3**).

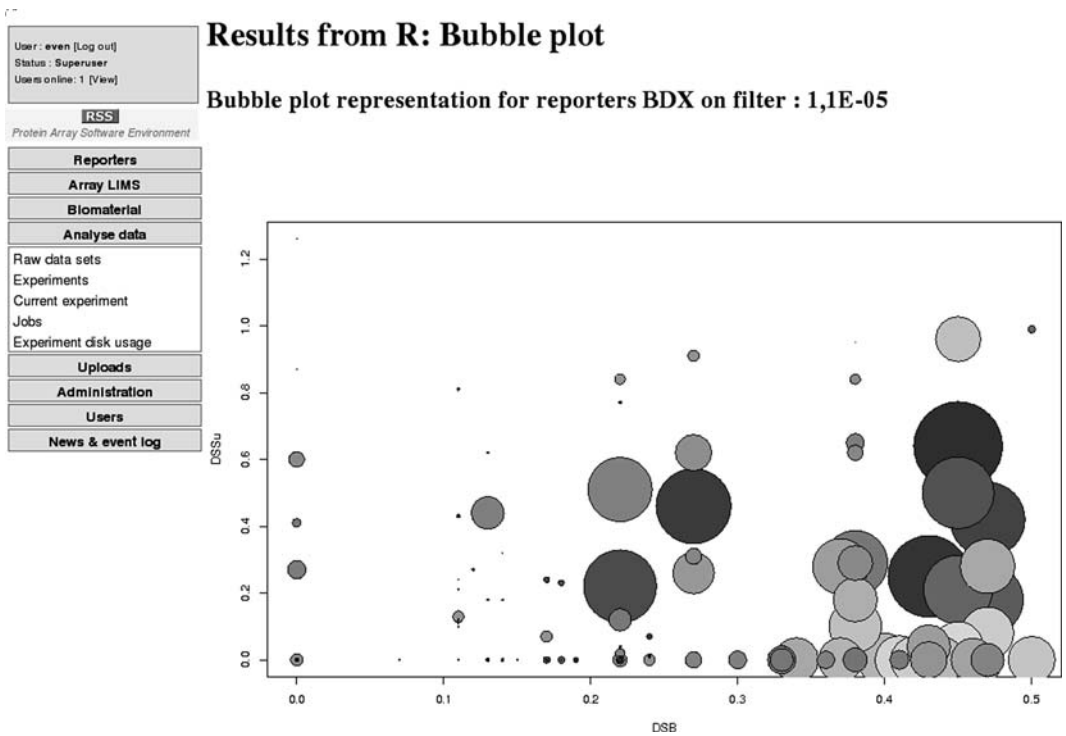


Fig. 24.3. Visualization plug-in for PASE.

3.6. Additional Analysis Tools

We have implemented in this platform special plug-ins for analyzing and viewing data integrated into the Web platform.

3.6.1. Rule Mining

Association rules were first formulated in 1994 by R. Agrawal and R. Srikant (8). An association rule is an expression of the form: IF C THEN P. The 'IF' part is called the rule condition (C) and the 'THEN' part is called the rule prediction (P). In the context of microarray experiments, the rule mining approach allows to:

- Find relations between proteins or peptides based on expression,
- Find relations between proteins in a subgroup of situations,
- Find relations between proteins analyzing comportment and other data (functional or structural characteristics).

Constructing rules is a real combinatorial problem, as a rule consists of the combination of several attributes; hence we adopt a combinatorial optimization approach using a multi-objective evolutionary algorithm for the resolution (9). This rule mining algorithm has been integrated to the plug-in architecture of PASE.

As an input, the complete list of proteins is provided depending on the bioassay creating in PASE. Then the Rule Mining plug-in will create an output file that contains a list of association rules (found by the algorithm) and a HTML file with graphical visualization of the results (*see* **Fig. 24.4**).

We can separate the plug-in execution into two steps:

- Rule mining algorithm (develop in C++): the resulting output is a file with all the significant rules (and their quality criteria) found.
- Visualization tools developed in R language for an easy interpretation of the results: barplot and hitmap of attributes (protein expression), classification of the rules ordered by protein names, etc.

