Integrated bacterial genomics for the discovery of novel antimicrobials

Hannes Loferer, Alexander Jacobi, Anton Posch, Christine Gauss, Sebastian Meier-Ewert and Bernd Seizinger

Sequencing of bacterial genomes has been progressing with breathtaking speed. Currently, the genomes of 23 bacterial species are sequenced, with approximately 40 more sequencing projects in progress. Industrial research is now facing the challenge of translating this information efficiently into drug discovery. This review will summarize the impact of bacterial genomics, bioinformatics and second-generation genomic technologies on target identification, assay development, lead optimization and compound characterization.

One of the greatest achievements of medicine in the 20th century is the use of antimicrobial drugs to control infectious diseases. Initially, the 'antibiotic era' was dominated by penicillin and the sulphonamides. Within two decades following the introduction of penicillin, most of the existing classes of antibiotics had been discovered by systematic screening of natural product libraries. Furthermore, no new chemical classes of active antibiotics have been successfully introduced into the clinic for over 30 years (Ref. 1).

No less remarkable, however, is the versatility shown by microorganisms in overcoming the effects of antibiotics. Bacteria have developed a variety of resistance mechanisms coupled with the ability to mobilize the respective genetic information between bacterial strains and species[2]. Short-term measures, such as chemical modification of existing antibiotics and development of inhibitors of resistance genes, will have a significant impact on antibacterial therapy in the immediate future. However, it is evident that this field requires additional targets, innovative assay-development strategies and new chemical entities[3]. While the introduction of innovative chemistry will probably be triggered by combinatorial chemistry[4] and novel resources for natural products[5], there are high expectations for microbial genomics in accelerating target discovery and assay development[3]. This review describes the latest genomics-based technologies and their potential impact on antibacterial drug discovery.

Genomics and target discovery

Since the first whole-genome sequence of a self-replicating organism was reported, the whole-genome sequences of 22 more bacterial species have been published (http://www.tigr.org). The information gained from these projects has already had a major impact on both basic microbiology and its industrial applications, and has rapidly changed the way research is conducted in this field.

This review mainly focuses on antibacterial targets that are essential for bacterial growth. Latest developments in the areas of targets involved in bacterial virulence or resistance against antibacterial agents have been reviewed previously[3].

Large-scale genetics

Traditionally, the search for novel genes required for bacterial survival or virulence was based on several genetic methods involving random mutagenesis of a bacterial

[a]**Hannes Loferer***, [a]**Alexander Jacobi**, [a,b]**Anton Posch**, [a,b]**Christine Gauss**, [a]**Sebastian Meier-Ewert** and [a]**Bernd Seizinger**, [a]Genome Pharmaceuticals Corporation, Fraunhoferstrasse 20, D-82152 Martinsried/Munich, Germany. [b]TOPLAB, Gesellschaft für angewandte Biotechnologie, Fraunhoferstrasse 18a, D-82152 Martinsried/Munich, Germany. *tel: +49 89 85 65 26 50, fax: +49 89 85 65 26 10, e-mail: hannes.loferer@gpc-ag.com

 1359-6446/00/$ – see front matter PII: S1359-6446(99)01455-5

genome followed by screening for the relevant phenotype[6]. Previously, the most successful strategy for finding novel genes essential for viability was the isolation of conditionally lethal mutants. A conditionally lethal mutation causes the gene product to function normally under one set of environmental conditions (the permissive condition – e.g. growth at 37°C for *Escherichia coli*), but fail to function under another set of conditions (the non-permissive condition – e.g. growth at 42°C for *E. coli*). The genes that can mutate to conditional lethality (i.e. mutations that are lethal to a bacterium under non-permissive conditions) are generally genes that are essential for viability. However, such temperature-sensitive mutants do have their limitations[3]. It has been estimated that approximately one-third of proteins are difficult to mutate into a thermolabile form[7]. In addition, genes have been identified that are required for viability at high temperatures, and this will lead to false-positive assignment of some genes as being essential under all growth conditions[7]. In summary, despite the widespread success of using conditional mutants to identify essential genes, it can be assumed that a significant number of essential genes still remain undiscovered by this technology.

