GENOMICS

Helicobacter pylori: A surprisingly conserved bacterium

Robert E.W. Hancock, Richard Alm, James Bina, and Trevor Trust

It is one of the dogmas of the pharmaceutical industry that antiviral agents are developed to be virus specific, whereas antibacterials should be broad spectrum in nature. Very few bacteria justify themselves as standalone targets for pharmaceuticals. Such a designation requires that the organism have a unique biology and distinct modus operandi in infection (such that a given clinical syndrome virtually guarantees the presence of the bacterium), and that it causes infections of substantial economic significance.

Only two bacteria fit these criteria: *Mycobacterium tuberculosis* and *Helicobacter pylori*. The latter is the causative agent of a number of gastroduodenal diseases, including most cases of gastritis and duodenal ulcer and the majority of gastric ulcers, and has been implicated as a causative agent in stomach cancer. Indeed, the lucrative ulcer palliative industry—which manufactures blockers of stomach acidification to prevent pain resulting from ulcers—was built on the diseases that physicians once thought were caused by stress, but were shown, by Barry Marshall and Robin Warren[1], to be caused instead by *H. pylori*.

With this understanding has come the realization that successful ulcer therapy must involve an antibiotic or vaccine component directed against *H. pylori*. Unfortunately, *H. pylori*, although susceptible to antibiotics in the laboratory, is quite resistant to antibiotic therapy in the stomach/duodenum, and a successful vaccine has yet to appear. The potential for developing a successful antibiotic therapy or vaccine has been considered marginal, because examination of the macrorestriction pattern of four strains of *H. pylori* revealed almost complete heterogeneity in the physical genetic maps of these strains[2].

The genome of a clinical isolate of *H. pylori* (J99) had been sequenced in large part by Genome Therapeutics Corporation (licensed to ASTRA, Cambridge, MA). Recently, The Institute for Genomic Research (TIGR; Rockville, MD) released and published the complete genome of an unrelated strain of *H. pylori*[3]. The availability of genomic data from two diverse isolates has allowed some interesting and important comparisons to be made. Tomb et al.[3] seemed to confirm the potential for chromosomal variation in *H. pylori* with hypotheses of potential slipped-strand mispairing and recombination events because of the existence of repetitive sequences within a family of genes, which were predicted to encode outer-membrane proteins (OMPs), although such events have yet to be demonstrated experimentally.

The genome of a clinical isolate of *H. pylori* (J99) had been sequenced in large part by Genome Therapeutics Corporation. Recently, The Institute for Genomic Research released and published the complete genome of an unrelated strain of *H. pylori*. The availability of genomic data from two diverse isolates has allowed some interesting and important comparisons to be made.

It should be remembered that OMPs are often major antigens that in some pathogens determine serotype specificity. Some, called porins, are involved in mediating the passage of antibiotics into the cell, thereby playing a critical role in determining antibiotic susceptibility. This family of *H. pylori* OMPs was first introduced into the literature in 1995 as a family of four pore-forming proteins (porins HopA-D), which displayed extensive N-terminal sequence homology[4]. Tomb et al.[3] expanded this family to 32 proteins in an OMP family based on extensive C-terminal homology with (27 proteins), or without (5 proteins), a loose derivative of the conserved N-terminus. These proteins are clearly biologically significant, as family members have been shown to be porins and adhesins.

The composition of the outer membrane of *H. pylori*, which has a large array of moderately expressed proteins, differs from that of many other Gram-negative bacteria, which generally contain a smaller number of dominant proteins. Our own analysis reveals that there are at least two additional members of this family (HP0101 and HP1066), as well as four OMPs (HP0605, HP0839, HP1125, and HP1327) lacking these homologous sequence signatures. The 32 OMP proteins of Tomb et al. can be divided into subfamilies based on a combination of both multiple sequence and block alignments. Using this strategy there appears to be two independent subgroups, one containing 11 OMPs and a second containing 7 OMPs.

