# Structure-based discovery of antibacterial drugs

*Katie J. Simmons*[*‡], *Ian Chopra*[*§] & *Colin W. G. Fishwick*[*‡]

Abstract | The modern era of antibacterial chemotherapy began in the 1930s, and the next four decades saw the discovery of almost all the major classes of antibacterial agents that are currently in use. However, bacterial resistance to many of these drugs is becoming an increasing problem. As such, the discovery of drugs with novel modes of action will be vital to meet the threats created by the emergence of resistance. Success in discovering inhibitors using high-throughput screening of chemical libraries is rare. In this Review we explore the exciting opportunities for antibacterial-drug discovery arising from structure-based drug design.

Pharmacophore
A set of structural features in a molecule that are recognized at a receptor site and are responsible for the biological activity of the molecule.

*Antimicrobial Research Centre, University of Leeds. ‡School of Chemistry, University of Leeds. §Institute of Molecular and Cellular Biology, University of Leeds, Leeds, LS2 9JT, UK. Correspondence to C.W.G.F. e-mail: colinf@chem.leeds.ac.uk
doi:10.1038/nrmicro2349

The treatment of bacterial infections through the administration of chemotherapeutic agents, which began in the 1930s, was one of the most profound medical advances of the twentieth century. The origins of almost all of the antibacterial drugs in use today lie in empirical screening programmes to identify inhibitors on the basis of their ability to prevent bacterial growth and can be traced to the so-called 'golden period' of antibacterial-drug discovery between the 1940s and 1970s[1,2]. Subsequent development of these drugs, or agents derived from them, has produced an impressive global reduction in the burden of disease caused by bacterial infection.

Unfortunately, the widespread emergence of resistance to antibiotics in pathogenic bacteria over the past 30 years is now a serious threat to global public health and could undermine the major advances achieved in the treatment of infection[3–9]. Paradoxically, as the problems accompanying the emergence of resistance to existing drugs increase, there has been a decline in the discovery and development of new antibacterials. The reasons for this situation are complex[1,3–5] but, in part, reflect technical difficulties associated with the identification of suitable novel compounds for development as candidate antibacterials.

In the past, many successful antibacterial agents were sourced from empirical screening of natural products or of synthetic chemical libraries[6,7]. However, in recent years empirical screening has not returned suitable pharmacophores for development. Indeed, in the past 40 years only two new structural types, daptomycin and linezolid, have been introduced to the clinic following their discovery using empirical screening methods[8].

The determination of complete bacterial genome sequences and the parallel development of other techniques such as proteomics inspired a new genomics-based approach to drug discovery from the mid 1990s[9]. By March 2009, crystal structure data were available for more than 600 individual proteins derived from bacteria (see the Protein Data Bank and TargetDB databases). Many companies sought to identify novel antibacterial agents from high-throughput screening (HTS) campaigns using purified enzyme targets that were validated by genomic approaches as being essential for the organism. It was thought that the era of exploiting novel natural products or continually modifying existing compounds into improved analogues had passed and that novel agents directed against previously unexploited targets would be identified[10]. The major investment dedicated to the genomic approach for antibacterial-drug discovery reflected the optimism about its likely success. Small-molecule screening approaches had successfully identified promising lead compounds in other therapeutic areas, such as treatments for cancer, diabetes and asthma[11], and consequently the compound collections used for HTS of bacterial targets were largely composed of small synthetic molecules[12]. However, the success rate of the concerted genomic and HTS initiatives has been extremely low, and new strategies are required in order to develop the next generation of antibiotics[6].

The time may now be right to consider fresh approaches to antibacterial-drug discovery. The power of structure-based drug discovery (SBDD) (FIG. 1) has been demonstrated most clearly by the discovery of new therapeutics for HIV/AIDS, a case in which structural knowledge of the HIV protease enabled the successful

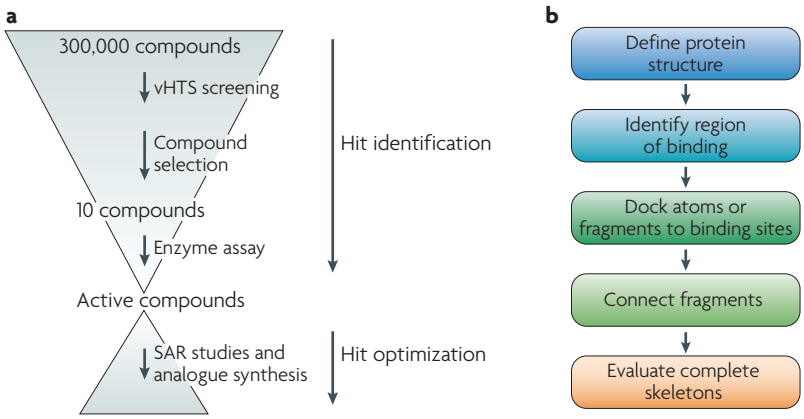


Figure 1 | **Protocols for high-throughput screening docking and *de novo* design.** **a** | A typical selection process for a compound purchased using virtual high-throughput screening (vHTS). **b** | A typical procedure for *de novo* design using programs such as SPROUT. SAR, structure–activity relationship.

design and development of five protease inhibitors that are now commercially available drugs[13–15]. Further successes that owe their origins to SBDD are drugs such as nelfinavir (Viracept; ViiV Healthcare )[16] and amprenavir (Agenerase; GlaxoSmithKline) (compounds 14 and 15; see Supplementary information S1 (figure))[17] for AIDS, zanamivir (Relenza; GlaxoSmithKline)[18] for influenza, the cyclooxygenase 2 (COX2; also known as PTGS2) inhibitors celecoxib (Celebrex; Pfizer)[19] and rofecoxib (Vioxx; Merck — although this was later withdrawn owing to safety concerns) (compounds 16 and 17; see Supplementary information S1 (figure))[20].

Although the potential of SBDD in antibacterial-drug discovery has yet to be fully realized, there are many validated molecular targets already available (FIG. 2; TABLE 1). The growing number of validated targets for which structural information has been obtained makes this approach increasingly attractive[21]. In this Review, we describe the use of SBDD for the development of novel antibacterial agents.

## The principles underpinning SBDD

The starting point for all structure-based design work, whether ligand- or protein-based, is the choice of a suitable target. An antimicrobial-drug target should be essential, have a unique function in the pathogen and exhibit an activity that can be altered by a small molecule. Tools are emerging to prioritize targets on the basis of their predicted suitability for SBDD. A recent study correlated the characteristics of protein-binding pockets with the frequency of the binding ligands identified by nuclear magnetic resonance (NMR)-based screening. This has lead to the creation of an algorithm that can predict the suitability of the binding pocket on the basis of the characteristics that can be identified from high-resolution protein structures, such as the rigidity of the binding site and its hydrophobic character. Such algorithms will allow future researchers to focus their efforts in drug discovery on proteins that are more likely to yield high-affinity ligands[22–24].

When choosing a target enzyme, it is important to be aware of the conformational variations of a protein during ligand binding. This reflects the fact that, unlike the structure of the protein in the crystal, in solution the protein is flexible and can undergo considerable conformational changes on ligand binding.

Until recently, most SBDD programs relied on a single high-resolution protein crystal structure. However, as this is only a snapshot of the protein 'frozen' in one form, it can cause problems when designing ligands, as the biologically active form of the protein may be conformationally different to the crystal form. Many studies emphasize the importance of allowing for protein and ligand flexibility when performing SBDD[25–27]. Unfortunately, modelling of molecular flexibility, especially for the protein, drastically increases the computer time required, often making it prohibitive in terms of time and technical capacity. Programs and techniques such as SLIDE[28], FlexE[29,30] and MCSA–PCR (multiconformation simulated annealing–pseudo-crystallographic refinement)[31] can be used to model protein flexibility, and most programs can now model ligand flexibility as well. A good example of the large degree of conformational flexibility seen during substrate binding is provided by MurA, which is involved in the synthesis of bacterial peptidoglycan. High-resolution crystal structures of MurA have been solved in the 'apo' form and also with bound substrate and substrate analogues. The inhibitor T6361 binds to MurA and blocks the conformational change that is normally induced by substrate binding. The crystal structure of the MurA–T6361 complex shows that the protein adopts a substantially different conformation to that seen in the structure of the MurA–substrate crystal[32–34]. This example demonstrates the challenge of predicting the conformation of the ligand-binding pocket, even if knowledge of the dynamic motion of the protein is available. This situation is not unusual, creating a challenge for both 'virtual HTS' and *de novo* SBDD. Several methods are emerging to overcome these difficulties but have yet to be completely evaluated[35].

