Antibiotic discovery: is it all in the genes?

The search for new classes of antibiotics to stem the global emergence of drug-resistant bacterial pathogens is perhaps one of the most important challenges facing the pharmaceutical industry. Nearly all contemporary classes of antibiotics were developed over 25 years ago during the 'golden era' of antimicrobial drug development [Knowles, D.J.C. *Trends Microbiol.* (1997) 10, 379–383]. Now, researchers working in industrial, government and university laboratories are scrambling to find new classes of antibiotics against novel molecular targets in bacteria.

In the past, antibiotic compounds were discovered fortuitously through massive anti-bacterial screening programmes with the specific mode of action being worked out much later. New advances in genetic engineering, over the past 20 years, have allowed for a paradigm shift in anti-infective R&D where drug discovery now starts at the level of the gene. In addition, fundamental improvements in electrical and mechanical engineering have led to smaller, faster and less expensive microprocessors and storage devices, which have in turn promoted advances in software engineering and laboratory automation. As a result, large volumes of relevant biological information are now available to address highly complex biological questions.

Bacterial genomics

This revolution in biological understanding is perhaps most evident in the field of genomic research. In 1995, the first genome of a cellular organism, the proteobacterium *Haemophilus influenzae*, was sequenced in its entirety [Fleishmann, R.D. *et al. Science* (1995) 269, 496–512]. Now, whole genomes have been sequenced from 16 species of bacteria and archaebacteria with several others nearing completion. Understandably, fewer eukaryotic genomes have been completed; nevertheless, progress has been impressive. The genome sequence of the yeast *Saccharomyces cerevisiae* [Jacq, C. *et al. Nature* (1997) 387 (6632 Suppl.), 75–78] was completely determined over a year ago and the first multi-cellular organism, the nematode *Caenorhabditis elegans*, is expected to be finished this year.

The Holy Grail of genomics

The enormous challenge of sequencing the human genome, the 'Holy Grail' of genomics, is currently under way in the public sector, and recently, two private ventures have proposed sequencing it to virtual completion by the year 2002 [Marshall, E. and Pennisi, E. *Science* (1998) 280, 994–995; Marshall, E. *Science* (1998) 281, 1121]. The genomes of several eukaryotic pathogens are also being sequenced, including the fungi *Candida albicans* (responsible for thrush and HIV-related infections) and *Plasmodium falciparum* (malaria), an apicomplexa protist. Many industrial interests have adopted the strategy of survey sequencing a wide variety of eukaryotic and bacterial pathogens to obtain rapidly information on potential anti-infective targets. It is very likely that 75–100 small genomes will have been sequenced, to varying levels of completion, by the end of the millennium (Fig. 1).

Integrating bioinformatics and evolutionary biology

If recent rounds of corporate restructuring, mergers and acquisitions are any indication, many companies are recognizing that the life sciences, driven by new information obtained through genomics and bioinformatics, will be an important sector in the future global economy [Enriquez, J. *Science* (1998) 281, 925–926]. However, for any therapeutic area, including anti-infectives, paradigms for the effective integration and utilization of genomic data in the R&D pipeline are still evolving. Perhaps the best way to discover new classes of anti-infectives rapidly lies in an area of research that is new to industrial R&D: evolutionary biology.

The importance of meshing evolutionary biology with anti-infective R&D is most evident if one considers the ideal characteristics of a potential molecular target for an antibiotic. Briefly, the target would:

- Occur over a broad spectrum of bacterial species – as accurate and rapid diagnosis of an infection in clinical situations is often difficult and a single treatment is more appealing and cost effective to the patient;
- Have no similar homolog in humans – thus avoiding cytotoxicity issues;
- Exist as a unique copy – as gene duplications can facilitate the evolution of antibiotic resistant loci;
- Be essential for the persistence of infection such that inhibition of the target would be lethal to the bacterium;
- Be amenable to high-throughput assays to convert efficiently the biological target into a sustainable chemical lead.

