Trends in Microbiology Vol. 10 No. 10 (Suppl.), 2002

The impact of genomics on anti-infectives drug discovery and development

Tanya Parkinson

Genomics and pharmacogenomics are signalling the start of a new era for the pharmaceutical industry. The successful integration of these technologies into the drug discovery process provides the promise of increased efficiency for pharmaceutical companies, with higher confidence in the targets they pursue and smarter design of clinical trials. There are benefits too for the consumer, with the possibility of customized drug treatments leading to improved efficacy and fewer side-effects. This article reviews the impact of genomics at the various stages in the lifetime of a drug, through discovery, development and clinical use, focusing particularly on anti-infectives.

Tanya Parkinson
Pfizer Global Research and Development, Sandwich, Kent, UK CT13 9NJ.
e-mail: tanya_parkinson@sandwich.pfizer.com

Infectious diseases are the number one cause of premature death in the world (http://www.who.int/infectious-disease-report/index.html). A large number of antimicrobial drugs are used clinically, but their effectiveness is being eroded by the development of resistance, and concerns over safety. The need for newer and safer antimicrobials therefore continues unabated. Recent years have seen an explosion in genome sequencing of both microbial pathogens and their human host, which is helping us to understand the complex interactions involved in the infection process. The pharmaceutical industry is exploiting this information to identify better targets for treating infectious diseases and to improve understanding of patient responses to a drug (pharmacogenomics).

Target selection

Table 1 summarizes the various genomic approaches to target selection. One of the major challenges in drug discovery for the treatment of infectious diseases is to identify targets that are essential for the microbe to survive, but which are absent, or significantly divergent, in their mammalian host. For viral diseases, the small genome and relatively few viral proteins make this process fairly straightforward. However, for bacterial and fungal pathogens, there is a much larger potential pool from which to select targets. Various large-scale mutagenesis approaches are available for identification of essential fungal and bacterial genes. These provide a much more efficient method of identifying and validating targets compared with the traditional gene-by-gene approach. It is also possible to tailor these approaches to look at specific areas of metabolism that have a higher probability of delivering targets. For example, cell-wall biosynthesis is the target for nikkomycin and echinocandin antifungals, but many steps in the synthesis and assembly of the cell wall are as-yet unknown. By using mutagenesis approaches, it is possible to identify mutants with defective cell walls, which might reveal new targets in this area [1].

Signature-tagged mutagenesis

Signature-tagged mutagenesis provides a method for simultaneous identification of multiple genes required for survival of the pathogen in a mammalian host [2]. The method relies on construction of a transposon-mutagenized bacterial or fungal library in which the transposon contains a small variable sequence tag that can be amplified by PCR. The mutant cells are individually arrayed into microtitre plates, then pooled and used to infect a suitable animal model. At the same time, colony or dot blots are made from the plates. Cells are recovered from the animal model and PCR is used to make labelled-tag probes, which are hybridized to the blots. Mutants unable to survive will not be present in the pool of cells from the animal and can be recovered from the original arrays for further analysis. This method has been used to identify genes required to establish an infection from a variety of pathogenic bacteria [3] and fungi [4,5], but has the disadvantage of losing any mutations in genes that are essential *in vitro*. To capture such genes would require a conditional mutagenesis strategy, such as constructing a library of genes under the control of a regulatable promoter, or a library of temperature-sensitive mutants.

In vivo expression

Another way to identify putative virulence genes is by using *in vivo* expression technology. This is a promoter-trap method for identifying pathogen genes that are induced under *in vivo* conditions [2]. A reporter gene, which can