Although the Protein Data Bank is rapidly expanding, there is a substantial gap between the number of structures available in the database and the number of known gene sequences. TABLE 2 gives a cross section of the number of individual protein targets for which structural data is available versus the number of genomic ORFs for eight prominent bacterial pathogens.

For cases in which the crystal structure of the particular bacterial protein is not available, construction of a homology model is often possible, provided that a crystal structure is available for a protein with substantial sequence similarity to the protein of interest. Approaches to producing these models vary from purely *ab initio* methods[36] based on only physical and chemical principles to models based on sequence and structural information. Advanced homology models use experimentally determined structures to predict the conformation of another protein that has a similar amino acid sequence. Homology modelling involves four steps: fold assignment, sequence alignment, model building and model refinement. Several computer packages are available to perform this process automatically; for example, the SWISS-MODEL software

Algorithm
A finite sequence of instructions; an explicit, step-by-step procedure for solving a problem, often used for calculation and data processing.

Scoring function
A fast, approximate mathematical method that is used to predict the strength of the non-covalent interaction (also referred to as the binding affinity) between two molecules following docking. Scoring functions are normally parameterized (or trained) against a data set consisting of experimentally determined binding affinities between molecular species that are similar to the species that the user wishes to predict. For predictions of protein–ligand affinities, the tertiary structure of the protein, the active conformation of the ligand and the binding mode must be known.

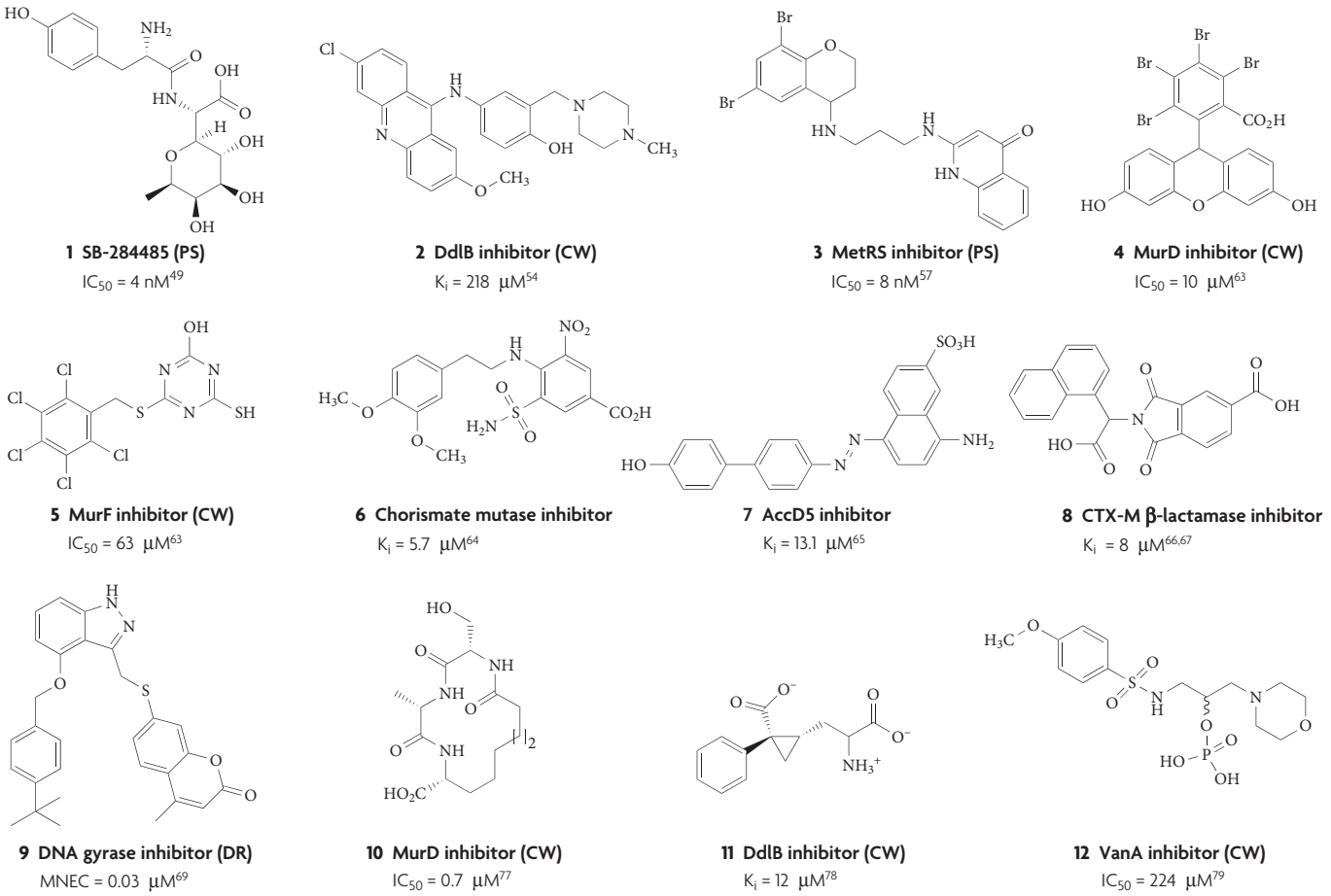


Figure 2 | **Inhibitors designed using structure-based drug discovery.** The structures of some of the key compounds that have been designed using structure-based drug design techniques. Abbreviations given in parentheses denote the molecular target for the compounds: cell wall synthesis (CW), DNA replication (DR), protein synthesis (PS) or RNA synthesis (RS). DdlB, D-alanine–D-alanine ligase; $IC_{50}$, half-maximal inhibitory concentration; $K_i$, inhibition constant; MetRS, methionyl-tRNA synthetase; MNEC, maximal non-effective concentration; VanA, vancomycin resistance protein A.

is a fully automated homology-modelling server available through the ExPASy Proteomics Server[37–40]. CASP (Critical Assessment of Techniques for Protein Structure Prediction) is a competition involving protein structure prediction that has taken place biannually since 1994. The competition provides users of structure prediction servers with an opportunity to assess the quality of the various methods and servers[41,42].

Table 1 | **Targets and classes of antibacterial drugs**

| Drug target | Drug classes |
|---|---|
| Cell wall synthesis | β-Lactams, bacitracin, cycloserine, fosfomycin and glycopeptides |
| Cell membrane integrity | Daptomycin and polymyxins |
| Nucleotide biosynthesis | Sulfonamides and trimethoprim |
| DNA replication | Quinolines, nitrofurans and nitroimidazoles |
| RNA synthesis | Rifamycins |
| Protein synthesis | Aminoglycosides, chloramphenicol, fusidic acid, ketolides, macrolides, oxazolidinones, streptogramins, tetracyclines and mupirocin |

## Methods for SBDD

Three main methods are available to assist in the identification of new putative ligands on the basis of structural information.

In the first approach, termed 'substrate- and known inhibitor-inspired design', the structures of substrates, known inhibitors or cofactors for the particular target enzyme are modified to become inhibitors by maximizing complementary interactions in the target site[13–15,43,44]. In the second approach, databases containing the structures of small molecules are docked into a region of interest *in silico* and scored according to their predicted interactions in the target site. Many programs are available to perform such 'virtual HTS', all of which have different docking algorithms and different scoring functions. The third important approach involves *de novo* design of inhibitor scaffolds. Fragments of molecules are positioned in chosen sites in the target protein and then linked *in silico* to give complete molecules. These molecules can then be scored and ranked for factors such as their predicted binding affinity, their molecular complexity or their expected synthetic

Table 2 | **Structural genomics in bacterial pathogens**

| Species | Total number of ORFs* | Number of ORFs cloned‡ | Number of purified proteins‡ | Number of solved structures in PDB‡ |
|---|---|---|---|---|
| *Escherichia coli* | 5,402 | 4,562 | 1,974 | 233 |
| *Pseudomonas aeruginosa* | 2,230 | 2,262 | 791 | 92 |
| *Haemophilus influenza* | 1,191 | 959 | 264 | 20 |
| *Staphylococcus aureus* | 2,043 | 1,762 | 465 | 35 |
| *Streptococcus pneumoniae* | 1,340 | 1,174 | 364 | 48 |
| *Enterococcus faecalis* | 2,363 | 2,106 | 580 | 63 |
| *Mycobacterium tuberculosis* | 2,315 | 1,094 | 341 | 112 |
| *Helicobacter pylori* | 841 | 537 | 173 | 14 |

PDB, Protein Data Bank. *Data taken from the genome database PEDANT (14 October 2009) . ‡Data taken from TargetDB (14 October 2009) using the built-in query tools and requesting the information under each of the column headings for each species listed.

accessibility (FIG. 1). The resulting designed inhibitor scaffolds are then synthesized in the laboratory and subjected to biological evaluation. Recent examples of these three approaches in the context of antibacterial-drug discovery are discussed below.