The first three points have a solid basis on the key assumption of molecular evolution: that genes, hence encoded proteins and RNAs, from diverse organisms are related because of shared ancestry. Bioinformatics, relying on methodologies for comparative genomics and evolutionary inference, can suggest potential targets that fulfill these initial three criteria and, in turn, provide testable hypotheses for laboratory experiments to validate the fourth criterion of *in vivo* essentiality. The final criterion – ease of assay development – can be evaluated by combined knowledge of biochemistry and screening technologies.

How is this achieved?

As part of the anti-infective research program at SmithKline Beecham (SB), the following approach is used to identify and prioritize *in silico* molecular targets in two key Gram-positive bacterial pathogens, *Streptococcus pneumoniae* and *Staphylococcus aureus*. First, an extensive database of all available partial and complete genomic sequences from both public and proprietary sources is built. Initially, this database focused on protein coding regions or translated open reading frames

 1359-6446/98/$ – see front matter © Elsevier Science. All rights reserved. PII: S1359-6446(98)01273-2

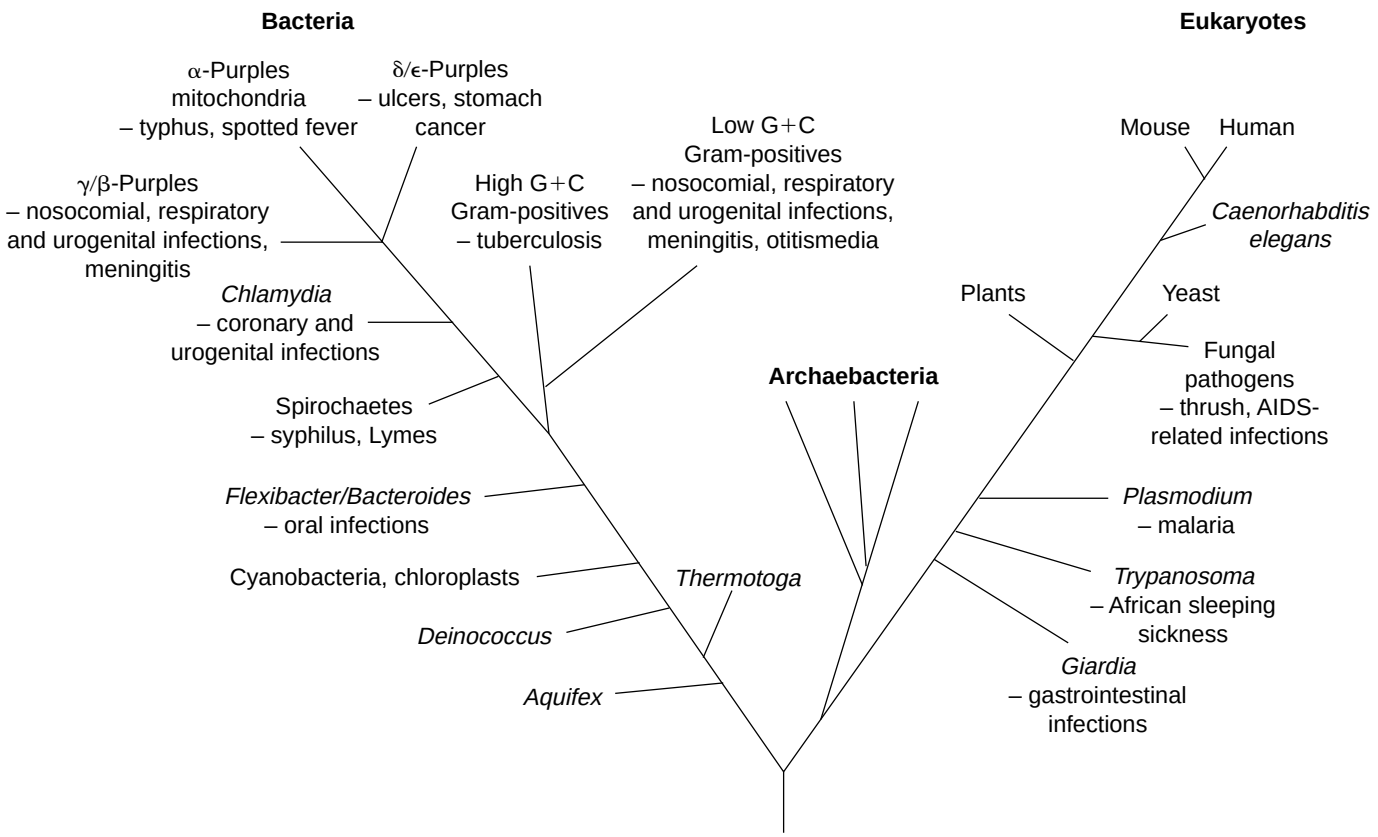

Figure 1. *Schematic rRNA phylogenetic tree showing the relative relationships among organisms whose genomes either have been or will be sequenced to near completion. Superimposed upon the tree are the major therapeutic areas in infectious diseases.*

(ORFs). ORFs were either obtained from public databases or predicted from DNA sequences. The inclusion of genome data from pathogenic bacteria and also nonpathogenic bacteria, archaebacteria and eukaryotes allowed for the development of an extensive network of gene relationships.

Homology relationship between ORFs

As of September 1998, the SB Microbial Bioinformatics Database had 47 separate genome databases representing 38 different species (for some species, sequence data are available from multiple strains or sources). An ORF by ORF array of sequence similarity scores (smallest sum probabilities) was built using the Basic Local Alignment Search Tool (BLAST) version 2.0 [Altschul, S.F. *et al. Nucleic Acids Res.* (1997) 25, 3389–3402] to establish a homology relationship between all available ORFs in every genome.