Until recently, transposon mutagenesis (i.e. the inactivation of genes by insertion of a transposable genetic element) of bacterial genomes was not a powerful approach for the identification of essential genes, because the frequency and randomness of insertions was too low. A breakthrough in using transposon mutagenesis to map genes required for cellular viability was achieved by combining *in vitro* transposition with natural transformation of certain bacterial species[8]. In this 'genomic analysis and mapping by *in vitro* transposition', or GAMBIT, the respective bacterial genome is covered by an overlapping set of long-range PCR products (≈10,000 base pairs). Each PCR product undergoes *in vitro* transposition, a procedure that leads to almost completely random and high-frequency insertions into the target DNA (Ref. 9). Subsequently, the bacterial strain is transformed by the mutagenized PCR product and the locations of transposon insertions are mapped by PCR. As several insertions are obtained within each open reading frame (ORF), genes essential for survival under standard laboratory conditions can be recognized by their lack of transposon insertions. Hence, within a reasonable time period, most essential genes of a small-genome bacterium, like *Haemophilus influenzae*, can be mapped. Furthermore, the transposon mutants obtained by GAMBIT can be used to identify genes necessary for non-standard conditions (such as virulence), this being similar to the signature-tagged mutagenesis approach[10].

Comparative genomics

Bacterial genome sequencing has triggered a complementary approach to target discovery that is directed rather than random, consisting of comparative genomics combined with bacterial genetics[11]. Extensive genomic information gained from many evolutionarily distant bacterial species has made the automated comparison of bacterial genomes a powerful tool for categorizing genes and their respective products[12,13]. Thus, focused lists of target candidates are generated by comparative genomics that are then rapidly validated using bacterial genetics. The gene categories generated by this approach enable such a preselection of target candidates on a whole-genome scale; in other words, targets can be defined according to the required characteristics for a given antibacterial treatment.

For example, genes that have orthologs in many evolutionarily distant organisms are target candidates for broad-spectrum applications. Similarly, genes can be selected that are present only in a small subset of the bacterial species sequenced to date, thereby representing possible targets for narrow-spectrum antibacterial compounds. This target category is of particular importance for the treatment of chronic infections, as narrow-spectrum drugs would reduce both the spread of drug resistance and the side effects caused by destruction of the commensal bacterial flora, both of which are major disadvantages of long-term treatment with broad-spectrum antibiotics. Targets can also be selected according to their putative functions, although it should be noted that this approach, in particular, is highly dependent on accurate functional annotation of genomes and experimental validation is still a necessity[11].

The genomes of many closely related organisms are still to be sequenced. However, it is already being appreciated that information can be gained from comparative genomics within a given species[14,15] or even from strain variants[16]. For example, the comparative analysis of the genomes of *Chlamydia trachomatis* and *Chlamydia pneumoniae* has generated testable hypotheses of genes that might be responsible for the differences in tropism and pathologies between these two organisms[14]. In parallel to analyzing differences in closely related genomes by sequencing, DNA-array technologies have enabled the possibility of comparing genomes by hybridization. As exemplified by the comparative hybridization analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* by Bacille Calmette-Guérin (BCG), genomic regions that are different between pathogenic and non-pathogenic variants of a species can be rapidly identified[17]. These regions contain genes that are likely to be of relevance for the development of antibiotics and/or vaccines.

In addition to categorizing genes solely by sequence-based comparisons (e.g. BLAST – basic linear alignment search tool[18]), the coding sequences elucidated in new sequencing projects can be compared with reference databases of cellular pathways created mainly using the biological information known about *E. coli* and *Bacillus subtilis*[19]. Using this method, the metabolic capabilities of a newly sequenced organism can be assessed, and putatively essential pathways and missing components of pathways identified.

Genes of unknown function

One of the most intriguing results from the bacterial sequencing projects completed so far is that a significant fraction of the genes have unknown function[20]; in other words, these genes have been identified solely by sequencing and have not been previously characterized by any genetic or biochemical approach[21]. Even in well-studied model organisms such as *E. coli*, ≈38% of all genes surprisingly fall into this category[20,21]. Given the drawbacks listed earlier of random screens for temperature-sensitive mutations, it can be speculated that, despite the extent of genetic studies on *E. coli* and *B. subtilis*, many targets still remain undiscovered among the genes of unknown function. Indeed, in a pilot study that investigated 26 FUN (Function UNknown) genes that are broadly conserved among diverse bacterial species (including currently the smallest bacterial genome of *Mycoplasma genitalium*), six novel genes essential for the growth of *E. coli* and *B. subtilis* were identified[11]. One of these genes was earlier eliminated from a postulated minimal gene set required for life based on its annotation as a host-interacting protein[12].

