

X-ray analysis and NMR have defined the 3D structures of many drug targets and their complexes with substrate analogues and other inhibitors. Has this really transformed the way we discover drugs?

Keynote review: Structural biology and drug discovery

Miles Congreve, Christopher W. Murray and Tom L. Blundell

It has long been recognized that knowledge of the 3D structures of proteins has the potential to accelerate drug discovery, but recent developments in genome sequencing, robotics and bioinformatics have radically transformed the opportunities. Many new protein targets have been identified from genome analyses and studied by X-ray analysis or NMR spectroscopy. Structural biology has been instrumental in directing not only lead optimization and target identification, where it has well-established roles, but also lead discovery, now that high-throughput methods of structure determination can provide powerful approaches to screening.

- ▶ Discussion of the use of structural biology in drug discovery began over 35 years ago, with the advent of knowledge of the 3D structures of globins, enzymes and polypeptide hormones. Early ideas in circulation were the use of 3D structures to guide the synthesis of ligands of haemoglobin to decrease sickling or to improve storage of blood [1,2], the chemical modification of insulins to increase half-lives in circulation [3] and the design of inhibitors of serine proteases to control blood clotting [4]. An early and bold venture was the UK Wellcome Foundation programme focussing on haemoglobin structures established in 1975 [1]. However, X-ray crystallography was expensive and time consuming. It was not feasible to bring this technique 'in-house' into industrial laboratories, and initially the pharmaceutical industry did not embrace it with any real enthusiasm.

In time, knowledge of the 3D structures of target proteins found its way into thinking about drug design. Although, in the early days, structures of the relevant drug targets were usually not available directly from X-ray crystallography, comparative models based on homologues began to be exploited in lead optimization in the 1980s [5]. An example was the use of aspartic protease structures to model renin, a target for antihypertensives [6]. It was recognized that 3D

TOM L. BLUNDELL

Tom L. Blundell is Sir William Dunn Professor of Biochemistry and Chair of Biological Sciences, University of Cambridge. He previously held positions at Oxford and Sussex Universities, and the Bernal Chair of Crystallography at Birkbeck College, University of London. Blundell is a Member of EMBO and Academia Europaea, a Fellow of the Royal Society and of the Academy of Medical Sciences. He has Honorary Fellowships at Linacre and Brasenose Colleges, University of Oxford, and at Birkbeck College, London, a Professorial Fellowship at Sidney Sussex, University of Cambridge, and Honorary Doctorates from 14 universities. Tom Blundell was CEO of AFRC and BBSRC (1991–1996). He is President of the Biosciences Federation and Chairman of the Royal Commission on Environmental Pollution. Blundell has consulted with a number of large pharmaceutical companies, including Pfizer, Parke Davis, SKB, Zeneca and Glaxo. From 1997 to 2004, he was a non-executive director of Celltech and Chairman of their Science and Technology Committee. Blundell is a co-founder of Astex Technology, which is concerned with the discovery of new medicines.



MILES CONGREVE



CHRISTOPHER W. MURRAY



Miles Congreve
Christopher W. Murray
Astex Technology,
436 Cambridge Science Park,
Milton Road,
Cambridge,
UK, CB4 0QA

Tom L. Blundell*
Department of Biochemistry,
University of Cambridge,
Tennis Court Road,
Cambridge,
UK, CB2 1QW
*e-mail:
tom@cryst.bioc.cam.ac.uk

structures were useful in defining topographies of the complementary surfaces of ligands and their protein targets, and could be exploited to optimize potency and selectivity [7]. Eventually, crystal structures of real drug targets became available; AIDS drugs, such as Agenerase and Viracept, were developed using the crystal structure of HIV protease [8,9] and the flu drug Relenza was designed using the crystal structure of neuraminidase [10]. There are now several drugs on the market that originated from this structure-based design approach; Hardy and Malikayil [11] list >40 compounds that have been discovered with the aid of structure-guided methods and that have entered clinical trials [11]. Seven of these compounds had, in mid-2003, become approved and marketed drugs (Table 1). It is arguable that additional drugs could also be included in Table 1, for example, other marketed HIV protease inhibitors, but, for the purposes of this review, Table 1 is intended to represent the overall contributions of structural biology for various target classes. Gleevec was launched in 2001 for the treatment of chronic myelogenous leukaemia. It has been reported that the crystal structure of Gleevec complexed with BCR-ABL fusion protein kinase assisted in understanding retrospectively how structural mutations in ABL can circumvent the anticancer activity of the compound [12,13]. Another anticancer kinase inhibitor drug, Tarceva, was approved by the FDA in 2004 for the treatment of locally advanced or metastatic non-small cell lung cancer. In addition, Exanta, which was approved in Europe in 2004 as the first oral anticoagulant since warfarin, was marketed almost 60 years ago, although this compound has not been approved in the USA.

The structure-based design methods used to optimize these leads into drugs are now often applied much earlier in the drug discovery process. Protein structure is used in target identification and selection (the assessment of the 'druggability' or tractability of a target), in the identification of hits by virtual screening and in the screening of fragments. Additionally, the key role of structural biology during lead optimization to engineer increased affinity and selectivity into leads remains as important as ever. Each of these topics will be outlined, using the field of kinase drug discovery as an example of the role of structure in lead optimization.

Target identification from sequence–structure homology recognition

Structural information about proteins can give clues on the membership of families and superfamilies. Although this approach is often classed as 'fold recognition', it is more properly termed 'sequence–structure homology recognition'. The recognition of HIV protease as a distant member of the pepsin–renin superfamily and the subsequent modelling of its 3D structure and design of inhibitors elegantly exemplifies the applicability of this approach [14,15]. In general, putative relatives are identified, the sequences aligned and the 3D structures modelled. This

is usually helpful in identifying binding sites and molecular function, if key residues are conserved.

Such approaches can be classified either as profile methods or threading. The profile methods introduce structural information into traditional sequence comparison algorithms, either using structure-dependent propensities [16,17] or substitution matrices and gap penalties [18–20]. By contrast, threading fits a probe sequence onto the backbone of a known structure, evaluating the compatibility between the sequence and the proposed structure by means of a set of empirical potentials that are derived from well-resolved protein structure data [21]. This method remains a powerful tool for fold recognition, but sequence–structure comparison strategies using profiles offer better homology recognition performance [22]. Combined algorithms have been reported; for example, GenTHREADER uses the sequence comparison method to generate the sequence–structure alignment and then evaluates the alignment using threading potentials [23].