 PII: S0966-842X(02)02427-7

Table 1. Genomic technologies for target selection

Technology	Use	Advantages	Disadvantages
Signature-tagged mutagenesis	Identification of genes required for pathogen survival in animal models	Identifies essential genes *in vivo*	Limited to genes that are not essential *in vitro* Resource required to construct initial library
In vivo expression technology	Identification of pathogen genes induced *in vivo*	Identifies genes induced *in vivo*	Increased expression does not necessarily mean the gene is essential *in vivo* Further validation required Resource required to construct initial library
Microarrays	Understanding host response to pathogens Correlating gene expression with pathogenicity Identifying molecular targets of antimicrobial compounds Inferring function of unknown genes	Low resource: no library construction required Looks at all genes simultaneously	Increased expression does not necessarily mean the gene is essential Further validation required Technically difficult to use for pathogens grown *in vivo*
Comparative genomics	Identification of pathogenicity-related genes Identification of antigens for vaccine development Selecting targets conserved across multiple pathogens Selecting targets with lowest homology to human proteins	No experimental work required: *in silico* analysis	Spectrum and selectivity analysis based on linear sequence analysis could be misleading
Structural genomics	Selecting targets conserved across multiple pathogens Selecting targets with lowest homology to human proteins Inferring function of unknown proteins	Spectrum and selectivity analysis based on 3D analysis of the active site	High resources required to solve crystal structures

be a gene known to be essential for survival (e.g. *purA* for *Salmonella typhimurium*), or alternatively a fluorescent marker (e.g. green fluorescent protein), is randomly integrated into the genome of the test organism. Promoters that are activated *in vivo* result in expression of the reporter gene. Such strains can be recovered from the animals and characterized to identify the promoter and corresponding gene. These technologies have been widely applied to pathogenic bacteria [6] and, on a more limited scale, in pathogenic fungi, to examine *in vivo* expression of the family of secreted aspartyl proteases in *Candida albicans* [7].

Microarrays

DNA microarrays offer an alternative method for identification of virulence genes. They show significant advantages over signature-tagged mutagenesis and *in vivo* expression technology in that no complicated genetic manipulations are required to construct mutant libraries. A highly informative microarray experiment would be to compare profiles of *in vitro* and *in vivo* grown pathogens. However, there are significant technical challenges associated with isolating RNA from pathogens grown *in vivo*. Low numbers of pathogens and the problem of specifically isolating pathogen RNA without isolating contaminating mammalian RNA make this a difficult experiment. Sub-culture of the pathogen would solve the problem, but would also result in loss of the specific gene expression responses induced by *in vivo* conditions. More commonly, investigators use microarrays to compare pathogenic and non-pathogenic isolates. This type of study has been used to correlate severity of clinical disease with genome deletions in *Mycobacterium tuberculosis* [8], and to identify the *cag* pathogenicity island of *Helicobactor pylori* as a key factor in inducing gastric inflammation [9].

Microarrays also offer the possibility of studying the host response to pathogens. This could suggest ways in which the immune system can be stimulated or augmented to achieve an antimicrobial effect. It might also enable the generation of knockout mice with increased susceptibility to particular infections, thus facilitating studies on virulence genes or profiling of antimicrobial drugs. Of particular relevance to viral infections is the identification of host genes that are required to support the viral life cycle. As viruses have small genomes encoding relatively few genes, they are dependent on host proteins to support their life cycle. Microarray experiments have been carried out on a wide-range of viruses with the aim of identifying putative targets for antiviral therapy [10].

With these types of microarray studies, it is important to bear in mind that upregulation of a gene transcript

might not necessarily correspond to an increase in the levels of active protein encoded by the gene. Regardless of whether it does, there is no certainty that inhibition of the function of that protein will prevent infection by the pathogen. There is still a need to follow-up the results of such studies with focused target validation experiments, such as antisense. However, the advantage the technology brings is in directing such validation experiments to a smaller subset of potential targets.

Microarrays can also be used to identify the molecular target of antimicrobial compounds of unknown mechanism or to determine the function of uncharacterized genes. Changes in gene expression in response to antimicrobial compound treatment are often indicative of the mechanism-of-action of the compound. Thus, the azole antifungal compounds that inhibit ergosterol biosynthesis induce expression of genes in the ergosterol biosynthesis pathway [11], and isoniazid, which inhibits a fatty acid synthase complex (required for mycolic acid synthesis) induces expression of genes involved in fatty acid and mycolic acid synthesis in *M. tuberculosis* [12]. By examining the transcriptional profile induced by treatment with an antimicrobial compound of unknown mechanism, it could be possible to infer the mechanism-of-action. Similarly, for genes of unknown function, it should be possible to construct a mutant strain and compare the transcriptional profile to a library of known mutant and compound-treated profiles. This approach has been applied to the *Saccharomyces cerevisiae* gene *YER044c* whose function was unknown. Transcriptional profiling revealed a pattern of gene expression similar to that of mutants in ergosterol biosynthesis, and further biochemical profiling confirmed that this gene product was involved in the ergosterol synthesis pathway [13].