Genomics and assay development

Traditionally, screening for novel antimicrobial compounds in industry has been performed by testing large libraries of natural products for their ability to kill bacteria. Many of the antibiotics used today were discovered this way. This classical approach is again enjoying increased interest because of improved chemical diversity through combinatorial chemistry[4] and the exploitation of unexplored resources of natural products[5]. However, this approach has disadvantages such as low sensitivity and the fact that the targets of the respective compounds are unknown. The latter point has also created interest in applying genomic technologies to investigate the mode of action of antibacterial compounds.

In an alternative approach, individual proteins involved in well-studied essential pathways are purified and biochemical *in vitro* assays set up to screen compound libraries for inhibitors. Empirically, many of the primary hits of these screens will have to be modified to enable penetration into the bacterial cell. Furthermore, the biological function (i.e. biochemical activity) of the respective target protein needs to be known to enable this type of assay development.

As already outlined, the latest developments in whole-genome genetic analysis (e.g. GAMBIT or the *B. subtilis* functional analysis program) will lead to the identification of most genes essential for growth and survival of several bacterial species within the near future. However, detailed functional information that will enable classical assay development will only be available for a minority of new targets. Thus, to use this extensive target information efficiently, innovative assay development strategies are required that are applicable to a broad functional variety of antibacterial targets. 'Second-generation' genomics technologies promise to play a central role in defining those strategies, some of which will now be described.

Surrogate markers

One approach towards the development of a generic assay is the identification of surrogate markers. Within the context of antibacterial drug discovery, surrogate markers are defined as genes that are deregulated as a specific response towards the inactivation of a given essential target. Eventually, inactivation of the respective target should occur through the action of a small molecule. However, for the identification of surrogate markers, target inactivation can be achieved through *in vivo* expression of small-peptide inhibitors (i.e. surrogate ligands) or by shifting a conditional mutant towards non-permissive conditions. Besides isolating temperature-sensitive alleles of the respective gene, conditional mutants can be generated by positioning a complementary copy of the essential gene under the control of a tightly regulated promoter[11]. The availability of such conditional mutants enables the analysis of the phenotypic consequences for a bacterial cell caused by the inactivation of an essential gene. An example of such an experiment is shown in Fig. 1. Thus, samples harvested before and after depletion of the target of interest can be compared to investigate its molecular effect on the bacterial cell. Characterization of these types of responses using genomic technologies such as RNA expression profiling or proteome analysis enables the identification of genes/proteins that are deregulated as a consequence of target inhibition. Hence, such deregulated genes/proteins are indicative of the biological activity of a given target and are defined as surrogate markers.

One crucial issue regarding this approach is whether such target- or pathway-specific responses can be detected