Once a homologue of known structure has been identified, it can be modelled using a variety of comparative (homology) modelling procedures, for example, those that use a fragment-assembly approach, such as Composer [24] or SWISS-MODEL [25], or alternatively a restraint-based approach, such as Modeller [26]. These give good models if the sequence identity is >30% but the accuracy falls off sharply when it is lower, mainly because of the difficulties in obtaining good alignments, in predicting shifts of core residues and in building loops [27].

The use of homology models is important in the development of kinase inhibitors [28]. Over 500 kinases have been identified in the human genome and, so far, only ~50 of these have had their structures determined. In cases where the structure is not known, homology models have been useful for optimizing the affinity of inhibitors [28]. This not only includes increasing affinity for the kinase target(s) of interest but also decreasing affinity for related kinases and identifying kinases to be used in selectivity screens on the basis of structural homology in the region where the inhibitor binds [29]. Homology models can also be successfully used as a starting point for virtual screening.

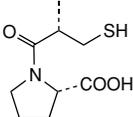
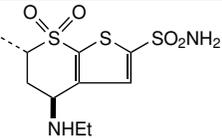
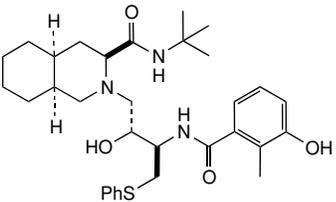
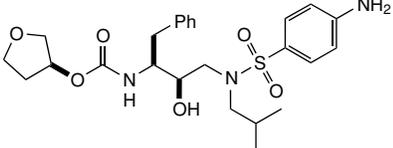
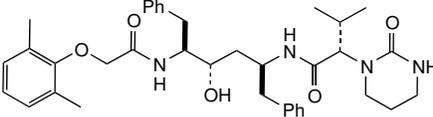
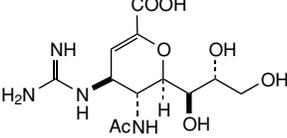
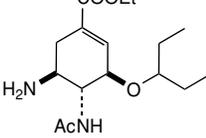
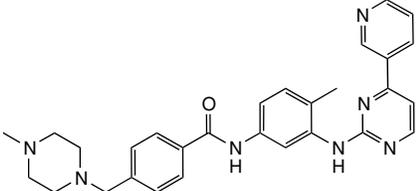
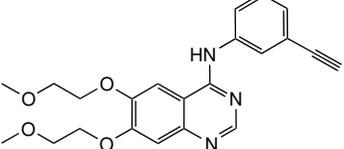
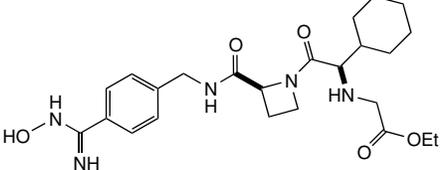
Structural genomics and drug targets

Although these comparative approaches have proved helpful, an experimental structure will usually be more accurate. The possibility of using high-throughput crystallography for defining structures of all gene products in an organism, known as structural genomics, has recently become a reality and there are several worldwide initiatives to define 3D structures of representative protein family members in several genomes [30–33]. Structures defined by these structural genomics initiatives should be useful not only as a basis for ligand design but also for homology recognition on the basis of structure–structure comparisons.

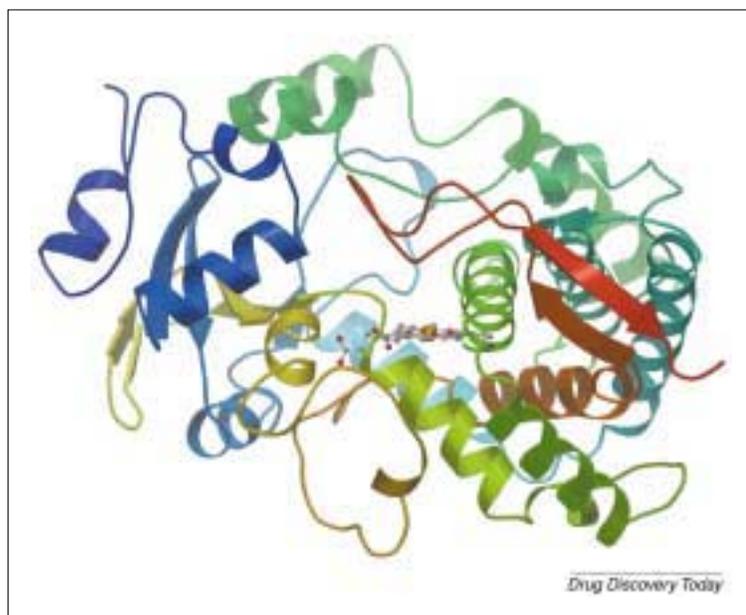
There is intense interest in automating all steps of protein crystallography – expression, characterization, crystallization

TABLE 1

Drugs derived from structure-based approaches

Trade name ^a	Generic name	Structure	Mechanism	Disease area	Date launched	Company	Refs
Capoten	Captopril		Angiotensin-converting enzyme	Hypertension	1981	Bristol-Myers Squibb	[118]
Trusopt	Dorzolamide		Carbonic anhydrase	Glaucoma	1995	Merck	[119]
Viracept	Nelfinavir		HIV protease	HIV/AIDS	1999	Agouron (Pfizer) and Lilly	[120]
Agenerase	Amprenavir		HIV protease	HIV/AIDS	1999	Vertex and GSK	[121]
Aluviran	Lopinavir		HIV protease	HIV/AIDS	2000	Abbott	[122]
Relenza	Zanamivir		Neuraminidase	Influenza	1999	Monash University and GlaxoSmithKline	[123]
Tamiflu	Oseltamivir		Neuraminidase	Influenza	1999	Gilead and Roche	[124]
Gleevec	Imatinib		BCR-ABL	Chronic myelogenous leukaemia	2001	Novartis	[125,126]
Tarceva	Erlotinib		EGFR	Non-small cell lung cancer	2004	OSI and Genentech	[127]
Exanta	Ximelagatran		Thrombin	Venous thromboembolic events	2004 (Europe only)	AstraZeneca	[128]

^aAll trade names are registered trademarks.