3.6.2. Statistical Analysis

PASE provides advanced statistical methods for protein microarrays. These methods are implemented using the R software in a three-tier-architecture. Besides the univariate descriptive statistics,

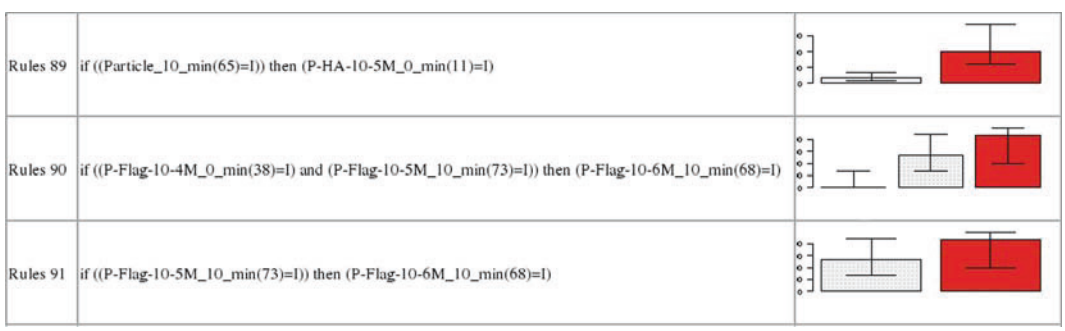


Fig. 24.4. Rule mining plug-in results.

mixed linear models and clustering techniques are implemented for normalizing data and detect clusters of proteins with particular biological properties.

1. *Data normalization*

Data normalization is a key point in this context where data are subject to experimental variations. Hence, we propose for PASE, advanced statistical methods named linear mixed models in order to get data comparable and to separate (estimate) the effects of different sources of variance for the intensity fluorescence (chip support, background noise, protein). Analysis of variance tests associated with linear mixed model provide the significance of these parameters (*see* **Fig. 24.5**).

2. *Classification*

After data normalization step, in order to detect target proteins with particular biological properties, clustering techniques are generally used. The particularity of microarray data is their repeated character and the dependence structure of replications within and between microarrays. The limits of univariate clustering techniques are well known (loss of information, low robustness) and the multivariate approach appears indeed to be more effective. Several approaches are proposed in literature based on

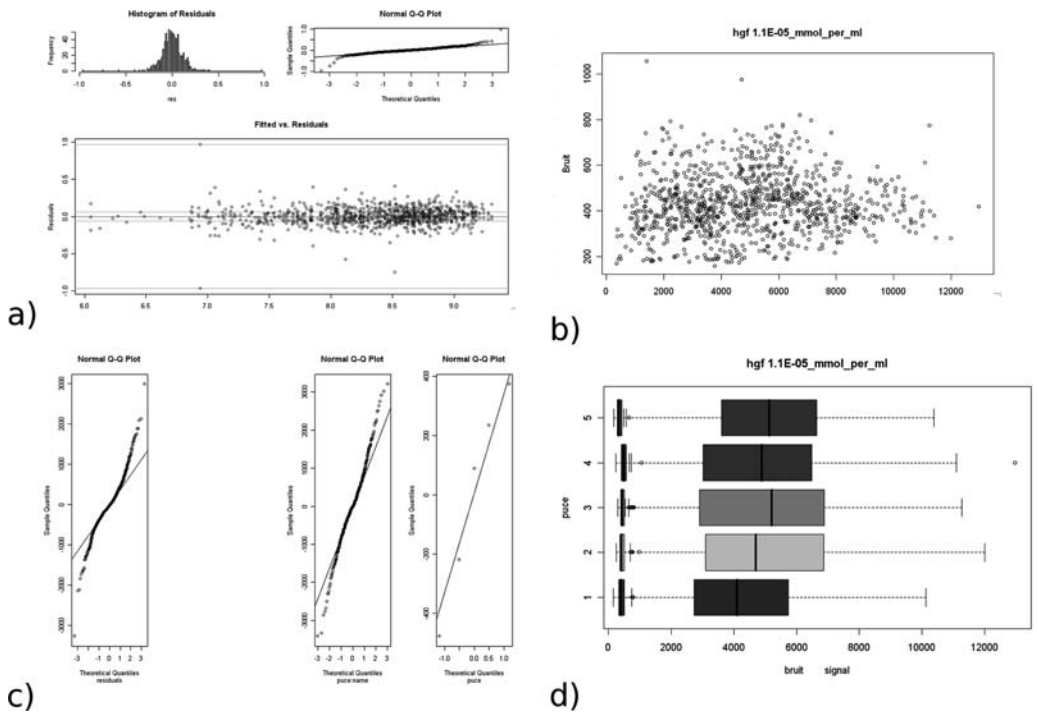


Fig. 24.5. **(a)** Linear mixed model quality; **(b)** background effect; **(c)** chip effect and interaction normality; **(d)** chip and signal/background intensity interaction.