When the genes encoding these OMPs are further analyzed, some interesting observations can be made. First, in the TIGR strain, there are three pairs of genes that encode very similar proteins: *Omp16* and *Omp17* are 85% identical; *Omp12* and *Omp22* are 99% identical (two amino acids extra in signal sequence of *Omp22*); and *Omp5* and *Omp29* are 100% identical. The proteins in these last two pairs are spaced almost 500 kb away from their cognate partners on the chromosome. Analysis of the ASTRA J99 sequence also reveals the presence of these three pairs of genes, displaying similar high levels of identity within each group. First, the J99 equivalents of the *Omp12* and *Omp22* proteins also display 99% identity to each other, yet only 89% identity to the TIGR proteins.

Moreover, we have shown that a number of strains isolated from diverse geographical areas and distinct clinical manifestations all appear to contain two copies of the *Omp5/Omp29* gene. Second, in both strains there are three pairs of genes from this family that are very close together on the chromosome: *Omp7* and *Omp8* are 135 nt apart; *HopC* and *HopB* are 25 nt apart; and *Omp25* and *Omp26* are 26 nt apart. The closest alignments to *Omp7* and *Omp8* are *Omp25* and *Omp26*, respectively, suggesting that these genes may have arisen by an ancestral duplication and transposition of an entire locus.

Robert E.W. Hancock (bob@cbdn.ca) and James Bina (jeb@biosys.net) are in the Department of Microbiology and Immunology at the University of British Columbia, Vancouver, B.C. Canada, V6T1Z3. Richard Alm and Trevor Trust are at ASTRA Research Center Boston Inc., 128 Sidney St., Cambridge, MA 02139-4239.

Does this then confirm the genomic malleability of *H. pylori*? We would suggest not. Our comparison of the Genome Therapeutics/ASTRA sequence with that of the TIGR strain and other sequences in the literature shows that many of the genes mentioned above are highly conserved in *H. pylori*. For example, the HopC/HopB pair of genes that are closely juxtaposed on the chromosome are 46% identical to each other but are >95% identical to their corresponding proteins in three different *H. pylori* isolates. In fact, many of the members of this family of proteins, while not as conserved at the nucleotide level, share >90% identity at the amino acid level between these two different strains.

Thus, we predict that this family of OMPs, representing more than 1% of the entire chromosomal sequence, are probably not agents for scrambling the *H. pylori* chromosome. Rather, this substantive sequence conservation represents a sequence signature for regulating, processing, and interacting with other proteins. Our best guess at present comes from modeling the transmembrane topology of members of the general (nonspecific) porin family. Four such porin proteins have been crystallized and revealed to be 16-stranded β-barrels comprising antiparallel β-strands[5].

While there is no conservation of sequence between species for such β-strands, a general motif of alternating hydrophobic and hydrophilic residues is evident, especially in the N- and C-terminal β-strands, which are involved in subunit:subunit interactions (porins are trimers). Tomb et al. have already noted that the conserved C-termini demonstrate alternating hydrophobic residues, and our own modeling suggests that the most conserved stretches of amino acids can be predicted to be transmembrane β-strands. Thus, it is a strong possibility that this dramatic conservation of a sequence signature serves as the conserved scaffolding for a family of β-barrel proteins.

Overall, we suggest that the *H. pylori* genome is not as plastic as was once assumed, and that the potential for capitalizing on these genomic leads to discover novel vaccines and therapeutics remains high.

1. Marshall, B.J. and Warren, R. 1984. *Lancet* **1:**1311–1315.
2. Jiang, Q., Hiratsuka, K., Taylor, D.E. 1996. *Mol. Microbiol.* **20(4):**833-842.
3. Tomb, J.F. et al. 1997. *Nature* **388:**539–547.
4. Exner, M.M. et al. 1995. *Infect. Immun.* **63(4):**1567–1572.
5. Cowan, S.W. et al. 1992. *Nature* **358:**727–733.