**FIGURE 1**

Crystal structure of human cytochrome P450 3A4. Overall fold of P450 3A4. The N-terminus is shown in blue and the C-terminus in red, with intermediate parts of the sequence in green and yellow. The haem, which is the site of oxidative metabolism of many drugs, is depicted as a ball and stick model at the centre. An understanding of the binding site of P450s and protein–ligand complexes of drug molecules bound in the active site of these proteins is improving understanding of drug metabolism. Figure generated using Astex Viewer™ (Astex Technology).

and structure determination: these have been reviewed extensively and they are only summarized here. For example, technology developments for genomic sequencing have provided superb tools that enable DNA sequencing and expression-construct design to be automated with small sample volumes [34]. Although expression in *Escherichia coli* is quick, low cost, high yielding and lacks post-translational modification, many eukaryotic proteins express poorly or form inclusion bodies that must be refolded [30–33]. Significant improvements have been obtained using insect cell or yeast expression, enabling many key drug targets, such as protein kinases, to be produced routinely. Cell-free protein expression is being used to tackle gene products that are toxic to the host organism or difficult to express [35]. Although the purification of proteins is much more dependent on specialist knowledge, inclusion of an affinity tag has led to the establishment of generic purification approaches in this area [36,37].

There have been significant and impressive advances in crystallization through provision of commercial crystallization solutions and the use of liquid-handling robots that can dispense preformulated screens or create custom ones [38–40]. The dispensing of protein drops has taken advantage of new technologies, reducing dispensing volumes to 25–100 nl of protein for each crystallization condition [41]. X-ray data collection has been greatly facilitated by the introduction of crystal mounting and alignment robots, such as ACTOR [42], enabling unattended collection of data 24 h a day [43]. Many exciting

developments in high-throughput structure determination are occurring in the structural genomics and synchrotron centres [30–33], including attempts to solve the phase problem using mathematical approaches that do not require labelling the protein with a heavy atom or anomalous scatterer [44,45] and automatic interpretation of the electron density, where resolution is better than 2.3 Å [46]. Attempts are being made to describe the structural heterogeneity in the form of anisotropic motion and discrete conformational substrates, often of functional importance [47].

As a result of these initiatives, structural biology has been able to tackle more difficult problems than have previously been amenable to structural elucidation. A good example of this has been the progress on cytochrome P450s (P450s). P450s are responsible for the majority of the primary metabolism of drug molecules in the human body. The mammalian P450s linked with metabolism are membrane-associated and, despite significant efforts for many years, the structure of the first mammalian P450 was not solved until 2000 [48]. Subsequently, improvements in manipulating the mammalian enzymes, coupled with the power of modern structural biology, have enabled the determination of several human P450s known to be important in drug metabolism, culminating in the determination of the P450 3A4 crystal structure and its co-complexes with relevant ligands (Figure 1) [49,50]. The P450 3A4 isoform is implicated in the metabolism of ~50% of marketed drugs and its structure determination might herald a new role for structure-based design in the modulation of the metabolic properties of candidate drug molecules.

Identifying ligand binding regions

As structural genomics projects produce more 3D structures, with many of the proteins recognized only from their gene sequences, it is becoming increasingly important to develop computational methods that will identify sites involved in productive intermolecular interactions that might give clues about functions and binding sites. If computational methods can give an indication of a function and also a binding region, for example, by homology, then these can be tested by site-directed mutagenesis, which has proved to be a powerful structure-based approach.

Sequence motif databases, such as PROSITE [51], identify specific residues that are likely to be involved in function, but 3D descriptors of functional sites have an advantage in that the sites themselves are usually made from discontinuous regions of the protein sequence [52]. There have been several attempts to predict functional and/or interaction sites computationally, for example, by finding steric strain or other types of high-energy conformations that often occur at active sites [53,54] or through identifying clefts that can accommodate ligands [55]. Almost all protein functional sites arise through mutation and

Darwinian selection, and hence they will be the most highly conserved regions of a protein [56–58]. The most widely used method based on evolutionary conservation of sequence is ‘evolutionary trace’ [59], in which residues that are conserved are highlighted on the structure. However, restraints leading to conservation of sequence can arise from protein function and structure and a recent approach differentiates evolutionary restraints on protein function from those on sequence and structure [60].

Identification of hits and leads

Hit identification is a costly and resource intensive exercise. For more than a decade, random HTS, in which a large compound collection is interrogated in a high-density assay format, has been the technique most often used to identify hits. The objective has been to identify compounds with IC_{50} values lower than 10 μ M. A great deal of infrastructure and logistics are required, not only for the screening process but also for the assembly, storage and maintenance of a quality controlled compound file. Furthermore, in an attempt to increase the number and quality of the compounds available for screening, the field of combinatorial chemistry was born during the 1990s. Many large companies have invested in establishing infrastructure and expertise in the synthesis of large libraries, using solid-phase chemistry approaches and solution-phase libraries coupled with high-throughput purification platforms [61–63].

Despite these intensive efforts, generating quality hits remains challenging. Hits-to-leads chemistry, following on from HTS, is required to assess and prioritize the results and to synthesize close analogues of active structures to establish outline SARs and build confidence in a hit series. However, many active wells in a screen against a target do not always guarantee the identification of a high-quality lead series [64].

The relatively disappointing results of HTS against specific classes of targets have led to the concept of ‘target tractability’, in which knowledge of the historical successes of a class of targets is used to assess the probability of success for other family members. This assessment is based on the number and drug-likeness of ligands for a target class, as well as knowledge of the binding sites of family members using protein structure information [65]. As a consequence of this prioritization of particular protein families as candidates for screening, a move to a more knowledge-driven overall approach has evolved. Rather than performing a large, costly, high-throughput screen, a more focused screening campaign is often more appropriate for the most tractable targets for which there is information about ligands, and perhaps also protein structure.

First, it will be possible to use *in silico* screening approaches (virtual screening and *de novo* design) to select a subset of samples from the larger compound file or to design a new lead using knowledge of previously identified ligands. Second, the energies of the combinatorial

chemist can be spent producing target-class-specific compound libraries, directed by known ligands or by knowledge of the active site of at least one member of the protein family [63]. Finally, a biophysical screen of small molecules, or ‘fragments’, might be possible, from which a lead can be derived using knowledge of how the fragment binds in the active site of the target. Several examples of each of these approaches will be outlined here.