### Substrate- and known inhibitor-inspired design

This approach uses the structural modification of known biologically active substrates and natural product-based inhibitors[45]. For example, SB-219383 (compound 18; see Supplementary information S1 (figure)) is a potent and specific inhibitor of bacterial tyrosyl-tRNA synthetase (TyrRS) and was originally identified from the fermentation broth of *Micromonospora* sp. NCIMB 40684 (REFS 46–48). To simplify the chemical structure of the molecule, the bicyclic ring was cleaved to yield a compound that retained potent TyrRS inhibition, and the addition of a butyl ester group led to improved potency. Another simpler molecule, SB-284485 (compound 1; FIG. 2) was also derived without losing inhibitory activity[49], therefore providing an excellent template for further structural modifications.

### Virtual screening approaches

The technique of molecular docking has been well established for some time. However, recent advances in both hardware and software algorithms have made possible the rapid docking of very large collections of small molecules into the chosen molecular target. The speed of some of these programs is such that up to 100,000 molecular structures can be docked per day when using a cluster of parallel processors. However, as with all docking algorithms, the scoring function that is used to assess the validity of specific docking poses is paramount, and each program has its own unique scoring function. These scoring functions will necessarily place different weightings on the various factors involved in ligand binding. As none of these functions is considered faultless, a consensus scoring approach is the best way to identify potential lead molecules. Consensus scoring uses several scoring functions to predict binding affinity. If a compound is predicted to bind tightly to a chosen protein using several docking algorithms, this provides higher confidence in the prediction. Among the most prominent docking programs currently available for virtual screening are AutoDock, Glide (Schrödinger), GOLD (The Cambridge Crystallographic Data Centre, Cambridge, UK), DOCK and eHiTS (SimBioSys Inc., Toronto, Canada). A selection of the most successful programs applied to the discovery of potential new antibacterials is discussed below (see also Supplementary information S2 (table) for additional resources).

***AutoDock.*** AutoDock is a suite of automated docking tools designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known three-dimensional structure. AutoDock consists of two main programs: 'autodock' performs the docking of the ligand to a set of grids describing the target protein, which are pre-calculated by 'autogrid'. In addition, a graphical front-end tool, AutoDockTools, is available to set up, visualize and analyse the results of dockings performed using AutoDock.

The most recent version of the software, AutoDock 4.0, uses the AMBER force field as well as a free-energy scoring function based on a linear-regression analysis and a diverse set of protein–ligand complexes with known inhibition constants.

AutoDock uses a genetic algorithm to generate a range of docking poses that can be clustered according to their energetic similarity. Several studies have shown that in docking calculations the most populated clusters of the docked-ligand conformation are better predictors of the native state than the lowest-energy cluster[50–53].

A recent example of the application of AutoDock to antibacterial-drug discovery is the structure-based virtual screening of the UK National Cancer Institute (NCI) 'diversity set' of 2,000 compounds using a crystal structure of D-alanine–D-alanine ligase (DdlB) from *Escherichia coli*, a key enzyme in peptidoglycan biosynthesis. Docking results were obtained as a list of compounds ranked according to their mean estimated binding affinity to the protein. This was determined by the calculated average free energy of binding for the most populated cluster of docked poses. Using this approach, the top 130 compounds were tested in an *in vitro* assay for inhibition of *E. coli* DdlB and several hits were identified. Three of these hits have novel scaffolds; two

Force field
The functional form and parameter sets used to describe the potential energy of a system of particles.

Free energy
The calculated difference between the internal energy of a system and the product of its absolute temperature and entropy.

Linear-regression analysis
Any approach to modelling the relationship between one or more variables denoted *Y* and one or more variables denoted *X*, such that the model depends linearly on the unknown parameters to be estimated from the data.

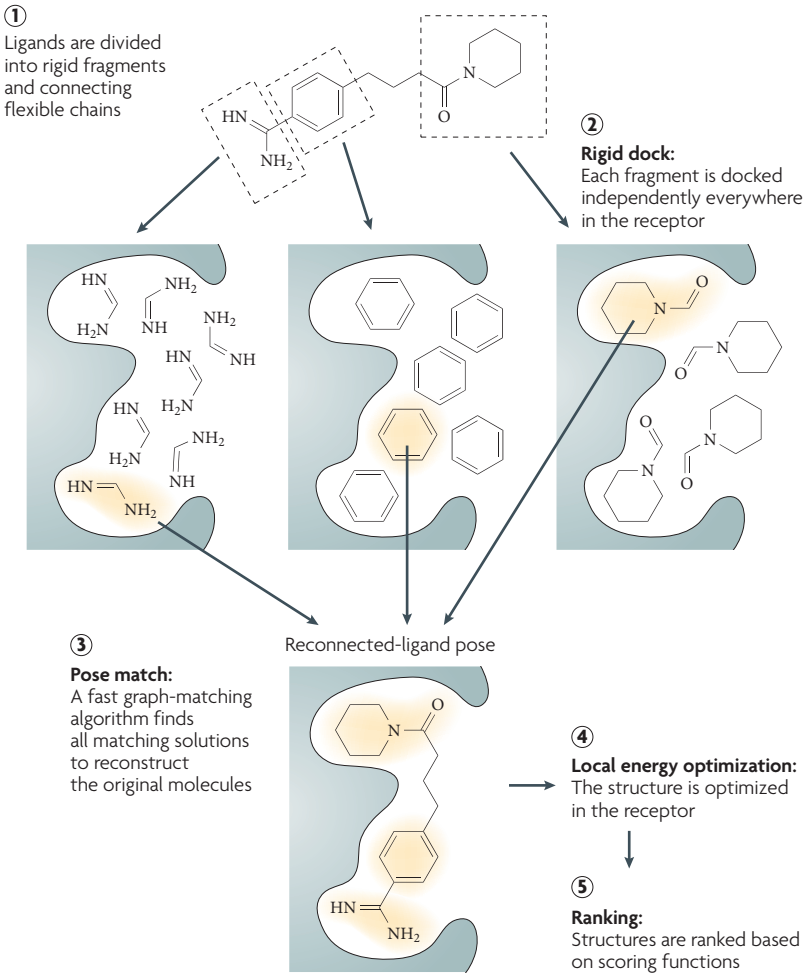


Figure 3 | **The eHiTS docking strategy.** In eHiTS, ligands are divided into rigid fragments and connecting flexible chains (step 1). These fragments are docked individually into the binding site of the target receptor (step 2) and a fast graph-matching algorithm finds all matching solutions to reconstruct the original molecule (step 3), which can then be optimized (step 4), scored and ranked (step 5).

Toxicity
The degree to which a substance is able to damage an exposed organism.

ADME
Absorption, distribution, metabolism and excretion. An acronym in pharmacokinetics and pharmacology describing the disposition of a pharmaceutical compound in an organism.

Log *P*
The partition coefficient (*P*), usually expressed as a log value, is a measure of the hydrophobic character of a substance, such that *P* = concentration of the substance in octanol / concentration of the substance in water.

of these (compounds 19 and 20; see Supplementary information S1 (figure)) are competitive inhibitors, competing with ATP, with inhibition constant ($K_i$) values in the low micromolar range, whereas the third (compound 2; FIG. 2) inhibited the enzyme in a non-competitive manner. Additionally, compounds 2 and 20 possessed some antimicrobial activity and are therefore promising hits for further optimization[54].

***Catalyst.*** Discovery Studio (Accelrys) provides a suite of software that has many applications, including SBDD, toxicity prediction, protein modelling and virtual HTS. The built-in virtual HTS screening tool, Catalyst, has been used in the development of novel antibacterial agents.