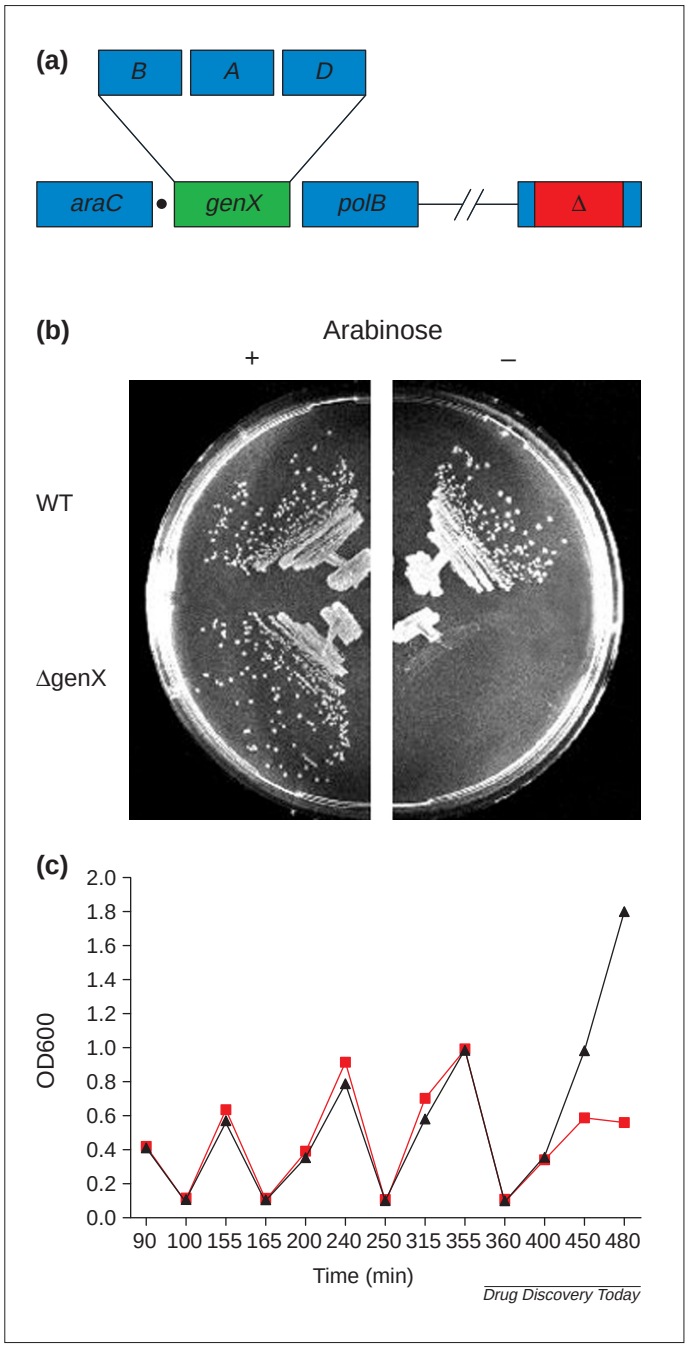

Figure 1. *Construction of a conditional mutant and depletion of the respective gene product. (a) Using plasmid pRDC15 (Ref. 10), an essential gene of interest is positioned at the* ara *locus of* Escherichia coli*, rendering the expression of the gene and its growth dependent on arabinose. In the presence of arabinose, the deletion of the corresponding wild-type (WT) locus can be achieved. (b) For an arabinose-dependent strain of* E. coli*, the presence of arabinose represents the permissive condition whereas its absence equals the non-permissive condition. (c) An essential target protein under investigation can be depleted from the cell by shifting from arabinose in the growth medium to glucose (tight repression of ara promoter, represented by an arrow). Following depletion of the essential protein from OD600, growth arrest is observed for the mutant (red line) and normal growth is measured for the WT control (black line).*

or whether generic stress responses will dominate across most mutants. Analyses by the current authors of several conditional mutants showed that specific deregulation of protein expression can be observed together with common responses. Figure 2 shows one example of conditional mutants in two FUN genes that have been subjected to proteome analysis.

How does the identification of surrogate markers lead to an assay for compound screening? Genes that are identified as surrogate markers can be linked to a reporter gene and its deregulation can then be assayed as a measure of target/pathway inhibition. Such a target-specific whole-cell assay will combine the advantage of selection for cell-permeable compounds inherent in killing assays with the target-specificity and sensitivity achieved by *in vitro* assays. Hence, the important additional advantage is that no detailed functional information concerning the target is required.

Surrogate ligands

Another generic assay development strategy is the identification of surrogate ligands for a given target. Surrogate ligands are short peptides that bind with high affinity (low micromolar to nanomolar) to a target protein and thereby inhibit its function. Based on such surrogate ligands, *in vitro* assays can be designed whereby compound libraries are screened for small molecules that competitively displace the peptide and thus occupy the peptide's binding site on the target protein.

There are several approaches for isolating surrogate ligands from random peptide libraries, the most prominent being phage display technology[22,23]. Here, a library of random peptides is displayed on the surface of filamentous phages. Binding peptides are isolated by a procedure called 'biopanning', where the purified target protein is immobilized on a solid support and phages carrying binding peptides are isolated from the phage library by repeated cycles of adhesion and washing. Sequencing the appropriate segment of the DNA of each captured phage provides the primary sequence of peptides that binds to the target. Isolation of high-affinity binding peptides using phage display has been described for several different protein classes, as will now be summarized.