Structure-guided design and virtual screening

Selecting (or designing) compounds *in silico* that bind to a protein active site is difficult. First, the *in silico* method must solve the docking problem by finding the optimum binding orientation for the compound in the active site of the protein. This means that it must predict the correct ligand conformation and orientation, in addition to any protein movement that is induced by the ligand, although for most applications of docking the protein is assumed to be rigid. Many methods and programs have been developed and tested in docking applications [66]. Docking accuracy is usually assessed by the ability to reproduce the experimentally determined binding mode of a ligand as the highest-ranking solution starting from a random ligand geometry but using the correct conformation of the protein. Currently, state-of-the-art docking programs correctly dock ~70–80% of ligands when tested on large sets of protein–ligand complexes [67,68].

The second challenge is that the *in silico* method must score the compound so that its relative affinity can be judged versus other compounds. In virtual screening, a large number of commercially available (or synthetically accessible) compounds are docked and ranked by *in silico* methods and the highest-ranking compounds are selected for acquisition (or synthesis) and experimentally tested for activity against the target protein. The method would work best with an accurate and rapidly calculated scoring function but, to date, only crude approximations are available [69].

Despite these difficulties, *in silico* methods are useful and influential in structure-guided design. For example, virtual screening will be effective if there is a significant enrichment of true hits in a selected subset of compounds compared with a subset selected via another mechanism and this condition does not require a highly accurate scoring function or a rigorous treatment of receptor flexibility. Successful examples of virtual screening in the identification of novel hits and the demonstration of significant enrichment have been described (two examples are presented in Figure 2) and there have been many other reports in the literature [69–74].

The identification of novel and potent inhibitors of DNA gyrase by 3D structure-biased ‘needle’ screening was described by Boehm *et al.* [75]. DNA gyrase is an antibacterial target that is an essential prokaryotic type II topoisomerase with no mammalian equivalent; the enzyme catalyses the ATP-dependent introduction of negative

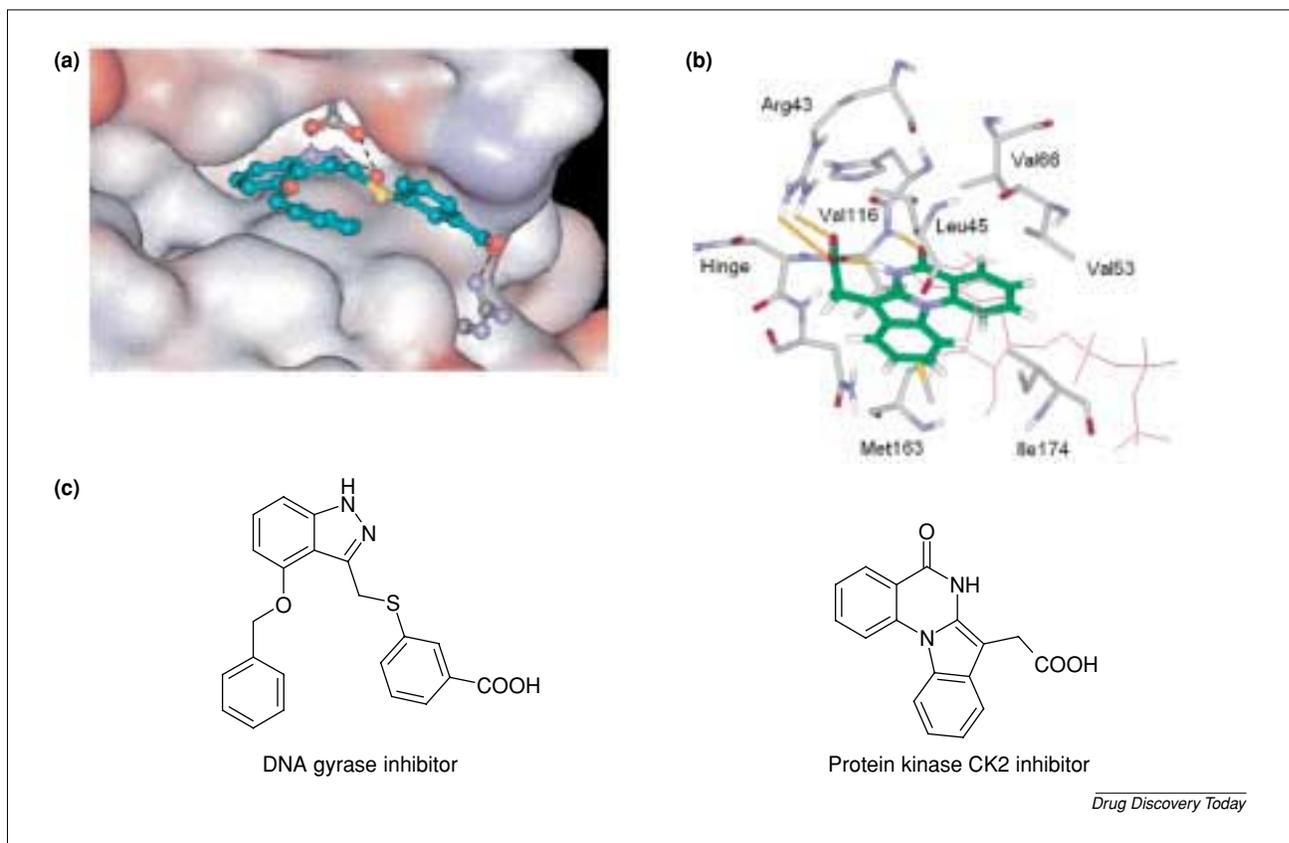


FIGURE 2

The use of virtual screening to identify potent ligands. (a) X-ray structure of a potent indazole-based inhibitor of DNA gyrase in the active site of the enzyme. Screening of ‘needles’ identified the indazole as a potential binder, and structure-guided chemistry ultimately afforded potent inhibitors. The interactions shown are the indazole forming H-bonds with Asp73 and Wat45 and the carboxylic acid forming a salt bridge to Arg136. Reproduced, with permission, from [75]. (b) High-throughput docking against a homology model of protein kinase CK2 followed by focused screening identified a potent and selective kinase inhibitor. The ligand is depicted in a binding pose predicted by DOCK in the ATP-binding site of the homology model of human CK2a. Also shown is the binding mode of ATP (magenta). Reproduced, with permission, from [75]. (c) 2D-representation of the ligands.

supercoils into bacterial DNA, with the B subunit of the A_2B_2 complex being the ATPase. The screening approach depended on an analysis of the key interactions of known ligands using protein–ligand structures. This identified that H-bonds with the side chain of Asp73 and a conserved water (Wat45) were important. Using this information, a pharmacophore was developed for *in silico* screening to identify 600 small molecules or ‘needles’ (now more conventionally referred to as fragments). These compounds were screened at a high concentration in an ATPase assay (up to 0.5 mM), which led to the identification of 150 actives that were clustered into 14 classes. To validate the hits, several additional assays were used, including NMR screening and, finally, X-ray analysis of the promising compounds. Seven of the 14 hit templates were then progressed to chemistry optimization, using information on the protein–ligand complex to direct the synthesis. The strategy ultimately led to the identification of an indazole derivative with potency tenfold higher than the DNA gyrase inhibitor novobiocin.