parametrical models such as mixtures model-based clustering (10–11),or the non-parametric methods such as hierarchical classification or the K-means algorithm (12).

Due to the repeated data structure, we propose for PASE the hierarchical classification approach based on the empirical cumulative distribution function (ECDF). It is well known that ECDF describes completely and in a unique way a probability distribution. In addition, ECDF plays an important role in non-parametric statistics due to the robustness of methods dealing with. An example of ECDF associated with a probe p represented by 15 repeated measures of intensity (five microarrays with three replications each) is presented in **Fig. 24.6**.

The Kolmogorov–Smirnov distance (13) appears then as a natural choice for the distance to be used in the hierarchical classification algorithm. At each step of the algorithm, a cluster is represented by an ECDF associated with the protein's measures belonging to the cluster. The number of clusters is automatically computed using a "knee method" (14). This approach is implemented in the PASE application and an example of output is presented in **Fig. 24.7**.

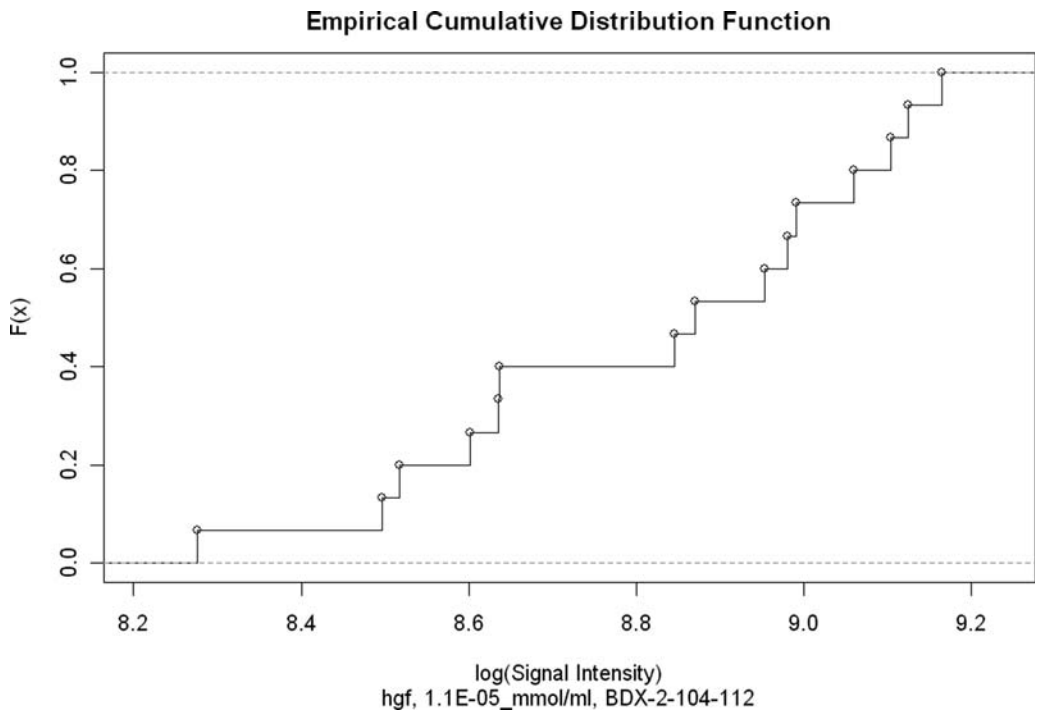


Fig. 24.6. Protein = ECDF.