A high-throughput approach to crystallography identified several small-molecule inhibitors of the *Staphylococcus aureus* methionyl-tRNA synthetase (MetRS), an enzyme that catalyses the highly specific attachment of methionine onto cognate methionine-specific tRNA[55]. The identified inhibitors all interact with an important amino acid, Asp51, and with two specific hydrophobic pockets. On the basis of this information, a four-point pharmacophore model was constructed using Catalyst. This model was used as a search query against the diverse compound collection from ChemDiv, which contains around 250,000 compounds. A total of 461 molecules from the Catalyst search were identified as potential hits. These were docked into the *S. aureus* MetRS structure and re-scored using the LigandFit scoring function[56]. Of the 31 compounds that were subsequently selected for biological testing, 22 displayed greater than 50% enzyme inhibition at a concentration of 100 μM, which is a hit rate of 71%. The most potent inhibitors, compounds 3 and 21, are shown in FIG. 2 and Supplementary information S1 (figure), respectively. This method provides an efficient way of finding new leads from a known active compound and compares favourably to random biological assaying of large compound libraries, which give hit rates of between 0.1% and 0.5%, on average[57].

***eHiTS.*** This is a recently developed virtual HTS software package that takes individual compounds from a large library and calculates the optimal conformation that each of these ligands can adopt in a targeted protein cavity. The program then calculates a score for each structure according to the geometries of the ligand and the complementarities of 'surface points' on the receptor and ligand. Complementary surface points receive a positive score, whereas repulsive surface points receive a penalty score. Additional terms are used in the final scoring function to further reflect all factors involved in binding, such as steric clashes, depth of the cavity, solvation, conformational-strain energy of the ligand, intramolecular interactions in the ligand, and entropy loss due to 'frozen' rotatable bonds[58–60].

eHiTS takes a unique approach to the docking problem, having an innovative docking algorithm and a novel system for the scoring function. The approach involves breaking ligands into rigid fragments and the connecting flexible chains and then docking each rigid fragment to every possible place in the cavity (FIG. 3). Recent examples of the use of eHiTS include screens for inhibitors of the *Mycobacterium tuberculosis* shikimate kinase (AroK), inhibitors of TyrRS and inhibitors of MurD and MurF.

*M. tuberculosis* AroK catalyses the phosphorylation of shikimate to shikimate-3-phosphate. An eHiTS screening protocol was used recently to identify potential inhibitors of AroK[61]. Screening compounds were extracted from the FAF-Drugs (Free ADME/tox filtering) collection to give 214,492 compounds following filtering using ADME and toxicity filters. FAF-Drugs is an online service that allows users to process their own compound collections through simple ADME/tox-filtering rules such as molecular mass, polar surface area, log *P* or number of rotatable bonds. From this collection of over 200,000 compounds, docking in the AroK active site using eHiTS identified 644 small molecules that were predicted to bind more tightly to AroK than the natural substrate and that are potential inhibitors with half-maximal inhibitory concentration ($IC_{50}$) values in

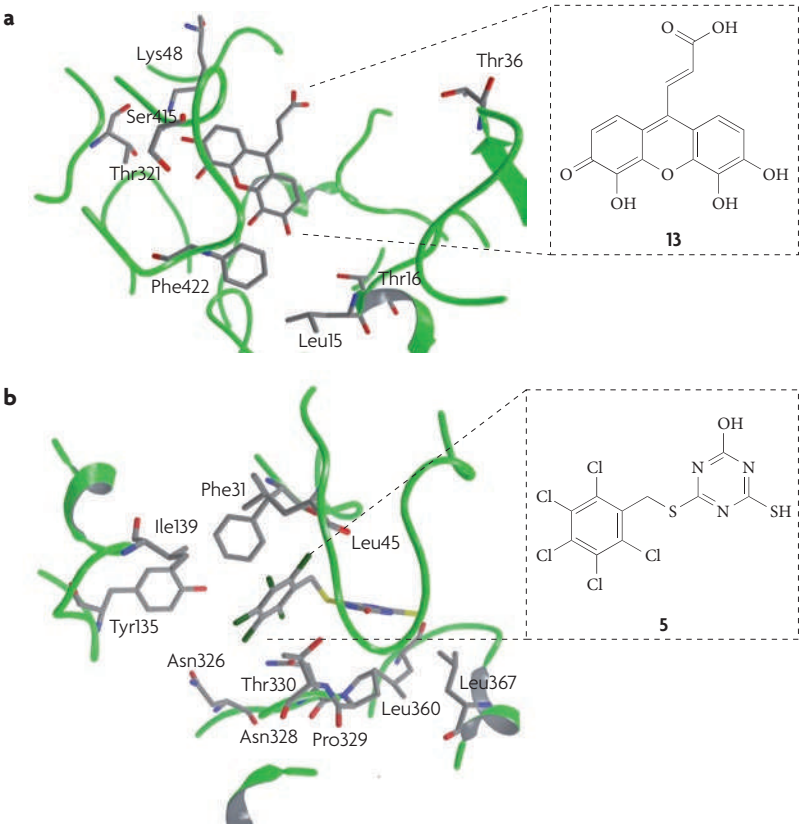


Figure 4 | **Inhibitors designed using eHiTS. a** | An inhibitor of MurD (compound 13), discovered using eHiTS, docked in the MurD active site. The inhibitor binds in the same pocket as the substrate (not shown) and makes key interactions with the labelled residues. The MurD backbone is shown as green ribbons. **b** | An inhibitor of MurF (compound 5), discovered using eHiTS, docked in the active site of the enzyme. The inhibitor interacts with key labelled residues, and the protein backbone is shown as green ribbons.

the low-micromolar range. The top 200 compounds were examined in further detail using the graphical user interface CheVi (SimBioSys) to determine key interactions and shape complementarity with the binding pocket of the enzyme.

In another study, a series of 340 potential inhibitors of TyrRS were docked into the active sites of both the human and staphylococcal TyrRS proteins using eHiTS version 5.3. This study sought to find molecules that bind more strongly to the staphylococcal enzyme than to the human enzyme. Therefore, the scores for each ligand docked into the two separate active sites were compared, and the ligands with the greatest difference in predicted binding affinity for the two enzymes were then examined in further detail. This led to the identification of ten potential inhibitors with a stronger affinity for staphylococcal TyrRS than for the human protein. These compounds have not yet been screened against the bacterial enzyme, but this approach would seem to offer potential for the design of selective antibacterial agents[62].

The Mur enzymes are essential for steps in peptidoglycan biosynthesis in bacteria[63]. eHiTS was used to screen 1,990 compounds from the NCI diversity set using crystal structures of MurD and MurF as targets, and the 50 top-scoring compounds for each enzyme were selected for biological evaluation[63]. For MurD, 4 of the top 50 compounds showed $IC_{50}$ values of below 250 μM (3 of these are shown here as compounds 4, 13 and 21; FIGS 2,4a and Supplementary information S1 (figure)).

Only one of the compounds selected from the MurF screen (compound 5; FIGS 2,4b) showed significant inhibitory activity (that is, had residual activity at 250 μM below their $IC_{50}$ values). The lower hit rate for MurF might be attributed, in part, to the compounds being assayed against *E. coli* MurF rather than *Streptococcus pneumoniae* MurF, the crystal structure of which was used to perform the eHiTS screening runs.

***UNITY.*** UNITY, a module in the SYBYL (Tripos) molecular modelling software, is a search and analysis system for exploring chemical and biological databases. It can be used for locating compounds that match a pharmacophore or fit a receptor site. UNITY's two-dimensional searching capabilities offer exact, substructure and similarity searching. Conformationally flexible three-dimensional searching rapidly finds molecules that can satisfy queries regardless of the conformation stored in a database. Structural queries may be based on complete molecular structures, molecular fragments, pharmacophore models or the receptor site.

Using UNITY, a virtual screen was carried out for inhibitors of *M. tuberculosis* chorismate mutase[64], an enzyme that catalyses the conversion of chorismate to prephenate in the tyrosine and phenylalanine biosynthesis pathway. Starting from a known inhibitor of a homologous enzyme, a three-dimensional pharmacophore search of a database of 15,000 compounds was performed. Of the 15 highest-scoring molecules, 4 demonstrated inhibition in the enzyme assay. The most potent molecule (compound 6) is shown in FIG. 2.