Vangrevelinghe *et al.* [76] reported the identification of a potent lead molecule by virtual screening of the Novartis corporate collection for the protein kinase target casein

kinase II (CK2), for which few ligands had previously been reported. A homology model of the human CK2 protein was produced using published data on the protein from *Zea mays*, which has 72% sequence similarity with human CK2. A large subset of the Novartis compound file was screened (400,000 compounds) using DOCK [77]. The virtual hits were filtered further by application of a consensus scoring filter and visual inspection of the remaining compounds to remove compounds for which there was no predicted H-bond to the hinge region of the kinase. Only 12 compounds were progressed to further testing, of which four were hits at a concentration of 10 μ M. The most potent hit was an indoloquinazolinone analogue with a potency of 80 nM. The compound was additionally found to be selective against a panel of 20 other kinases.

It is interesting to consider current and future trends in virtual screening approaches. Clearly, there is still much to be done in the area of improved ranking and scoring functions. It remains a source of frustration that a skilled modeller can look at high-ranking solutions from a docking program and dismiss many of them as ‘simply wrong’. Many researchers are turning towards scoring functions

that are targeted to a particular protein or protein class as a practical method of improving performance. However, although this approach is certainly useful, it is not entirely satisfactory [78–81]. Receptor flexibility remains an active area of research but current common practice is to employ functions that are reasonably tolerant of mild clashes and to use multiple receptor conformations [82–84]. The modelling of water molecules that form hydrogen bonds with the protein is also being explored with automated methods that will enable ligands to either displace a water molecule or form a hydrogen bond with it, depending on which case is the most energetically favourable [85]. Given the success that the pharmaceutical industry is currently experiencing with structure-based drug design and virtual screening, it should also be expected that many more examples of virtual screening and docking applications will appear in the literature.

De novo design and structure-guided library chemistry

Existing compound collections might not always contain molecules with an optimal fit for a given target protein or the compounds themselves could be of limited novelty. *In silico* methods that can design new ligands are, therefore, potentially useful. *De novo* design methods position fragments in the binding cleft of protein targets and then ‘grow’ them to fill the space available, optimizing the electrostatic, van der Waals and hydrogen bonding interactions [86–90]. Stahl and co-workers [91] have recently published a careful validation of the program Skelgen, developed by De Novo Pharmaceuticals, and this work gives a good indication of the state-of-the-art performance that can be expected from *de novo* design. As with virtual screening, there are serious challenges in the correct positioning and scoring of designed molecules but, additionally, *de novo* design can often produce molecules that are difficult to synthesize. For these reasons, there has been a tendency to initiate *de novo* design from a moiety (or fragment) that is already known to bind well to the protein target. This means that more control can be exercised over the ‘synthesisability’ of the designed compounds and that the designed compounds are more likely to bind to the receptor because they are based on an active scaffold.

A related concept is the use of structural knowledge to guide the design of combinatorial libraries. The use of structural knowledge parallels the noticeable shift in combinatorial chemistry over the past few years, away from large ‘diversity’ library synthesis to much more targeted approaches [63]. This is, in part, the result of disappointing results in identifying hits from large libraries using HTS approaches, and also because of a bias towards particular target classes that are now seen as ‘tractable’ to small molecule inhibitors. The move towards more targeted approaches is also supported by recent papers demonstrating superior hit rates for structure-guided approaches

when compared directly with diverse libraries [92,93]: other applications of structure-guided library chemistry have been discussed elsewhere [94,95].

An example of a targeted approach has been reported by Liebeschuetz *et al.* [96], in which the program PRO_SELECT was used to design candidate inhibitors of the blood coagulation enzyme factor Xa. Starting from an amidine anchor to bind in the S1 specificity pocket of the protease, templates were designed and library members enumerated, from which a small library was prepared. One hit from this library was progressed through two further rounds of library synthesis, culminating in the identification of two related inhibitors, each with an IC₅₀ of 16 nM. The X-ray crystal structure of one of the analogues confirmed the predicted binding mode.

Adams *et al.* [97] described the use of a homology model of Janus kinase 3 (JAK3) to guide library design, starting from a moderately active ligand. The oxindole hit was explored by the synthesis of a 700-membered library using 14 available oxindoles with selected aldehydes, based on dockings against the homology model. The library resulted in the identification of leads with low nanomolar potency in the kinase assay.

Fragment screening

A novel development in structural biology that has evolved in recent years is fragment-based screening. Typically, fragments are small organic molecules (100–250 Da) that exhibit low binding affinities (~100 μM to 10 mM) against target proteins, and, as such, would not usually be identified by HTS. However, the impressive advances in high-throughput NMR and crystallography have opened up the possibility of using structural information in lead discovery for HTS of protein–fragment complexes. Once a ‘hit’ has been identified in this way, and its exact binding mode elucidated, the fragment could provide a novel template that can be developed into a more complex, higher potency ligand. Despite the low affinity (assessed in a bio-assay) of these small molecule hits, fragments tend to exhibit high ‘ligand efficiency’; a high value for the average free energy of binding per heavy atom (i.e. excluding hydrogens) [98]. This property makes fragments attractive starting points for iterative medicinal chemistry optimization.

NMR spectroscopy was the first structural technique to be exploited for use in fragment screening using a method termed ‘SAR by NMR’ [99]. Perturbations to the NMR spectra of a protein are used to indicate that ligand binding is taking place and to give some indication of the location of the binding site. The ligands can be large molecules or lower molecular weight fragments. The experiments are typically performed using high concentrations of protein (200 μM) and ligand (1–10 mM). Similarly, the SHAPES approach, which uses compound scaffolds derived from those most commonly found in known therapeutic agents, exploits NMR to detect binding of a limited, but

diverse, library of small molecules to a potential drug target [100]. NMR detection of low (μM – mM) affinity binding is achieved using either differential line broadening or transferred nuclear Overhauser effect NMR methods. Recent reviews have emphasized the synergy between NMR screening methods and crystallography in applications to inhibitor design [101,102].