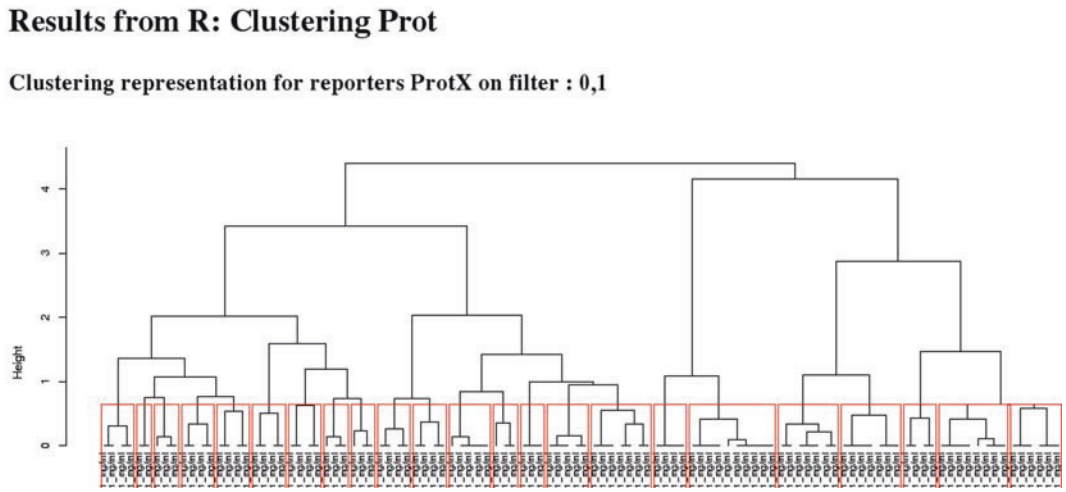


Fig. 24.7. Hierarchical clustering tree.

4. Notes

1. See http://www.mged.org/Workgroups/MIAME/miame.html for a list of MIAME compliant platforms.
2. ASE installation is very similar to the BASE 1.x one. The explanations are a PASE adaptation of the readme installation in the BASE 1.x version.
3. Therefore you should have a line in httpd.conf with 'User pase.' Also setting the group to the primary group of that user is a good idea (e.g., Group pase).
4. The default prefix is /usr/local/base. PASE will install a number of files and directories in the prefix directory, so make sure you put, e.g., /usr/local/base instead of /usr/local. If your PASE Unix user is not called 'pase,' you need to specify –with-base-user=..., and if your PASE data directory is not in PREFIX/data, you should add –with-data-dir=path/to/data.
5. If errors occur during the configuration phase, the file config.log will most likely be useful in tracking down the source of the problems. If you receive the error 'C++ compiler cannot create executables,' the problem could be that you don't have libstdc++ installed.
6. If such a file already exists the new file will be called config.inc.php.new.
7. It should be fairly self-explanatory, and normally you would not need to change much more than the database password and the admin contact info.

8. For SuSE and RedHat it has support for insserv and chkconfig, respectively. On other systems you might have to create the needed symlinks manually.
9. The documentation includes a lot of useful information about the project and the program, like documentation/manuals, download-pages, forum link, contact information, and much more. A forum is provided for PASE users. You will find much information about PASE installation and useful documentation. Follow the link on PASE's home page to the forum or go directly to http://www.genopole-lille.fr/paseforum/index.php.
10. To enter the demonstration, use *pase* as login and *pase* as password. The *pase* user account has read privileges to all data on the demo server.
11. Reporters can be proteins, peptides, or something else depending on the array platform you use.
12. Due to the huge heterogeneity of data in PASE, a system of flexible annotations has been created visible in the 'Reporters type' page. All reporters' type can have a number of annotations or associated information connected to them (e.g., sequence, purification protocol, purity).
13. Such information is location of microtiter plates, microtiter plate identities, their transformations from plate to plate, well annotations (quality control information), etc.. . .
14. For instance, a lab that makes high-speed protein purification may have several types of plates to manage and track: plate before purification, plate after purification, and array print plate. For each plate type, date, comments, quality control steps, sequence verification can be stored. Currently only 96- or 384-well plates can be used. A tool to merge information of 96- or 384-well plates is also available.
15. This file must contain information like: plate ID, ID of row and column of each position of the plate (e.g., A1, B6, H8), reporter ID, and full name of reporter type. When you upload plates containing new reporters, these reporters will be created in PASE.
16. In PASE 1.1, only BioRobotics MicroGrid II TAM file is integrated. A generic tool to Create TAM file has also been developed. The TAM file is needed to get the GAL file (file used in software like Imagene or Genepix). This GAL file is present when you select an array design with an integrated TAM file.
17. The four possible types are number, integer (e.g., culture temperature, cell size), text, or enumeration = group (e.g., protein status positive/negative, treatment a, b, or c).