Comparative genomics

Once potential targets have been identified, comparative genomics can be used to check for related proteins in other key pathogens to identify targets that could have broad-spectrum antimicrobial activity. Similarly, comparison of pathogen and human genome sequences can help identify targets with the lowest potential for host toxicity. This approach has been used to identify aminoacyl-tRNA-synthesis targets in pathogenic bacteria [14] and has also been applied to the selection of antifungal targets [15]. However, these types of full-length sequence analyses can be misleading. Small-molecule inhibitors frequently bind in the active site of enzymes, which are not generally comprised of consecutive amino acids from a linear sequence. Therefore, when evaluating spectrum and selectivity, it is more relevant to use 3D protein structure information to guide target selection. Structural genomics aims to provide a 3D structure for every protein in an organism. This is a major technical challenge that will not be achievable in the near future. Instead, industrial and academic consortia are seeking a representative structure for every existing protein fold, currently estimated at <1000 for soluble proteins [16].

Structural genomics

In addition to impacting on the selection of targets with respect to spectrum and selectivity, structural genomics can identify new targets by assigning function to previously uncharacterized proteins, based on structural homologies. An example of this is the *E. coli* Gab protein: structural analysis revealed that it was a member of the non-haem iron (II)-dependent oxygenase superfamily [17]. Structural analysis of proteins can also have a major impact on the design of antimicrobial compounds. For example, the crystal structure of the bacterial ribosome has revealed strategies for circumventing antibiotic resistance associated with aminoglycosides that target the A-site of the 16S RNA [18].

Another potential area where genomics could lead to smarter target selection is in the identification of target proteins for vaccine development. As mentioned previously, putative virulence factors can be identified by comparative genomics or *in vivo* methods. However, in addition, several computer programs are now available to search for secreted or membrane proteins that could be putative antigens. This *in silico* approach has been applied to the selection of antigens for Group B *Neisseria menigitidis* vaccine development [19,20] and *Chlamydia pneumoniae* vaccine development [21].

Genetic polymorphisms and therapeutic drug concentrations

Variations in drug metabolizing enzymes and drug transporters can result in a wide variation of drug concentrations between patients. Higher-than-expected drug levels can lead to serious side-effects, whereas lower-than-expected drug levels can compromise efficacy and lead to resistance development. Many polymorphisms in the major drug-metabolizing enzymes of the cytochrome P450 (CYP) family have been reported that can result in slow, normal or fast metabolism, with high, normal or low drug levels, respectively [22]. Polymorphisms in the major multidrug transporters might also influence drug concentrations. For example, patients carrying an MDR1 3435 TT genotype have low levels of P-glycoprotein expression in peripheral blood mononuclear cells, and low plasma levels of the antiretroviral drugs, nelfinavir and efavirenz [23]. Drug toxicity can also be related to polymorphisms in genes other than those encoding