The high-throughput methods developed for crystallization of proteins and for solving multiple protein–ligand complexes have made it possible to screen hundreds or even thousands of small molecules in an attempt to identify fragments that bind to protein targets and then to define precisely their binding sites: the approach depends on soaking crystals with single molecules or mixtures (or cocktails). Because protein crystals contain extensive channels filled with solvent that make up around 50% of their volume, small molecules will usually diffuse rapidly into the crystals and interact as if they were in solution, as long as the binding site is not occluded by the crystal packing. The small molecules can then be visualized using difference Fourier techniques by collecting sets of X-ray data on each soaked crystal under identical conditions. Automatic procedures can be used to facilitate the rapid solution of the structure of protein–ligand complexes by interpreting and analysing the X-ray data without the need for manual intervention. For a cocktail, the various molecules can be fitted to the difference electron density and ranked; this enables complete automation of the system, once the initial protein crystals have been characterized and the structures solved.

There is an increasing number of case studies of the use of X-ray crystallography as a tool to identify fragment ‘hits’. An early use of X-ray crystallography for screening was described by Verlinde *et al.* [103], who exposed crystals of trypanosomal triosephosphate isomerase to mixtures of compounds in the search for inhibitors. Nienaber *et al.* [104] have outlined the CrystaLEAD™ method of X-ray-based screening of fragment libraries, where cocktails of up to 100 molecules were soaked against the serine protease urokinase. The potential for identifying useful inhibitors against a challenging target using a fragment approach is illustrated by Lesuisse *et al.* [105] in their studies with the Src SH2 domain. Structural GenomiX (SGX) has established an integrated technology platform for lead identification using high-throughput protein structure determination – FAST™ (Fragments of Active Structures Technology; www.stromix.com). The method enables the identification of potent, and selective, low molecular weight inhibitors of important drug targets. Hirth and Milburn [106] have described a process for the design of useful templates called Scaffold-Based Drug Discovery™, which uses X-ray analysis of protein–ligand co-crystals.

A fragment-based approach called Pyramid™ relies on the use of cocktails of small numbers of fragments at high concentrations (even as high as 200 mM) and automated molecular fragment matching and fitting using AutoSolve®

to rank candidate fragments in a cocktail [107,108]. Fragment hits against a range of enzyme targets identified using this approach have recently been described by Hartshorn *et al.* [109], and the results indicate the generality and practicality of the method. In a follow-up paper, Gill *et al.* [110] have shown how fragment hits can be evolved into potent drug-like inhibitors of p38 α mitogen-activated protein kinase. Compound **1** (Figure 3) was identified from a fragment screen of 327 ‘drug fragments’ using high-throughput crystallographic screening: on consideration of the binding interactions of **1** in the active site of the kinase, this compound was considered an attractive start-point for chemical optimization, despite its low potency (IC_{50} of 1.3 mM). After several rounds of structure-based drug design, **9** was discovered and was identified as a potent and selective inhibitor (IC_{50} of 65 nM): **9** binds by an induced fit process in which the DFG (activation) loop of the kinase underwent a significant movement to accommodate the ligand. The DFG motif marks the start of the activation loop, which is a flexible region involved in regulation of substrate binding. The area of fragment-based lead discovery has recently been thoroughly reviewed, and many more examples of the approach are described in these articles [111,112].

Kinase drug discovery

The field of kinase drug discovery is outlined here as an example of the use of structural biology during the lead optimization phase of medicinal chemistry projects. Arguably, the structural biology of kinases has had more impact on drug discovery than for many other target classes and the importance of kinase structural biology is likely to persist in the foreseeable future. The human protein kinase family consists of >500 sequences in the genome. These enzymes catalyse the transfer of the terminal phosphate group from ATP to a specific serine, threonine or tyrosine in the protein substrate. This phosphorylation of proteins is part of the intracellular signalling cascade triggered by specific extracellular signals or stresses on the cell. Kinase signalling is therefore involved in many cellular processes, such as gene expression, metabolic pathways, apoptosis and cell growth and differentiation. The protein kinase family has two major subfamilies, namely the tyrosine kinases and the serine–threonine kinases, and both have received considerable attention as drug discovery targets for a whole host of diseases, including cancer, diabetes and inflammation [113,114].

Indeed, this is only a relatively recent shift in the focus of the drug discovery community. Targeting the ATP-binding site of protein kinases was once considered extremely challenging, because the high degree of structural conservation in this region of the active site was thought to preclude the development of sufficiently selective agents for therapeutic use. However, within the past few years there has been a large number of low molecular weight, potent, ATP-site binders reported, many of which show