18. A biological sample can be manipulated before the incubation with the microarray. These manipulations can be followed using PASE.

Acknowledgments

This project was not possible without discussions between biologists and bioinformaticians, many thanks for these interactions. Therefore, we thank CNRS, Pasteur Institute of Lille, University of Lille I and II and gratefully acknowledge the "Réseau National des Génopoles," "Région Nord Pas de Calais," "Feder," and "Ministère de la Recherche" for financial support.

References

1. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J and Vingron M. (2001) Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. *Nat Genet.* **29(4)**, 365–71.
2. www.mged.org/Workgroups/MIAME/miame.html.
3. Lao HS, Carl T, Johan V-C, Sofia G, Åke B and Carsten P. (2002) BioArray Software Environment: A Platform for Comprehensive Management and Analysis of Microarray Data. *Genome Biol.***3(8)**: software0003.1–0003.6.
4. Online informations: http://www.genopole-lille.fr/bioinformatique/PASE. Review: 2006, March, 14.
5. Pamelard F, Laurence P, Desmet R, Dhaenens C, Melnyk O and Talbi E-G (2006) PASE: a webbased platform to manage, store and analyze data coming from polypeptide chips experiments, *JOBIM conf.*
6. Duburcq X, Olivier C, Desmet R, Halasa M, Carion O, Grandidier B, Heim T, Stievenard D, Auriault C and Melnyk O (2004) Polypeptide semicarbazide glass slide microarrays: characterization and comparison with amine slides in serodetection studies on. *Bioconjug Chem.* **15(2)**, 317–25.
7. Taylor CF, Paton NW, Lilley KS, Binz P-A, Julian RlK Jr, Jones AR, Zhu W, Apweiler R, Aebersold R, Deutsch EW, Dunn MJ, Heck AJR, Leitner A, Macht M, Mann M, Martens L, Neubert TA, Patterson SD, Ping P, Seymour SL, Souda P, Tsugita A, Vandekerckhove J, Vondriska TM, Whitelegge JP, Wilkins MR, Xenarios I, Yates JR and Hermjakob H (2007) The minimum information about a proteomics experiment (MIAPE). *Nature Biotechnol.* **25**, 887–893.
8. Agrawal R and Srikant R (1994) Fast algorithms for mining association rules. *In Proceedings of the 20th Intl. Conference on Very Large Databases,Santiago, Chile, Sept 1994.*
9. Khabzaoui M, Dhaenens C, and Talbi E-G (2004) A Multicriteria Genetic Algorithm to analyze DNA microarray data. *In Congress on Evolutionary Computation (CEC), volume II, pages 1874–1881, Portland, USA, IEEE Service Center.*
10. He Y, Pan W and Lin J (2006) Cluster analysis using multivariate normal mixture models to detect differential gene expression with microarray data. *Comput Stat Data Anal.* **51**, 641–658.
11. Celeux G, Martin O and Lavergne C (2005) Mixture of linear mixed models for clustering gene expression profiles from repeated microarray experiments. *Stat Modelling.* **5**, 243–267.
12. MacQueen J (1967) Some methods for classification and analysis of multivariate observations, In: LeCam, L.M., Neyman J. (Eds). *Proceedings of the Fifth Berkeley*

Symposium on Mathematical Statistics and Probability. University of California Press, 281–297.

13. Lehmann EL (1959) Testing statistical hypotheses, Wiley, New York.
14. Salvador S and Chan P (2004) Determining the Number of Clusters/Segments in Hierarchical Clustering/Segmentation, *Proc. 16th IEEE Intl. Conf. on Tools with AI*, pp. 576–584.

INDEX

Note: The letter 'n' followed by the locators denotes the note numbers.

G

H

I

K

L

M

N

O

P

R

S

T

W