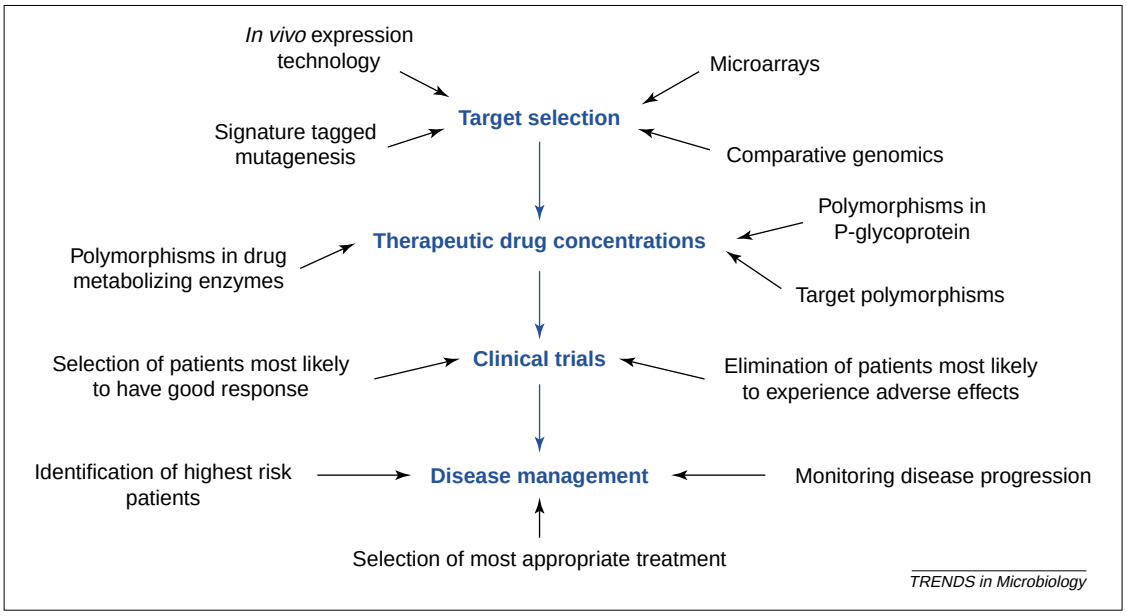

Figure 1. The impact of genomics on the drug discovery and development process

Genomic technologies impact on all stages in the lifetime of a drug, from target selection through to patient treatment.

drug-metabolizing enzymes. This is illustrated by the aminoglycoside antibiotics, which show ototoxicity (drug-induced hearing loss) as a side effect. A single nucleotide polymorphism (A1555→G) in the 12S rRNA gene has been linked to susceptibility to this side effect [24]. Another example is protease-inhibitor treatment in HIV, where a G→A polymorphism in the –238 promoter region of the TNF-α gene has been linked to lipodystrophy (a fat re-distribution syndrome) [25].

Genetic polymorphism of the drug target is another source of variability in drug efficacy. For anti-infectives, the key factor will be development of resistance by mutation of the drug target. DNA microarrays have been applied to studying the development of resistance of *Mycobacterium* to rifampin [26], and of HIV to protease and reverse-transcriptase inhibitor treatment [27]. However, as antiviral therapy moves away from targeting viral enzymes and towards focusing more on host targets, genetic polymorphisms in the host will become more important.

Pharmacogenomic therapeutic drug monitoring offers the possibility of genotyping patients to predict their response to multiple drugs, and therefore enables drug therapy to be personalized to suit each individual. This should ensure that every patient receives the optimal dosage, thus maximizing efficacy and minimizing adverse effects. An additional advantage for the patient is that the genotyping can be performed on a simple non-invasive sample, such as a swab of cheek cells.

Clinical trials

Predicting drug efficacy by genotyping will enable pharmaceutical companies to recruit patients into their clinical trials who are most likely to benefit from a drug. They will also be able to screen out patients most likely to suffer adverse reactions. As a result of this, clinical trials could become smaller, cheaper and faster to run.

But is this really good news? Tailoring drugs to specific segments of the population could fragment the market for pharmaceuticals. There are ethical considerations too. As many of the polymorphisms relevant to drug treatment can vary with ethnic origin, there could be issues over inclusion. Some drugs might be approved that are excluded for use with some ethnic groups. In addition, safety could be compromised if the clinical-trial group is small and genotypically similar, resulting in some side-effects going undetected. Of course, these considerations apply to drugs for any therapeutic indication and are not specific to anti-infectives. For more comprehensive reviews on the ethical implications of pharmacogenomics, see [28–30].