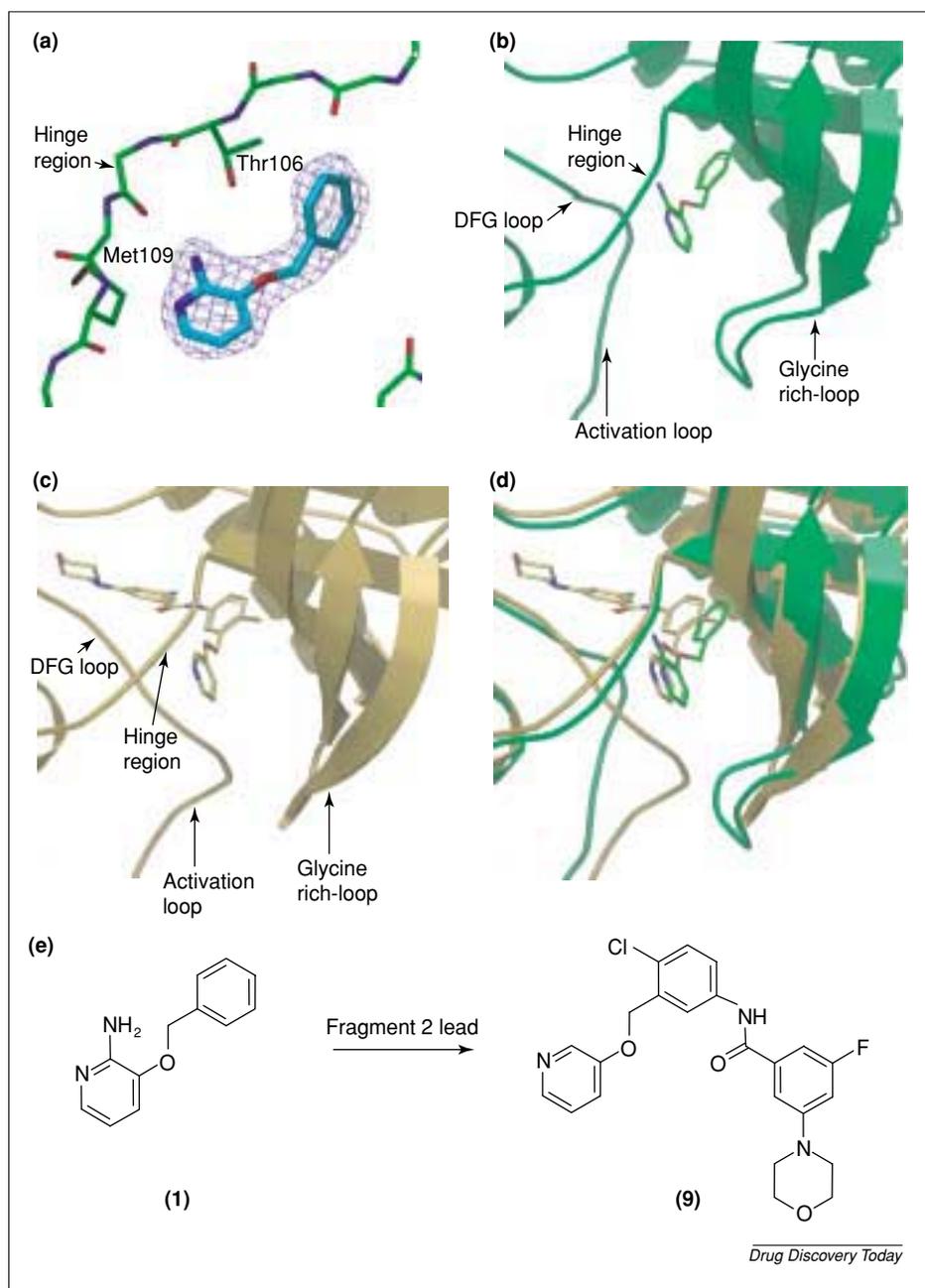


FIGURE 3

Fragment-based screening as a route to the identification of potent drug-like ligands. (a) Compound 1, discovered by fragment-based X-ray screening, is bound to the ATP-binding site of p38 α mitogen-activated protein kinase. The aminopyridine group forms a key H-bonding interaction with the backbone amide of Met109 in the hinge region and fills a hydrophobic pocket adjacent to Thr106 with the benzyl moiety. Compound 1 is a low affinity binder (IC_{50} of 1.3 mM), but each part of the molecule is interacting with the protein. Reproduced, with permission, from [110]. (b) Compound 1 shown in a ribbon version of the protein to illustrate the position of the activation loop and the glycine-rich loop. (c) Compound 9 was produced after several iterations of structure-based design, using the protein–ligand structure of 1 and structures of several of its substituted derivatives with higher potency. A large conformational change occurs on binding of 9, in which the so-called DFG loop of the kinase moves to accommodate binding of the hydrophobic amide substituent. During this movement, Phe169 moves ~ 10 Å and partially occludes the ATP-binding site. Compound 9 is considerably more potent than 1, with an IC_{50} of 65 nM. (d) Overlay of structures from (b) and (c) to illustrate the protein movement on binding of 9. (e) 2D-representations of 1 and 9. (b), (c) and (d) were generated using Astex Viewer™ (Astex Technology).

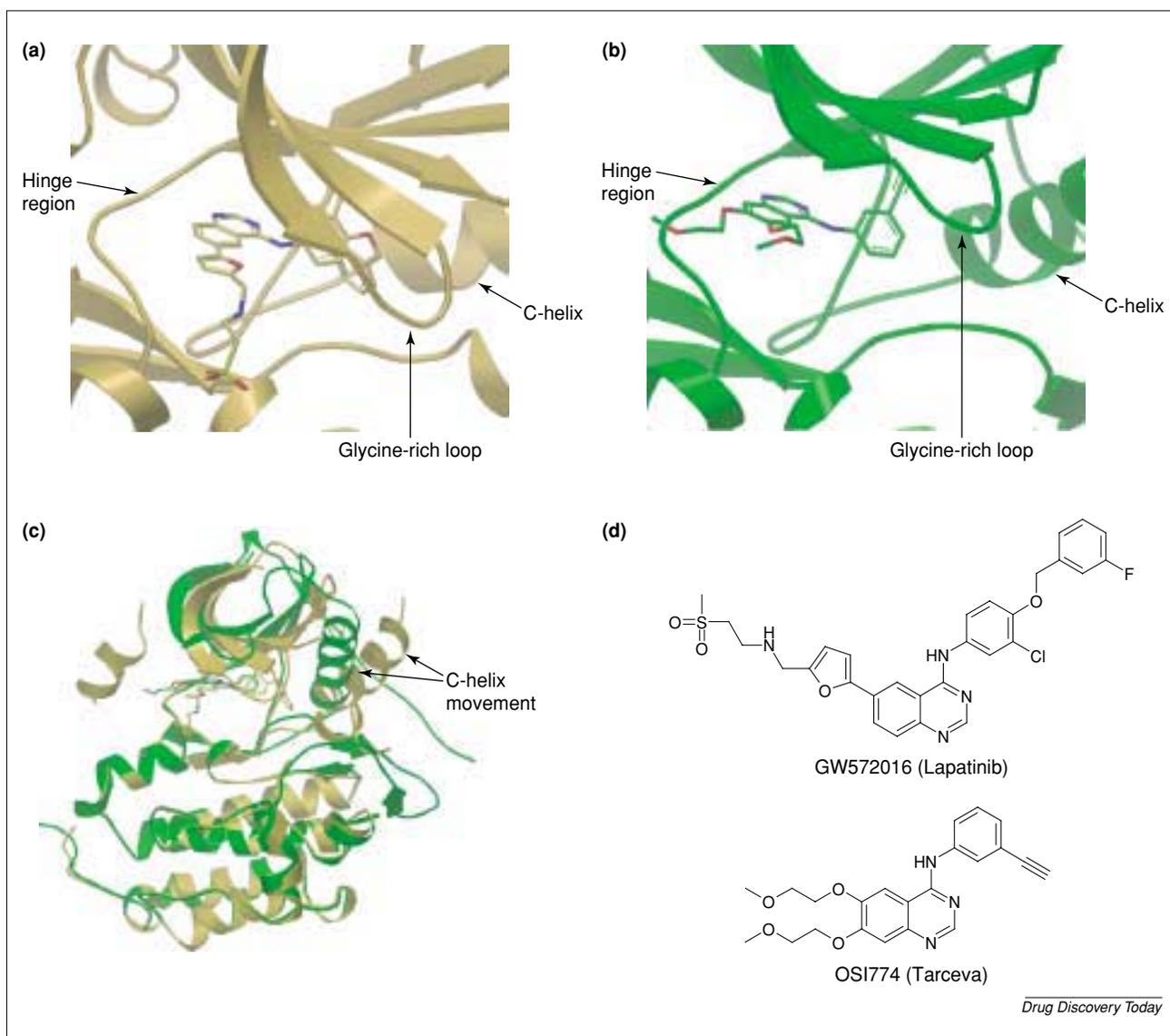
selective drugs can be developed, culminating with the first kinase inhibitor anticancer agent, Gleevec, reaching the market place [13,116].