Agonists that activate the cytokine erythropoietin (EPO) receptor were isolated from random phage display peptide libraries[23]. These agonists were represented by a 14-amino

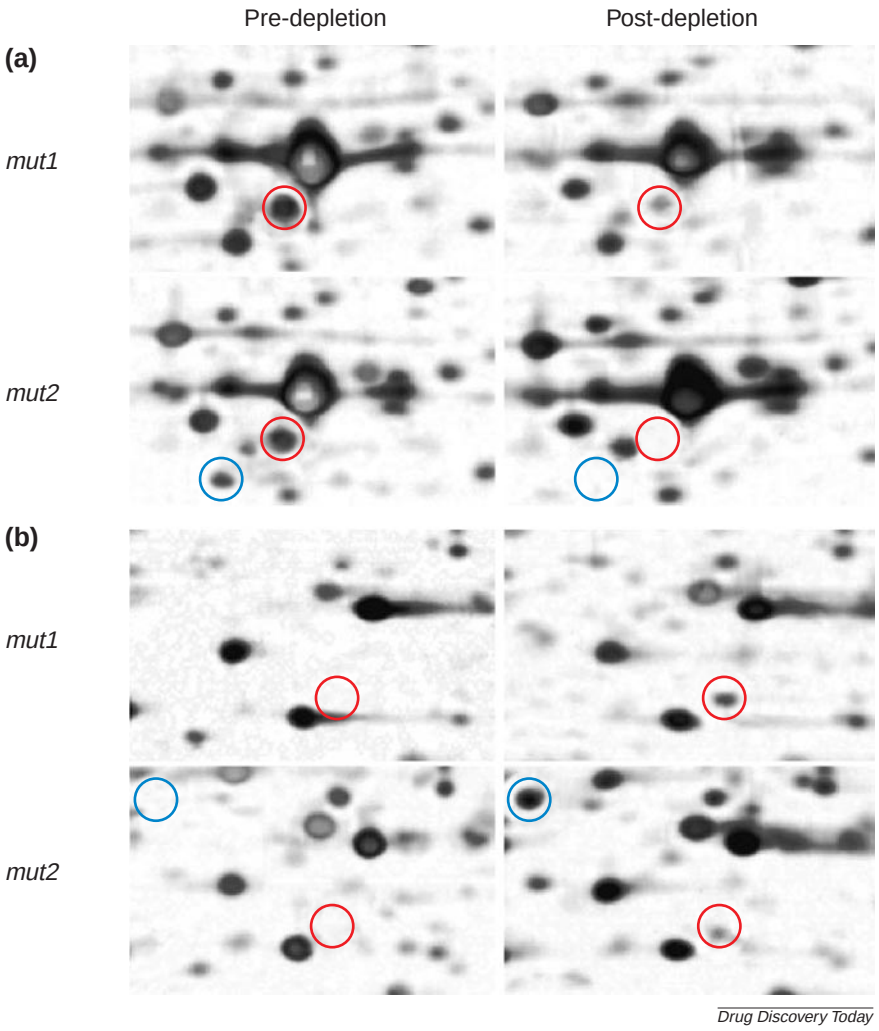

Figure 2. *Comparative proteome analysis of two conditional mutants (*mut1 *and* mut2*) in essential targets of unknown function (TUFs) existing under permissive (pre-depletion) or non-permissive (post-depletion) conditions. (a) and (b) represent different sections from the two-dimensional protein gel. Protein spots up- or down-regulated after a shift to non-permissive conditions in both mutants are circled in red. Those spots that are deregulated only in one of the two mutants are marked by blue circles. The latter proteins are candidates for surrogate markers. However, if only proteome data are available, deregulation has to be confirmed at the gene expression level.*

acid consensus sequence of cyclic peptides containing a disulphide bond. The amino acid sequences of these peptides were not found in the primary sequence of EPO. Furthermore, the signalling pathways activated by these peptides appeared to be identical to those induced by the natural ligand. Another example of isolating small-peptide ligands for a protein of therapeutic relevance was the nuclear hormone receptor for oestrogen[22]. In this study, affinity-selected peptides were used as probes to detect the conformational changes that occur within the oestrogen receptor when it binds various ligands. Furthermore, small-peptide ligands were isolated for a variety of enzymes. These peptides bound to the respective active sites and competitively displaced known inhibitors[24]. Therefore, there is great potential for the isolation of binding peptides that serve as lead structures for many functionally diverse proteins, thus qualifying this strategy for assay development of targets where functional information is lacking. However, these experimental strategies aim to isolate surrogate ligands *in vitro*.

