



Microbial Genomics and Drug Discovery

Garry Myers

Microbial Genomics Group
The Institute for Genomic Research, USA

Since the completion of the first microbial genome sequence of *Haemophilus influenzae* in 1995, >125 microbial genomes have been published and at least another 340 are in progress (<http://wit.integratedgenomics.com/GOLD/>). This global effort has focused primarily on pathogens, which encompass the majority of all genome projects and has generated a large amount of raw material for *in silico* analysis. Additionally, multiple strains of the same organism, or multiple species of the same genus are being sequenced or have been completed, opening the possibility to use comparative genomics tools for discovering novel drug targets. A major challenge in the post-genomic era is to fully exploit and decipher this new wealth of information to fulfill the urgent need for novel anti-microbial agents.

Prioritizing targets using bioinformatics technologies is now commonplace in the industry but represents just the first step in a very long process. It is now possible to use bioinformatics tools to analyze genome sequence data and identify novel putative targets with potential for therapeutic intervention. The subsequent steps, including target validation, assay development and small molecule library screening still represent the most time -and labor-intensive parts of the drug discovery process.

Targets can be identified using a variety of bioinformatic screening approaches including a) screening for highly-conserved proteins b) screening for specific virulence proteins c) screening for surface proteins (Drugs designed to attack such targets do not have to penetrate the cell, thus simplifying chemical design) d) using structural genomic approaches to find possible targets for already-existing compounds (and develop new rationally designed drugs) e) detecting genes critical for pathogenic life stages. The recent genome sequencing of the *B. anthracis* genome uncovered several potential targets for new drugs, including proteins predicted to be on the outside of the spore and vegetative cell, proteins similar to other virulence factors in other organisms and also determinants apparently necessary for survival of the bacterium in macrophages - a critical stage in infection. This class included genes encoding germination determinants and proteins that may mitigate the oxygen radical-based killing of the organism by the macrophage.

Genomics can be applied also to evaluate suitability of potential targets, using for criteria. a) Essentiality: The target must be essential for the growth, replication, viability or survival of the microorganism in the laboratory, as primary assays for antimicrobial activity are typically performed under such conditions b) Selectivity: The microbial target should not have any well-conserved homolog in the mammalian host to be treated, in order to reduce cytotoxicity issues. The recent completion of the human and mouse genomes therefore represents a major step for the drug industry. c) Spectrum: Targets can be separated into two groups depending upon their distribution among microorganisms and the expected spectrum of potential inhibitors: (1) Targets that are present in a large number of

pathogens should be useful in the discovery of broad-spectrum agents. (2) Targets specific to one or a subgroup of pathogens can be used to develop agents that may have narrower spectra of activity d) Practicality: Once a protein has been selected using bioinformatics tools, a significant amount of work is generally required to determine if it represents a feasible target for drug discovery. For instance, membrane proteins tend to be poorly soluble when overexpressed in vitro and present a special challenge in assay development e) Functionality: Information on the function of the target can be critical in generating viable chemical leads from a particular screening assay. For example, functional information can help rationally design a library of chemical compounds to screen for target inhibition.

With the steadily decreasing costs of “traditional” Sanger sequencing and the advent of novel technologies looming (for instance, high-density microarray-based resequencing), the future promises much greater amounts of high-quality sequence information. Almost certainly microbial pathogens will continue to be sequenced in greater and greater numbers. This pile-up of very closely related whole genome sequences will offer a great deal of information to researchers developing drugs. Firstly will be increased knowledge of the association of specific proteins with virulence - implying an essential function. Also, the availability of multiple alleles of potential targets will allow design of molecules that target regions of the protein that are conserved across the spectrum of virulent organisms.

Finally, it is important to point out that the sequencing of non-pathogens also makes a significant contribution to drug discovery. Not only do these genomes serve as an important negative control in screens for drug targets but due to the increasing intermeshing of biology through databases, novel functions discovered in non-pathogens can be applied to virulent organisms that are the targets for eradication.