***DOCK.*** DOCK is an open-source molecular-docking software package that is frequently used in SBDD. Historically, the DOCK algorithm addressed rigid-body docking using a geometric-matching algorithm to superimpose the ligand onto a negative image of the binding pocket. Important features that improved the algorithm's ability to find the lowest-energy binding mode have been added over recent years, including force field-based scoring, on-the-fly optimization, an improved matching algorithm for rigid-body docking and an algorithm for flexible-ligand docking. In one study using DOCK, over 4 million compounds were screened for predicted binding to the biotin- or propionyl CoA-binding pockets of AccD5[65], an essential *M. tuberculosis* acyl CoA carboxylase carboxyltransferase subunit. One of the nine top-scoring compounds identified by DOCK (compound 7; FIG. 2) had an $IC_{50}$ of 10 μM against the enzyme. DOCK has also been used for a fragment-based screening programme against the β-lactamase CTX-M-9 (REFS 66,67). A fragment subset of 67,489 compounds were docked into the active site of the enzyme, and compounds were selected from the top of the ranking list. From the 69 fragments investigated (which included compound 8; FIG. 2), 10 exhibited $IC_{50}$ values in the micromolar

On-the-fly optimization
A fast, dynamic procedure used to make a system or design as effective or functional as possible.

DNA gyrase
An enzyme that unwinds DNA so that the DNA can duplicate.

range — a hit rate of 14.5%. The same group also carried out a similar study looking for fragment inhibitors of the β-lactamase AmpC[68]. A library of 137,639 fragments from the ZINC database were docked, and 48 fragments that were highly ranked were purchased, 23 of which had $K_i$ values ranging from 0.7 to 9.2 mM. The 48% hit rate compared very favourably with lead-like docking and HTS against the same enzyme[68].

### *De novo* design using protein structures

*De novo* ligand design has been continually improving since its invention in the early 1990s, reflecting both the tremendous developments in computational power and the continual improvement of efficient computational algorithms (FIG. 1). Most of the *de novo* design tools follow a similar pipeline in terms of operation. The most widely used *de novo* design programs are LUDI and SPROUT (Keymodule Ltd, Leeds, UK) (discussed below), SkelGen (De Novo Pharmaceuticals Ltd), Flux[69], GANDI and BOMB (Cemcomco, Madison, Connecticut, USA). Several of these *de novo* methods have been used to design novel antibacterial agents.

***LUDI.*** LUDI, another module in Discovery Studio, constructs possible new ligands for a given protein of known three-dimensional structure. This approach is based on rules about energetically favourable non-bonded-contact geometries between functional groups of the protein and the ligand; these rules are derived from a statistical analysis of the crystal packing of organic molecules.

Small fragments are docked into the protein's binding site in such a way that hydrogen bonds and ionic interactions can be formed with the protein and hydrophobic pockets are filled with lipophilic groups derived from the ligands. The program can then append further fragments onto a previously positioned molecular core. It is also possible to link several fragments together, through bridge fragments, to form a complete molecule. All putative ligands retrieved or constructed by LUDI are then scored using a simple scoring function that was fitted to experimentally determined binding constants of protein–ligand complexes[70].

*In silico* screening for potential inhibitors of DNA gyrase was performed using LUDI and Catalyst to screen the available-chemicals database (ACD) and part of the Roche compound inventory (a total of 350,000 compounds). LUDI was used to dock small 'needle' molecules into the binding site of the gyrase. Needle screening is a technique that can be used to identify low-molecular-mass inhibitors (with a molecular mass of <300 daltons) that can penetrate into deep and narrow channels and subpockets. The LUDI searches led to the identification of around 150 weak inhibitors (including molecules 23–25; see Supplementary information S1 (figure)). X-ray crystallography verified binding of the needles to the ATP-binding site. Following structure–activity relationship (SAR) studies to probe the structural requirements of the validated needles, a series of indazole-based inhibitors was identified, with the most potent compound (compound 9; FIG. 2) being ten times as active as novobiocin, with a maximal non-effective concentration (MNEC) of 0.03 μM[70].

***SPROUT.*** SPROUT uses a fragment-joining technique to generate structures that fit the steric and electronic constraints of a specific receptor site or pharmacophore hypothesis. During structure generation, atoms or fragments of molecules are placed at each of the 'target sites' and are then linked to produce molecular 'skeletons'. Molecular fragments are represented by templates, in which atoms are labelled purely by their hybridization state and are represented by vertices, and bonds are labelled as single, double, triple or aromatic. Several actual molecular fragments can be produced from each template by replacing the vertices with any element that can adopt the appropriate hybridization state.

The structure generation phase joins templates together to produce skeletons. Each skeleton represents several molecules, because each component template can represent several molecular fragments. Skeleton generation begins by selecting a template and positioning it at a target site so that it satisfies the steric requirements associated with that particular site. New molecular templates can then be added to further target sites, chosen by the user. These docked molecular templates are then joined using 'spacer' templates, again under the full control of the user. Once skeletal generation is complete, the resulting structures can be clustered and sorted using a range of parameters specified by the user. These include overall molecular properties such as molecular complexity and estimated binding affinity, as well as more detailed filtering options such as the presence or absence of certain molecular features. In this way, unwanted structures can be identified and discarded, leaving only structures that meet the requirements of the user[71–77].

SPROUT was used to design a series of novel macrocyclic inhibitors of the bacterial cell wall biosynthesis enzyme MurD[78]. SPROUT revealed that the binding cavity contains a hydrophobic pocket that is not used by the natural substrate. Simplification of the MurD substrate generated models of possible macrocyclic inhibitors that omit much of the sugar portion of the natural

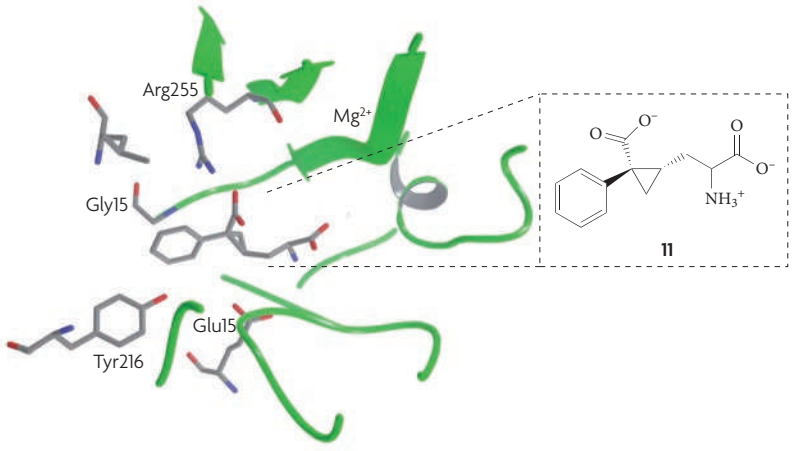


Figure 5 | **SPROUT-designed inhibitor modelled in D-alanine–D-alanine ligase.** The inhibitor (compound 11) interacts with key labelled residues in the D-alanine–D-alanine ligase (DdlB) of *Escherichia coli*, the backbone of which is in green.

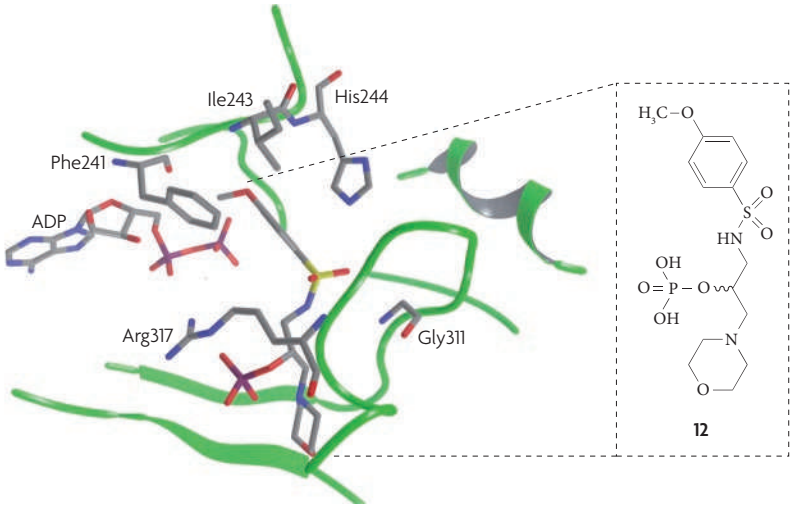


Figure 6 | **Inhibitor of vancomycin resistance protein A, designed using SPROUT.** The inhibitor (compound 12) binds to vancomycin resistance protein A (VanA) (green backbone) through the same pocket as the phosphate substrate. Key interactions with labelled VanA residues are shown.

substrate. A series of macrocyclic molecules based on these SPROUT-designed molecular skeletons were synthesized and evaluated as inhibitors of *E. coli* MurD. It was found that compound 10 (FIG. 2) exhibited the highest affinity for MurD.