In many instances, it is desirable to validate that the respective peptide(s) have a significant effect on the target protein *in vivo*. Thioredoxin, isolated from *E. coli*, serves as a structural framework for presenting peptides *in vivo*[25]. Using the yeast two-hybrid approach, a peptide aptamer was isolated that binds, and competitively inhibits, the cyclin-dependent kinase 2 (Cdk2)[26]. Expression of this peptide in human cells slows their progression through the G1 phase of the cell cycle. Furthermore, expression of inhibitory peptide aptamers directed against the essential cyclin-dependent kinases, DmCdk1 and DmCdk2 in *Drosophila*, caused adult eye defects typical of those caused by cell cycle inhibition[27]. These findings demonstrate that *in vivo* validation of surrogate peptide ligands is possible even in complex systems such as cultured mammalian cells or *Drosophila*. For bacterial targets that are essential for viability, it can be easily tested whether the peptide ligand binds to a functionally relevant site on the protein (i.e. the respective peptides have to be lethal or inhibit growth on induction of expression in *E. coli*).

Protein–protein interaction technologies

As already described, it can be expected that many validated drug targets from microbial genomics will be

identified in the near future for which detailed functional information is lacking. Although existing technologies enable assay development based on such targets, large-scale functional analysis helps to prioritize targets for screening programs by positioning them in the context of cellular pathways.

One historically very successful method of gaining functional information concerning proteins is the identification of protein–protein interactions using the yeast two-hybrid system[28]. This technology utilizes the fact that the DNA-binding and the *trans*-activating domains of the yeast transcriptional activator Gal4 can be separated, rendering the protein inactive. Activity of the separated domains can be regained by physical proximity. Hence, protein fusions to these domains are generated and physical interaction of the respective protein fusion partners can be identified on the basis of transcriptional activation of the reporter genes[28].

In the standard molecular biology laboratory, performing a yeast two-hybrid screen is very labour-intensive, because of the high frequency of false-positive hits that must be identified for every screen. The two largest problems in laboratory-scale yeast two-hybrid screens are the appearance of autoactivating protein domains fused to the DNA-binding domain vector and the effect of 'sticky' proteins that bind non-specifically to many other proteins[29]. Thus, solutions for efficiently eliminating these problematic clones are essential if large-scale analyses of protein–protein interactions are to be feasible.

Genome Pharmaceuticals Corporation (GPC) has developed an efficient procedure for parallel identification and elimination of autoactivators and sticky proteins. In this method, called PathCode™, replicas of putatively interaction-positive yeast clones (as identified by positive selection) are arrayed on nylon filters and allowed to grow on three different selective media in parallel. One medium selects for the presence of both plasmids and assays for protein interaction using the *lacZ* reporter. The remaining two media each counterselect against one of the two interaction plasmids and positively select for the other. The selection (for and against) is reversed between the two media. Clones that are *lacZ*-positive on any of the counterselective media are false-positive hits. Large numbers of clones are analyzed using digital image analysis procedures based on the software package BiochipExplorer™ (GPC). This technology opens up the possibility of analyzing large numbers of 'baits' or even performing library-versus-library screens for entire bacterial genomes.

Assay formats can be developed that enable screening for small molecules that inhibit and/or disrupt such protein–protein interactions ('reverse two-hybrid screening')[30]. Thus, the investigation of protein–protein interactions using the yeast two-hybrid system not only helps to gain functional information, but might directly be the basis for drug screening. However, it remains to be elucidated how many of the antibacterial drug targets will be involved in protein–protein interactions that can be disrupted by small molecules.