Disease management

The benefits of pharmacogenomic therapeutic drug monitoring for the patient are obvious. They are more likely to receive the optimal dose of a drug, and less likely to suffer adverse effects. But there are additional benefits that patients can derive from pharmacogenomics. Physicians will be able to identify individuals with increased susceptibility to diseases and apply appropriate prophylaxis. For example, polymorphisms have been associated with increased severity of parasitic diseases such as schistosomiasis and malaria [31].

Pharmacogenomics can also be used to predict disease progression. Most individuals infected with hepatitis B suffer an acute infection, which resolves itself in time. However 5–20% of infected individuals go on to develop

chronic infection, which can lead to liver disease and hepatocarcinoma. Recent studies have suggested that the chronic state is linked to a variant cytokine-receptor gene on chromosome 21 [31]. Similarly, disease progression in HIV-infected patients is linked to the CCR5 Δ32 deletion, with heterozygotes showing delayed disease progression, and homozygotes showing resistance to infection [32].

Another advantage of pharmacogenomics is that it helps physicians to select the most appropriate treatment for their patients. For example, individuals infected with hepatitis-C genotypes 2 and 3 respond much better to interferon and ribavirin treatment than those infected with genotype 1 [33]. Knowing which genotype the patient is infected with therefore enables the physician to select the most appropriate treatment regime. Virus genotyping is also crucial in HIV therapy, given that there are high rates of resistance to antiretroviral drugs, and a correlation between certain mutations and clinical response [34].

Concluding remarks

It is clear that genomics and pharmacogenomics have the potential to change the way drugs are discovered, developed and prescribed (Fig. 1). Target selection, drug monitoring, design of clinical trials, and patient treatment could all become more efficient. Indeed, we are already seeing some of these changes, and the prospect of personalized medicines moves ever closer. However, there will undoubtedly be a price to pay. These technologies are expensive and will require investment to develop and integrate them into current clinical practices. In our excitement and enthusiasm, we must also take care to consider the ethical implications of this new revolution.