The SPROUT software suite was also used to design a novel cyclopropyl-based inhibitor of the bacterial enzyme DdlB[79]. Enzymological evaluation of the designed inhibitor (compound 11; FIG. 2) showed that this molecule (as a diastereomeric mixture) inhibits the activity of DdlB with an apparent $K_i$ value of 12.5 (± 0.1) μM (FIGS 2,5).

The Van enzymes are responsible for resistance to vancomycin in *S. aureus* and *Enterococcus* spp. SPROUT was used in conjunction with the X-ray crystal structure of the *Enterococcus faecium* D-alanine–D-lactate ligase (VanA) to devise new VanA inhibitors based on a hydroxyethylamine template that was designed to mimic the tetrahedral reaction intermediates[80]. The most active compound (compound 12; FIG. 2) had an $IC_{50}$ of 224 μM against VanA. Owing to the similar topology of the active sites of VanA and DdlB, the set of generated compounds was also tested against *E. coli* DdlB. Compound 12 exhibited an $IC_{50}$ value of 110 μM against DdlB (FIGS 2,6).

Recent reports from groups in the pharmaceutical industry have commented on the difficulty in identifying inhibitors of bacterial RNA polymerase, a crucial enzyme in the construction of RNA chains from DNA templates, using traditional HTS approaches[9]. SPROUT was used to create the first *de novo*-designed RNA polymerase inhibitors (for example, compound 26; see Supplementary information S1 (figure))[81] using the X-ray crystal structure of RNA polymerase from *Thermus aquaticus*. Preliminary experimental data indicate that these molecules do inhibit bacterial RNA polymerase.

The production of inhibitors that are active in the low-micromolar concentration range after pure *de novo* design is now commonplace using this approach, and such inhibitors provide an excellent starting point for further optimization.

***SPROUT-LeadOpt.*** SPROUT-LeadOpt (Keymodule) is specifically designed to aid in the structural optimization process following the discovery of a lead molecule. Unlike SPROUT, SPROUT-LeadOpt is not a *de novo* design tool; rather, it produces structures that are similar to known lead molecules and that have improved predicted binding affinities. Core extension allows the user to explore different regions of a receptor's binding site by adding monomers to a core molecule. SPROUT-LeadOpt uses a set of synthetic-chemistry rules to identify 'reactive functional groups' on a core molecule. It then systematically generates extended structures in a combinatorial fashion by carrying out virtual synthetic chemistry using common synthetic reactions and a database of monomer structures. 'Monomer replacement' uses retrosynthetic fragmentation to identify 'monomers' that are present in one or more bound ligands. New molecular structures are then created by replacing these monomers with structurally related molecules taken from a 'monomer library'. The power of this approach was recently shown by the development of inhibitors of dihydroorotate dehydrogenase[82]. Although the enzyme is not a bacterial target protein, this work demonstrates the power of SPROUT-LeadOpt in the development of new lead scaffolds.

## Future perspectives

Undoubtedly, structure-based approaches for the identification of new inhibitors of both classical and novel bacterial target proteins will increase in future years. However, it should also be remembered that improved understanding of the molecular basis by which existing antibacterial agents act may provide insights for new SBDD approaches. For example, the structural basis of the interaction between the bacterial ribosome and several established antibiotics, such as tetracycline, streptomycin and chloramphenicol, has been elucidated in recent years[83–85].

In addition to ribosome–antibiotic structures, several high-resolution crystal structures have been obtained that show the binding of the antibiotics rifampicin and myxopyronin to bacterial RNA polymerase[86,87]. These structures are helpful not only for determining the mechanism of action of these antibiotics but also for the development of novel inhibitors or, in a ligand-based approach, for searching for structurally simplified analogues of these macrocyclic, natural product-based inhibitors. The binding position of the macrocycles in the crystal structures of RNA polymerase can also be used to identify suitable binding pockets in this large enzyme. Indeed, it may be possible to design molecules that simultaneously inhibit two or more functional sites in the enzyme, which could minimize the potential for the development of resistance.

The growing global burden of bacterial resistance to established antibacterial drugs is creating important but unmet medical needs, and in some cases isolates have been described that are resistant to all previously

appropriate chemotherapeutic agents. In recent years, HTS of chemical libraries has dominated the search for new antibacterial-drug leads. Unfortunately this has not been successful, and a consensus is now emerging that new approaches are required[9]. SBDD is still in its infancy. However, the technology described in this Review may provide the springboard for future, more successful attempts to discover the new classes of antibacterial drugs that will be required to fight infection in the twenty-first century and beyond.