Genomics and drug development

So far, bacterial genomics has had a major impact on identification and validation of targets and on assay development technologies for high-throughput screening. However, genomic technologies will also be crucial to subsequent stages of drug development such as lead optimization, toxicology and clinical studies. One technology with particularly high potential in these areas is the determination of cellular gene expression patterns using DNA arrays[31]. The global changes in gene expression of a given cell as a response to the effect of a compound can be viewed as a reflection of the mechanism by which a compound acts on the cell. In other words, compounds with similar effects on the cell's physiology could produce related changes in gene expression patterns. Using high-density oligonucleotide expression arrays representing nearly all the yeast genes, novel kinase inhibitors isolated from combinatorial chemical libraries were characterized by investigating their effect on yeast gene expression on a genome-wide scale[32]. In this study, two inhibitors with an identical *in vitro* inhibitory spectrum (Cdc28p and Pho85p) were compared with a structurally related compound that showed no inhibitory *in vitro* activity. In this study, 3% of the genes showed a greater than twofold change in transcript level when treated with each of the kinase inhibitors, whereas only 0.03% were affected following treatment with the control compound. Part of the set of genes affected by the inhibitors were loci involved in cell cycle progression and phosphate metabolism, consistent with the spectrum of kinases that were inhibited *in vitro*. Very few of the genes induced by the inhibitors were affected by the control compound, suggesting that many of the drug-sensing mechanisms might respond to signals associated with the function rather than the structure of the drug[32].

As already mentioned, pharmaceutical and biotechnological companies are screening large numbers of chemical libraries for compounds with antibacterial activity. Downstream development of these primary hits into leads would be greatly facilitated if efficient technologies for mode-of-action studies were available. Recently, the effect of the anti-tuberculosis drug isoniazid on global gene expression profiles was investigated[33]. One of the key

findings of this study was the induction of a set of genes encoding the pathway known to be affected by the drug (synthesis of the outer lipid envelope of mycobacteria)[33]. This result shows that it is, in principle, possible to correlate changes in gene expression patterns with a drug's mode of action. However, for analyzing entirely new chemical entities only described by their bactericidal or bacteristatic activity, extensive databases of the effects of reference compounds (of known mode of action) and conditional mutants on transcript profiles will have to be generated. The degree of resolution to which the mechanism of action of a currently undescribed compound can be pinpointed by gene expression profiling remains to be determined.

Another application of the same principle is the assessment of the risk of developing early hits from drug screening. For example, characterization using gene expression profiling of an anti-arteriosclerotic compound that, in cell culture, drastically reduced levels of low-density lipoprotein, revealed that the effect of this compound on gene expression strongly resembled that of a completely different class of compounds that was already shown to be toxic[34]. Thus, resources were not wasted on unsuitable drug candidates. Such approaches are also being discussed in the field of general toxicology and have already been defined as the subdiscipline of toxicogenomics[35].

Together with advances in the human genome project and genomic technologies, the development of anti-infective drugs in the future will become more efficient and targeted by defining patient sub-populations according to their suitability for a given treatment. This era of genetic susceptibility to infectious disease will streamline clinical trials by using more focused patient populations. For example, a small deletion in the gene for the interferon γ (IFNγ) receptor (*IFNGR1*) was found to be associated with dominant susceptibility to infections caused by BCG that is normally avirulent and used as a tuberculosis vaccine[36].

Concluding remarks

The increasing resistance of bacterial pathogens to present -day antibiotics and the lack of a robust pipeline of innovative antimicrobial substances demand innovative and more efficient approaches towards the development of anti-infective drugs. Bacterial genomics has so far greatly increased the rate with which novel targets are identified and validated. Furthermore, the chances that 'second-generation' genomic technologies will accelerate target identification and generic assay development are high. Genomics can also help to streamline later stages of the development of antimicrobials, such as lead optimization, toxicology and clinical trials. However, a concerted innovative application of genomic technologies and chemistries are required to decrease the lag-period between lead identification and marketing of a new drug. In all probability, the genomics-based technologies described in this review will contribute significantly to fulfill these tasks.

Acknowledgements

We thank David Bancroft for his valuable comments on the manuscript.