The results of these analyses were deposited into a relational database where various queries of the comparative array could be generated. The views of most immediate relevance to the identification of potential novel antibiotic targets were the degree of similarity of *S. pneumoniae* and *S. aureus* ORFs in the following comparisons: between the two species; with conspecific pathogens commonly found in the same infectious environments; with all other bacteria; and with humans (eukaryotes). On the basis of these comparisons, nearly 2000 predicted ORFs in both *S. pneumoniae* and *S. aureus* were prioritized for subsequent experiments that monitored the *in vivo* expression of genes in animal infection models, and determined *in vitro* and *in vivo* gene essentiality.

Limitations

However, many complexities of genome evolution still elude all known automated intergenomic comparison methods, including the SB approach.

Biodiversity

First, there is the incredible diversity of life itself. Typically, ~30% of all predicted ORFs from a bacterial or archaeal genome show little or no homology to another entry in public databases according to standard, high-throughput homology search tools. Although subsequent annotation efforts using more sophisticated and time-consuming methods have sometimes reduced that percentage, it is clear from all accounts thus far that a very significant portion of a genome is devoted to species-specific cellular processes. However, there are several unassigned

ORFs that are homologous across a very diverse range of species. Unassigned ORFs, that are specific to bacteria and demonstrated to be essential for infection, would comprise truly novel classes of antibiotic targets. Obviously, the major challenge in exploiting such targets would be the effective application of advanced computational and biochemical analyses to determine their precise cellular function.

Paralogous and orthologous genes

Second, sequence comparison methodologies often have difficulty sorting paralogous and orthologous gene relationships. Orthologous genes are those which are related by ancestry between species or organisms, while paralogous genes are related to genes in the same organism as well as genes in different organisms. At different cut-off values of similarity, automated procedures will find either orthologous or orthologous and paralogous members of a particular protein family. Many proteins have internal paralogous segments because they evolved through successive rounds of gene duplication, fusion and divergence, while some functionally distinct proteins may also share similar domains (i.e. GTP-binding proteins involved in translation [Cousineau, B. *et al. J. Mol. Evol.* (1997) 45, 661–670]. However, the levels of similarity between orthologous and paralogous proteins vary greatly between protein families making it difficult to develop generalized procedures for distinguishing between family members.