1. Chopra, I., Hesse, L. & O'Neill, A. J. Exploiting current understanding of antibiotic action for discovery of new drugs. *J. Appl. Microbiol.* **92**, S4–S15 (2002).
2. Knowles, D. J. C. New strategies for antibacterial drug design. *Trends Microbiol.* **5**, 379–383 (1997).
3. Spellberg, B. *et al.* The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America. *Clin. Infect. Dis.* **46**, 155–164 (2008).
4. Talbot, G. H. *et al.* Bad bugs need drugs: an update on the development pipeline from the Antimicrobial Availability Task Force of the Infectious Diseases Society of America. *Clin. Infect. Dis.* **42**, 657–668 (2006).
5. Chopra, I. *et al.* Treatment of health-care-associated infections caused by Gram-negative bacteria: a consensus statement. *Lancet Infect. Dis.* **8**, 133–139 (2008).
6. Overbye, K. M. & Barrett, J. F. Antibiotics: where did we go wrong? *Drug Discov. Today* **10**, 45–52 (2005).
7. Newman, D. J., Cragg, G. M. & Snader, K. M. The influence of natural products upon drug discovery. *Nat. Prod. Rep.* **17**, 215–234 (2000).
8. Shlaes, D. M. An update on tetracyclines. *Curr. Opin. Investig. Drugs* **7**, 167–171 (2006).
9. Payne, D. J., Gwynn, M. N., Holmes, D. J. & Pompliano, D. L. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nature Rev. Drug Discov.* **6**, 29–40 (2007).
**GlaxoSmithKline share their experience of evaluating more than 300 genes and 70 HTS campaigns over a period of 7 years; they analyse what they have learned and how this has influenced their antibacterials strategy.**
10. Fernandes, P. Antibacterial discovery and development — the failure of success? *Nature Biotech.* **24**, 1497–1503 (2006).
11. Golebiowski, A., Klopfenstein, S. R. & Portlock, D. E. Lead compounds discovered from libraries. *Curr. Opin. Chem. Biol.* **5**, 273–284 (2001).
12. Chan, P. F., Holmes, D. J. & Payne, D. J. Finding the gems using genomic discovery: antibacterial drug discovery strategies – the successes and the challenges. *Drug Discov. Today Ther. Strateg.* **1**, 519–527 (2004).
13. Dorsey, B. D. *et al.* L-735,524: the rational design of a potent and orally bioavailable HIV protease inhibitor. *J. Med. Chem.* **37**, 3443–3451 (1994).
14. Erickson, J. *et al.* Design, activity and 2.8Å crystal structure of a C2 symmetrical inhibitor complexed to HIV-1 protease. *Science* **249**, 527–533 (1990).
15. Roberts, N. A. *et al.* Rational design of peptide-based HIV proteinase inhibitors. *Science* **248**, 358–361 (1990).
16. Kaldor, S. W. *et al.* Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *J. Med. Chem.* **40**, 3979–3985 (1997).
17. Highleyman, L. New anti-HIV drugs in development. *BETA* **12**, 7–8 (1999).
18. McCauley, J. Relenza. *Curr. Biol.* **9**, R796–R796 (1999).
19. Stratton, M. S. & Alberts, D. S. Current application of selective COX-2 inhibitors in cancer prevention and treatment. *Oncology (Williston Park)* **16**, 37–51 (2002).
20. Mardini, I. A. & FitzGerald, G. A. Selective inhibitors of cyclooxygenase-2: a growing class of anti-inflammatory drugs. *Mol. Interv.* **1**, 30–38 (2001).
21. Schmid, M. B. Crystallizing new approaches for antimicrobial drug discovery. *Biochem. Pharmacol.* **71**, 1048–1056 (2006).
**This article describes the use of structural genomics, in the form of SBDD, to overcome the bottlenecks in the antimicrobial pipeline, such as identifying good leads.**
22. Halgren, T. A. Identifying and characterising binding sites and assessing druggability. *J. Chem. Inf. Model.* **49**, 377–389 (2009).
23. Villoutreix, B. O. *et al.* Free resources to assist structure-based virtual ligand screening experiments. *Curr. Protein Pept. Sci.* **8**, 381–411 (2007).
**This report provides links for *in silico* ADME/tox prediction tools, compound collections, some ligand-based methods, characterization and simulation of three-dimensional targets, and homology modelling tools, as well as links to software for druggable-pocket predictions, active-site comparisons, analysis of macromolecular interfaces, protein-docking tools to help identify binding pockets, and protein– ligand docking and scoring methods.**
24. Cheng, A. C. *et al.* Structure-based maximal affinity model predicts small-molecule druggability. *Nature Biotech.* **25**, 71–75 (2007).
25. Carlson, H. A. & McCammon, J. A. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* **57**, 213–218 (2000).
**An explanation of the need to accommodate an ensemble of protein configurations in drug design and the computational methods available for generating and manipulating that dynamic information.**
26. Jain, A. N. Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J. Comput. Aided Mol. Des.* **21**, 281–306 (2007).
27. Zhao, Y. & Sanner, M. F. FLIPDock: docking flexible ligands into flexible receptors. *Proteins* **68**, 726–737 (2007).
28. Schnecke, V., Swanson, C. A., Getzoff, E. D., Tainer, J. A. & Kuhn, L. A. Screening a peptidyl database for potential ligands to proteins with side-chain flexibility. *Proteins* **33**, 74–87 (1998).
29. Claussen, H., Buning, C., Rarey, M. & Lengauer, T. FlexE: efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **308**, 377–395 (2001).
30. Claussen, H., Buning, C., Rarey, M. & Lengauer, T. in *Rational Approaches to Drug Design: Proceedings of the 13th European Symposium on Quantitative Structure-Activity Relationships* (eds Höltje, H.-D. & Sippl, W.) 324–333 (Heinrich-Heine Universität, Düsseldorf, 2001).
31. Ota, N. & Agard, D. A. Binding mode prediction for a flexible ligand in a flexible pocket using multi-conformation simulated annealing pseudo crystallographic refinement. *J. Mol. Biol.* **314**, 607–617 (2001).
32. Skarzynski, T. *et al.* Structure of UDP-*N*-acetylglucosamine enolpyruvyl transferase, an enzyme essential for the synthesis of bacterial peptidoglycan, complexed with substrate UDP-*N*-acetylglucosamine and the drug fosfomycin. *Structure* **4**, 1465–1474 (1996).
33. Eschenburg, S., Kabsch, W., Healy, M. L. & Schonbrunn, E. A new view of the mechanisms of UDP-*N*-acetylglucosamine enolpyruvyl transferase (MurA) and 5-enolpyruvylshikimate-3-phosphate synthase (AroA) derived from X-ray structures of their tetrahedral reaction intermediate states. *J. Biol. Chem.* **278**, 49215–49222 (2003).
34. Eschenburg, S. *et al.* A novel inhibitor that suspends the induced fit mechanism of UDP-*N*-acetylglucosamine enolpyruvyl transferase (MurA). *J. Biol. Chem.* **280**, 14070–14075 (2005).
35. DeWeese-Scott, C. & Moult, J. Molecular modeling of protein function regions. *Proteins* **55**, 942–961 (2004).
36. Yang, J. S., Chen, W. W., Skolnick, J. & Shakhnovich, E. I. All-atom ab initio folding of a diverse set of proteins. *Structure* **15**, 53–63 (2007).
37. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201 (2006).
38. Bordoli, L. *et al.* Protein structure homology modeling using SWISS-MODEL workspace. *Nature. Protoc.* **4**, 1–13 (2009).
39. Kopp, J. & Schwede, T. Automated protein structure homology modeling: a progress report. *Pharmacogenomics* **5**, 405–416 (2004).
40. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* **31**, 3381–3385 (2003).
41. Jones, T. A. & Kleywegt, G. J. CASP3 comparative modeling evaluation. *Proteins* **37** (Suppl. 3), 30–46 (1999).
42. Tramontano, A., Leplae, R. & Morea, V. Analysis and assessment of comparative modeling predictions in CASP4. *Proteins* **45**, 22–38 (2001).
43. Chan, D. C. M., Laughton, C. A., Queener, S. F. & Stevens, M. F. G. Structural studies on bioactive compounds. 34. Design, synthesis, and biological evaluation of triazenyl-substituted pyrimethamine inhibitors of *Pneumocystis carinii* dihydrofolate reductase. *J. Med. Chem.* **44**, 2555–2564 (2001).
44. Varney, M. D. *et al.* Crystal structure-based design and synthesis of benz[*cd*]indole-containing inhibitors of thymidylate synthase. *J. Med. Chem.* **35**, 663–676 (1992).
45. Schmid, M. B. Seeing is believing: the impact of structural genomics on antimicrobial drug discovery. *Nature Rev. Microbiol.* **2**, 739–746 (2004).
46. Berge, J. M. *et al.* Inhibitors of bacterial tyrosyl tRNA synthetase: synthesis of four stereoisomeric analogues of the natural product SB-219383. *Bioorg. Med. Chem. Lett.* **10**, 1811–1814 (2000).
47. Berge, J. M. *et al.* Synthesis and activity of analogues of SB-219383: novel potent inhibitors of bacterial tyrosyl tRNA synthetase. *J. Antibiot. (Tokyo)* **53**, 1282–1292 (2000).
48. Stefanska, A. L. *et al.* SB-219383, a novel tyrosyl tRNA synthetase inhibitor from a *Micromonospora* sp. I. Fermentation, isolation and properties. *J. Antibiot. (Tokyo)* **53**, 345–350 (2000).
49. Brown, P. *et al.* Synthetic analogues of SB-219383. Novel *C*-glycosyl peptides as inhibitors of tyrosyl tRNA synthetase. *Bioorg. Med. Chem. Lett.* **11**, 711–714 (2001).
50. Prasad, J. C., Goldstone, J. V., Camacho, C. J., Vajda, S. & Stegeman, J. J. Ensemble modeling of substrate binding to cytochromes p450: analysis of catalytic differences between CYP1A orthologs. *Biochemistry* **46**, 2640–2654 (2007).
51. Källblad, P., Mancera, R. L. & Todorov, N. P. Assessment of multiple binding modes in ligand-protein docking. *J. Med. Chem.* **47**, 3334–3337 (2004).
**This study quantifies the effect of considering multiple docking solutions on the success rate of obtaining the crystallographic binding mode.**
52. Limongelli, V. *et al.* Ensemble-docking approach on BACE-1: pharmacophore perception and guidelines for drug design. *ChemMedChem* **2**, 667–678 (2007).
53. Kozakov, D., Clodfelter, K. H., Vajda, S. & Camacho, C. J. Optimal clustering for detecting near-native conformations in protein docking. *Biophys. J.* **89**, 867–875 (2005).
54. Kovac, A. *et al.* Discovery of new inhibitors of D-alanine:D-alanine ligase by structure-based virtual screening. *J. Med. Chem.* **51**, 7442–7448 (2008).
55. Finn, J., Stidham, M., Hilgers, M. & Kedar, G. C. Identification of novel inhibitors of methionyl-tRNA synthetase (MetRS) by virtual screening. *Bioorg. Med. Chem. Lett.* **18**, 3932–3937 (2008).
56. Venkatachalam, C. M., Jiang, X., Oldfield, T. & Waldman, M. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graph. Model.* **21**, 289–307 (2003).
57. Dove, A. Drug screening—beyond the bottleneck. *Nature Biotech.* **17**, 859–863 (1999).
58. Zsoldos, Z. eHiTS: exhaustive flexible ligand docking with customizable scoring function tailored to protein families. *SimBioSys Inc.* [online], http://www.simbiosys.ca/science/presentations/acs-2004/abstract.html (2004).