REFERENCES

1 Hancock, R.E.W. and Knowles, D. (1998) Are we approaching the end of the antibiotic era? *Curr. Opin. Microbiol.* 1, 493–494
2 Heinemann, J.A. (1999) How antibiotics cause antibiotic resistance. *Drug Discovery Today* 4, 72–79
3 Schmid, M.B. (1998) Novel approaches to the discovery of antimicrobial agents. *Curr. Opin. Chem. Biol.* 2, 529–534
4 Myers, P.L. (1997) Will combinatorial chemistry deliver real medicines? *Curr. Opin. Biotechnol.* 8, 701–707
5 Nisbet, L.J. and Moore, M. (1997) Will natural products remain an important source of drug research for the future? *Curr. Opin. Biotechnol.* 8, 708–712
6 Berg, C.M. and Berg, D.A. (1996) Transposable element tools for microbial genetics. In Escherichia coli *and* Salmonella (Neidhardt F.C. ed.), pp. 2588–2612, ASM Press
7 Schmid, M.B. *et al.* (1989) Genetic analysis of temperature-sensitive lethal mutants of *Salmonella typhimurium*. *Genetics* 123, 625–633
8 Akerley, B.J. *et al.* (1998) Systematic identification of essential genes by *in vitro* mariner mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* 95, 8927–8932
9 Devine, S.E. and Boeke, J.D. (1994) Efficient integration of artificial transposons into plasmid targets *in vitro*: A useful tool for DNA mapping, sequencing and genetic analysis. *Nucleic Acids Res.* 22, 3765–3772
10 Hensel, M. *et al.* (1995) Simultaneous identification of bacterial virulence genes by negative selection. *Science* 269, 400–403
11 Arigoni, F. *et al.* (1998) A genome-based approach for the identification of essential bacterial genes. *Nat. Biotechnol.* 16, 851–856
12 Mushegian, A.R. and Koonin, E.V. (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. U. S. A.* 93, 10268–10273
13 Tatusov, R.L. *et al.* (1997) A genomic perspective on protein families. *Science* 278, 631–637
14 Kalman, S. *et al.* (1999) Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. *Nat. Genet.* 21, 385–389
15 Cole, S.T. (1998) Comparative mycobacterial genomics. *Curr. Opin. Microbiol.* 1, 567–571
16 Alm, R.A. *et al.* (1999) Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 397, 176–180
17 Behr, M.A. *et al.* (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 284, 1520–1523
18 Altschul S.F. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410
19 Karp, P.D. *et al.* (1999) Integrated pathway-genome databases and their role in drug discovery. *Trends Biotechnol.* 17, 275–281
20 Hinton, J.C.D. (1997) The *Escherichia coli* genome sequence: The end of an era or the start of the FUN? *Mol. Microbiol.* 26, 417–422

21 Blattner, F.R. *et al.* (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453–1462

22 Paige, L.A. (1999) Estrogen receptor (ER) modulators each induce distinct conformational changes in ER_{α} and ER_{β}. *Proc. Natl. Acad. Sci. U. S. A.* 96, 3999–4004

23 Wrighton, N.C. (1996) Small peptides as potent mimetics of the protein hormone erythropoietin. *Science* 273, 458–464

24 Hyde-DeRuyscher *et al.* (1999) Detection of small-molecule enzyme inhibitors with peptides isolated from phage displayed combinatorial peptide libraries. *Chem. Biol.* 7, 17–25

25 Colas, P. *et al.* (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380, 548–550

26 Cohen, B.A. *et al.* (1998) An artificial cell cycle inhibitor isolated from a combinatorial library. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14272–14277

27 Kolonin, M.G. and Finley, R.L. (1998) Targeting cyclin-dependent kinases in *Drosophila* with peptide aptamers. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14266–14271

28 Colas, P. and Brent, R. (1998) The impact of two-hybrid and related methods on biotechnology. *Trends Biotechnol.* 16, 355–363

29 Bartel, P. *et al.* (1993) Elimination of false positives that arise in using the two-hybrid system. *Biotechniques* 14, 920–924

30 Huang, J. and Schreiber, S.L. (1997) A yeast genetic system for selecting small molecule inhibitors of protein–protein interactions in nanodroplets. *Proc. Natl. Acad. Sci. U. S. A.* 94, 13396–13401

31 Maier, E. *et al.* (1997) Automated array technologies for gene expression profiling. *Drug Discovery Today* 2, 315–324

32 Gray, N.S. *et al.* (1998) Exploiting chemical libraries, structure and genomics in the search for kinase inhibitors. *Science* 281, 533–538

33 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

34 Service, R.F. (1998) Microchip arrays put DNA on the spot. *Science* 282, 396–399

35 Nuwaysir, E.F. *et al.* (1999) Microarrays and toxicology: The advent of toxicogenomics. *Mol. Carcinog.* 24, 153–159

36 Jouanguy, E. *et al.* (1999) A human *IFNGR1* small deletion hotspot associated with dominant susceptibility to mycobacterial infection. *Nat. Genet.* 21, 370–378