Horizontal gene transfer

Third, horizontal gene transfer between distantly related species can result in some very surprising species and gene relationships. Of course, horizontal gene transfer has long been recognized as playing a role in the proliferation of antibiotic resistance genes between bacteria. However, evolutionary biologists now believe that the eukaryotic genome itself has many genes obtained via horizontal gene transfer from bacteria [Reviewed by Brown, J.R. and Doolittle, W.F. *Microbiol. Mol. Biol. Rev.* (1997) 61, 456–502]. Some eukaryotic genes, and perhaps entire metabolic pathways, show greater similarity to bacterial versions than to archaebacterial homologs, the presumed prokaryotic progenitor of eukaryotes. The ramifications with respect to potential drug cytotoxicity are clear – it might be more difficult to find a specific antibacterial compound against a bacterial target that has a close human homolog.

Summary

Even discounting these exceptional cases, however, the prioritization of *S. aureus* and *S. pneumoniae* ORFs has proven to be very useful by eliminating hundreds of costly anti-infective 'dead-ends' while highlighting many targets that may have escaped the experienced researcher's intuition. To improve the historical description or evolutionary relationships of the proteins in the SB database, more sensitive tools are applied that incorporate more information inherent in each protein, such as conserved amino acid residues or conserved secondary and tertiary structural domains. Also, phylogenetic analyses are used to resolve evolutionary relationships among different species as well as to distinguish between levels of orthology and paralogy. However, these techniques are not yet automated and still rely heavily on the computational biologist's intuition.

Molecular evolution and anti-infectives research seem to be an unlikely winning combination in today's competitive pharmaceutical environment. In practice, however, the application of molecular evolutionary principles has helped to direct experimental efforts towards molecular targets that are more likely to be sustainable in the drug development pipeline. In addition, the methods of evolutionary biology and comparative genomics will continue to play an important role in expanding our general knowledge about the biology of pathogenic bacteria.

James R. Brown and Patrick V. Warren
Microbial Bioinformatics
SmithKline Beecham Pharmaceuticals
1250 S. Collegeville Road
PO Box 5089, UP1345
Collegeville, PA 19426-0989, USA
tel: +1 610 917 6374
fax: +1 610 917 7901
e-mail: james_r_brown@sbphrd.com

High-throughput screening

Patent strategies

The need for pharmaceutical companies to launch new drugs more frequently and thus maintain a lead over competitors, requires R&D management to improve productivity and focus on more efficient means of rapidly identifying new drug candidates. This has led to more focused R&D portfolio strategies, the development of high-throughput screening (HTS) systems and more aggressive approaches to protection of intellectual property. The development of HTS systems for the selection of drug candidates has driven the establishment of an active patent literature in this field.

In a recent patent analysis Jakobsen, P.H., Kurtzhals, P. and Poulson, F. have reviewed patent strategies for HTS assays [*Exp. Opin. Ther. Patents* (1998) 8, 1157–1165]. These patents cover various aspects of HTS, including specific use of equipment and reagents, and assay design. From a biological perspective applications often focus on the use of specific protein/receptor interactions or other cellular effects in the assay. Other approaches include protecting the *in vitro* methods developed for screening factors that modulate gene expression, inhibit protein or mRNA biosynthesis or modulate intracellular transduction of an extracellular signal. In many case the patent applications contain broad functional claims that have yet to be legally challenged to establish their validity. There will be an increase in the filing of patent applications in this field. Given the broad diversity of the patent literature and the rapidity of developments within this field, it is incumbent on users of HTS systems to maintain an awareness of the developing HTS patent literature.

Andrew Lloyd

 1359-6446/98/$ – see front matter © Elsevier Science. All rights reserved. PII: S1359-6446(98)01272-0