59. Zsoldos, Z., Johnson, A. P., Simon, A., Szabo, I. & Szabo, Z. eHiTS: novel algorithm for fast, exhaustive flexible ligand docking and scoring. *SimBioSys Inc.* [online], http://www.simbiosys.com/science/presentations/acs-2002/abstract.html (2002).
60. Zsoldos, Z., Reid, D., Simon, A., Sadjad, B. S. & Johnson, A. P. eHITS: an innovative approach to the docking and scoring function problems. *Curr. Protein Pept. Sci.* **7**, 421–435 (2006).
61. Segura-Cabrera, A. & Rodriguez-Perez, M. A. Structure-based prediction of *Mycobacterium tuberculosis* shikimate kinase inhibitors by high-throughput virtual screening. *Bioorg. Med. Chem. Lett.* **18**, 3152–3157 (2008).
62. Eitner, K. *et al.* eHiTS-to-VMD interface application. The search for tyrosine-tRNA ligase inhibitors. *J. Chem. Inf. Model.* **47**, 695–702 (2007).
63. Turk, S. *et al.* Discovery of new inhibitors of the bacterial peptidoglycan biosynthesis enzymes MurD and MurF by structure-based virtual screening. *Biorg. Med. Chem.* **17**, 1884–1889 (2009).
64. Agrawal, H., Kumar, A., Bal, N. C., Siddiqi, M. I. & Arora, A. Ligand based virtual screening and biological evaluation of inhibitors of chorismate mutase (Rv1885c) from *Mycobacterium tuberculosis* H37Rv. *Bioorg. Med. Chem. Lett.* **17**, 3053–3058 (2007).
65. Lin, T. W. *et al.* Structure-based inhibitor design of AccD5, an essential acyl-CoA carboxylase carboxyltransferase domain of *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103**, 3072–3077 (2006).
66. Babaoglu, K. *et al.* Comprehensive mechanistic analysis of hits from high-throughput and docking screens against β-lactamase. *J. Med. Chem.* **51**, 2502–2511 (2008).
67. Morandi, S., Morandi, F., Caselli, E., Shoichet, B. K. & Prati, F. Structure-based optimization of cephalothin-analogue boronic acids as β-lactamase inhibitors. *Biorg. Med. Chem.* **16**, 1195–1205 (2008).
68. Teotico, D. G. *et al.* Docking for fragment inhibitors of AmpC β-lactamase. *Proc. Natl Acad. Sci. USA* **106**, 7455–7460 (2009).
69. Fechner, U. & Schneider, G. Flux (1): a virtual synthesis scheme for fragment-based de novo design. *J. Chem. Inf. Model.* **46**, 699–707 (2006).
70. Boehm, H. J. *et al.* Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. *J. Med. Chem.* **43**, 2664–2674 (2000).
71. Boda, K. & Johnson, A. P. Molecular complexity analysis of *de novo* designed ligands. *J. Med. Chem.* **49**, 5869–5879 (2006).
72. Gillet, V., Johnson, A. P., Mata, P., Sike, S. & Williams, P. SPROUT: A programme for structure generation. *J. Comput. Aided Mol. Des.* **7**, 127–153 (1993).
73. Gillet, V. J., Myatt, G., Zsoldos, Z. & Johnson, A. P. SPROUT, HIPPO and CAESA: Tools for *de novo* structure generation and estimation of synthetic accessibility. *Perspect. Drug Discov. Des.* **3**, 34–50 (1995).
74. Gillet, V. J. *et al.* SPROUT: recent developments in the *de novo* design of molecules. *J. Chem. Inf. Comput. Sci.* **34**, 207–217 (1994).
75. Law, J. M. S. *et al.* Validation of the SPROUT *de novo* design program. *J. Mol. Struct.* **666**, 651–657 (2003).
76. Mata, P. *et al.* SPROUT: 3D structure generation using templates. *J. Chem. Inf. Comput. Sci.* **35**, 479–493 (1995).
77. Zsoldos, Z., Szabo, I., Szabo, Z. & Johnson, A. P. Software tools for structure based rational drug design. *J. Mol. Struct.* **666**, 659–665 (2003).
78. Horton, J. *et al.* Macrocyclic inhibitors of bacterial cell wall biosynthesis enzyme MurD. *Bioorg. Med. Chem. Lett.* **13**, 1557–1560 (2003).
79. Besong, G. E. *et al.* A *de novo* designed inhibitor of D-Ala-D-Ala ligase from *E. coli*. *Angew. Chem. Int. Ed. Engl.* **44**, 6403–6406 (2005).
80. Sova, M. *et al.* Design and synthesis of new hydroxyethylamines as inhibitors of D-alanyl-D-lactate ligase (VanA) and D-alanyl-D-alanine ligase (DdlB). *Bioorg. Med. Chem. Lett.* **19**, 1376–1379 (2009).
81. Agarwal, A. K., Johnson, A. P. & Fishwick, C. W. G. Synthesis of *de novo* designed small-molecule inhibitors of bacterial RNA polymerase. *Tetrahedron* **64**, 10049–10054 (2008).
82. Davies, M. *et al.* Structure-based design, synthesis and characterization of inhibitors of human and *Plasmodium falciparum* dihydroorotate dehydrogenases. *J. Med. Chem.* **52**, 2683–2693 (2009).
83. Schlunzen, F. *et al.* Structural basis for the interpretation of antibiotics with the peptidyl transferase centre in eubacteria. *Nature* **413**, 814–821 (2001).
84. Pioletti, M. *et al.* Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine and IF3. *EMBO J.* **20**, 1829–1839 (2001).
85. Franceschi, F. & Duffy, E. M. Structure-based drug design meets the ribosome. *Biochem. Pharmacol.* **71**, 1016–1025 (2006).
86. Mukhopadhyay, J. *et al.* The RNA polymerase "switch region" is a target for inhibitors. *Cell* **135**, 295–307 (2008).
87. Belogurov, G. A. *et al.* Transcription inactivation through local refolding of the RNA polymerase structure. *Nature* **457**, 332–338 (2009).



**Acknowledgements**
Work in the authors' laboratories is supported by grants from the UK Medical Research Council (MRC), the UK Biotechnology and Biological Sciences Research Council (BBSRC), Heart Research UK and the European Community's Seventh Framework Programme FP7 (under grant agreement HEALTH-F4-2007-201924, EDICT Consortium).


**Competing interests statement**
The authors declare no competing financial interests.

## DATABASES

**Entrez Genome Project:**
http://www.ncbi.nlm.nih.gov/genomeprj
*Enterococcus faecium* | *Escherichia coli* | *Mycobacterium tuberculosis* | *Staphylococcus aureus* | *Streptococcus pneumoniae* | *Thermus aquaticus*
**UniProtKB:** http://www.uniprot.org
AccD5 | AmpC | AroK | DdlB | MurA | MurD | MurF | VanA

## FURTHER INFORMATION

**Colin W. G. Fishwick's homepage:** http://www.chem.leeds.ac.uk/People/Fishwick.html
**ACD:** http://cds.dl.ac.uk/cds/datasets/orgchem/suppliers.html
**AutoDock:** http://autodock.scripps.edu
**BOMB:** http://www.cemcomco.com/GenMol.html
**CheVi:** http://www.simbiosys.ca/chevi/index.html
**Discovery Studio:** http://accelrys.com/products/discovery-studio/index.html
**DOCK:** http://dock.compbio.ucsf.edu
**eHiTS:** http://www.simbiosys.ca/ehits/index.html
**ExPASy Proteomics Server:** http://expasy.org
**FAF-Drugs:** http://bioserv.rpbs.univ-paris-diderot.fr/Help/FAFDrugs.html
**GANDI:** http://www.biochem-caflisch.unizh.ch/download
**Glide:** http://www.schrodinger.com/products/14/5
**GOLD:** http://www.ccdc.cam.ac.uk/products/life_sciences/gold
**PENDANT:** http://pedant.gsf.de
**Protein Data Bank:** http://www.pdb.org/pdb/home/home.do
**SkelGen:** http://www.denovopharma.com/page2.asp?PageID=484
**SPROUT:** http://www.keymodule.co.uk/SPROUT.html
**SPROUT-LeadOpt:** http://www.keymodule.co.uk/SPROUT-LeadOpt.html
**SYBYL:** http://www.tripos.com/index.php?family=modules,SimplePage,,,&page=sybyl_sybyl
**TargetDB:** http://targetdb.pdb.org
**ZINC:** http://zinc.docking.org

## SUPPLEMENTARY INFORMATION

**See online article:** S1 (figure) | S2 (table)

**ALL LINKS ARE ACTIVE IN THE ONLINE PDF**