References

1 Lussier, M. *et al.* (1997) Large scale identification of genes involved in cell surface biosynthesis and architecture in *Saccharomyces cerevisiae*. *Genetics* 147, 435–450
2 Chiang, S.L. *et al.* (1999) *In vivo* genetic analysis of bacterial virulence. *Annu. Rev. Microbiol.* 53, 129–154
3 Mecsas, J. (2002) Use of signature-tagged mutagenesis in pathogenesis studies. *Curr. Opin. Microbiol.* 5, 33–37
4 Brown, J.S. *et al.* (2000) Signature-tagged and directed mutagenesis identify PABA synthetase as essential for *Aspergillus fumigatus* pathogenicity. *Mol. Microbiol.* 36, 1371–1380
5 Nelson, R.T. *et al.* (2001) Identification of virulence mutants of the fungal pathogen *Cryptococcus neoformans* using signature-tagged mutagenesis. *Genetics* 157, 935–947
6 Slauch, J.M. (2000) IVET and RIVET: use of gene fusions to identify bacterial virulence factors specifically induced in host tissues. *Methods Enzymol.* 326, 73–96
7 Kretschmar, M. (2002) Individual acid aspartic proteinases (Saps) 1-6 of *Candida albicans* are not essential for invasion and colonization of the gastrointestinal tract in mice. *Microb. Path.* 32, 61–70
8 Kato-Maeda, M. *et al.* (2001) Comparing genomes within the species *Mycobacterium tuberculosis*. *Genome Res.* 11, 547–554
9 Israel, D.A. *et al.* (2001) *Helicobacter pylori* strain-specific differences in genetic content, identified by microarray, influence host inflammatory responses. *J. Clin. Invest.* 107, 611–620
10 Früh, K. *et al.* (2001) Virogenomics: a novel approach to antiviral drug discovery. *Drug Discov. Today* 6, 621–627
11 Bammert, G.F. and Fostel, J.M. (2000) Genome-wide expression patterns in *Saccharomyces cerevisiae*: comparison of drug treatments and genetic alterations affecting biosynthesis of ergosterol. *Antimicrob. Agents Chemother.* 44, 1255–1265
12 Wilson, M. *et al.* (1999) Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. *Proc. Natl. Acad. Sci. U. S. A.* 96, 12833–12838
13 Hughes, T.R. *et al.* (2000) Functional discovery via a compendium of expression profiles. *Cell* 102, 109–126
14 Raczniak, G. *et al.* (2001) Genomics-based identification of targets in pathogenic bacteria for potential therapeutic and diagnostic use. *Toxicol.* 160, 181–189
15 Spaltman, F. *et al.* (1999) Computer-aided target selection – prioritizing targets for antifungal drug discovery. *Drug Discov. Today* 4, 17–26
16 Wolf, Y.I. *et al.* (2000) Estimating the number of protein folds and families from complete genome data. *J. Mol. Biol.* 299, 897–905
17 Chance, M.R. *et al.* (2002) Structural genomics: a pipeline for providing structures for the biologist. *Prot. Sci.* 11, 723–738
18 Fong, D.H. and Berghuis, A.M. (2002) Substrate promiscuity of an aminoglycoside antibiotic resistance enzyme via target mimicry. *EMBO J.* 21, 2323–2331
19 Pizza, M.G. *et al.* (2000) Whole genome sequencing to identify vaccine candidates against serogroup meningococcus. *Science* 287, 1816–1820
20 Tettelin, H. *et al.* (2000) Complete genome sequencing of *Neisseria meningitidis* serogroup B strain MC58. *Science* 287, 1809–1815
21 Grandi, G. (2001) Antibacterial vaccine design using genomics and proteomics. *Trends Biotechnol.* 19, 181–188
22 Brockmoller, J. *et al.* (2000) Pharmacogenetic diagnostics of cytochrome P450 polymorphisms in clinical drug development and in drug treatment. *Pharmacogenomics* 1, 125–151
23 Fellay, J. *et al.* (2002) Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter 1: a pharmacogenetics study. *Lancet* 359, 30–36
24 Bacino, C. *et al.* (1995) Susceptibility mutations in the mitochondrial small ribosomal RNA gene in aminoglycoside induced deafness. *Pharmacogenetics* 5, 165–172.
25 Pirmohamed, M. and Back, D.J. (2001) The pharmacogenomics of HIV therapy. *Pharmacogenomics J.* 1, 243–253
26 Troesch, A. *et al.* (1999) *Mycobacterium* species identification and rifampin resistance testing with high-density DNA probe arrays. *J. Clin. Microbiol.* 37, 49–55
27 Wilson, J.W. *et al.* (2000) Comparative evaluation of three human immunodeficiency virus genotyping systems: the HIV-GenotypR method, the HIV PRT GeneChip Assay, and the HIV-1 RT Line Probe Assay. *J. Clin. Microbiol.* 38, 3022–3028
28 Anderlik, M.R. and Rothstein, M.A. (2001) Privacy and confidentiality of genetic information: what rules for the new science. *Annu. Rev. Genomics Hum. Genet.* 2, 401–433
29 Issa, A.M. (2000) Ethical considerations in clinical pharmacogenomics research. *Trends Pharm. Sci.* 21, 247–249
30 Rothstein, M.A. and Epps, P.G. (2001) Ethical and legal implications of pharmacogenomics. *Nat. Rev. Genet.* 2, 228–231
31 Hill, A.V.S. (2001) The genomics and genetics of human infectious disease susceptibility. *Ann. Rev. Genomics Hum. Genet.* 2, 373–400.
32 Huang, Y. *et al.* (1996) The role of mutant CCR5 allele in HIV-1 transmission and disease progression. *Nat. Med.* 2, 1240–1243
33 Podzorski, R.P. (2002) Molecular testing in the diagnosis and management of hepatitis C virus infection. *Arch. Pathol. Lab. Medicine* 126, 285–290
34 Clevenbergh, P. *et al.* (2000) Persisting long-term benefit of genotype-guided treatment for HIV-infected patients failing HAART. The Viradapt study: week 48 follow-up. *Antivir. Ther.* 5, 65–70