Comparison of the binding modes of two clinically studied anticancer kinase inhibitors, OSI774 (Tarceva) and GW572016 (Lapatinib), in the ATP-binding site of the tyrosine kinase epidermal growth factor receptor (EGFR) [117] reveals that, although the two molecules are from the same chemical class (quinazolines), the protein–ligand structures have significant differences (Figure 4). In addition to variations in the key H-bonding interactions formed with the quinazoline ring and the shape of the ATP-binding site, Lapatinib binds to a relatively closed form of the ATP site, whereas Tarceva binds to a more open form. Additionally, a large pocket is created by a 9 Å shift in one end of the C-helix that accommodates the 3-fluorobenzyloxy group of Lapatinib. The shift in position of the C-helix is significant because it results in the loss of a highly conserved Glu–Lys salt bridge (Glu738 and Lys721), which is known to ligate the phosphate groups of ATP and is important in the mechanism of kinases. Lapatinib has, therefore, been described as binding to the ATP site in a conformation that resembles an inactive kinase structure, whereas Tarceva binds to a form that resembles structures of active kinases [117]. These differences have the consequence that Lapatinib has a slow off-rate from the active site, possibly as a result of the requirement for a protein conformational change on its dissociation, and it can be speculated that this effect on the enzyme kinetics influences the duration of drug activity *in vivo*.

Lapatinib also has activity against a related growth factor, ErbB-2, that could be an important enzyme in the pathophysiology of breast cancers. However, Tarceva shows a different kinase selectivity profile and lacks this activity. This difference in cross reactivity profile is likely to be because of, at least in part, the conformation of the kinase that the ligands have a tendency to bind to – inactive versus active. These findings suggest that small

good selectivity against a broad range of kinases [115]. A structural understanding of how these inhibitors bind has built confidence in the field that efficacious and relatively

changes in the chemical structures of kinase inhibitors can have profound consequences on the binding mode and kinetics of binding with target kinases.

**FIGURE 4**

Kinase drug discovery. (a) and (b) illustrate the protein–ligand binding modes of Lapatinib and Tarceva, respectively, in the EGFR tyrosine kinase ATP site, using similar ligand orientations; generated using Astex Viewer™ (Astex Technology) from Protein Data Bank structures 1XKK and 1M17 for Lapatinib and Tarceva, respectively. (c) The tertiary structure of the kinase is different for each ligand: Lapatinib binds to an inactive-like conformation, whereas Tarceva binds to an active-like conformation. The major differences close to the ligands are in the movement of the glycine-rich loop and shift of the C-helix. (d) 2D-representations of Lapatinib and Tarceva.

This is just one example of how an understanding of the structural biology of kinases is having a considerable impact on this flourishing field of research and is driving a detailed understanding of the properties of kinase inhibitors. Because of the close similarity of the ATP-binding site of kinases and the large number of kinase enzymes in the genome, it is unlikely that highly selective compounds for a kinase target of choice can ever be developed. However, ‘combinatorial inhibition’ or ‘spectrum selective’ inhibitors of multiple protein kinases potentially have an attractive profile for the treatment of cancer, because of the broad range of factors that cause cancer in any given patient population. The key challenges in kinase research are either developing relatively selective inhibitors or, instead,

inhibitors with several complementary activities. Structural biology is likely to remain a vital tool to achieve these ends.

Conclusions

Knowledge of the 3D structures of protein targets is now playing a major role in all stages of drug discovery. Its place in lead optimization is well established, with large teams of structural biologists recruited into all major pharmaceutical companies. The success of the method is evident from the drugs currently in use and new ones reaching the market, and it is clear that, in many companies, structure-guided approaches have become central to developing good drug candidates. Despite these trends,

structural biology remains a challenging field and breakthroughs are still required to elucidate the structures of many key proteins and protein families.

Structural biology has also played an important role in the exploitation of recent developments in the sequencing of the human genome and the genomes of model organisms and disease agents through target identification and validation. Target selection is heavily influenced by structural information as an indicator of the tractability of a given protein family to inhibition by small molecule

ligands. Additionally, a rational understanding of the selectivity of an agent against related proteins often depends on information on the protein–ligand complex. But the most exciting advances in recent years have come from the use of high-throughput X-ray analysis and NMR in structural screening, principally for fragments. It has now been demonstrated that these can be used to ‘grow’ useful lead molecules for most classes of drug target. Early results indicate that this approach can significantly accelerate the early stages of drug discovery.

References

- 1 Beddell, C.R. *et al.* (1976) Compounds designed to fit a site of known structure in human haemoglobin. *Br. J. Pharmacol.* 57, 201–209
- 2 Goodford, P.J. *et al.* (1980) The interaction of human haemoglobin with allosteric effectors as a model for drug–receptor interactions. *Br. J. Pharmacol.* 68, 741–748
- 3 Blundell, T.L. *et al.* (1972) The structure, chemistry and biological activity of insulin. *Adv. Protein Chem.* 26, 279–402
- 4 Davie, E.W. *et al.* (1991) The coagulation cascade: initiation, maintenance and regulation. *Biochemistry* 30, 10363–10370
- 5 Blundell, T.L. (1996) Structure-based drug design. *Nature* 384, 23–26
- 6 Sibanda, B.L. *et al.* (1983) Three-dimensional structure, specificity and catalytic mechanism of renin. *Nature* 304, 273–275
- 7 Campbell, S.F. (2000) Science, art and drug discovery: a personal perspective. *Clin. Sci.* 99, 255–260
- 8 Miller, M. *et al.* (1989) Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. *Science* 246, 1149–1152
- 9 Lapatto, R. *et al.* (1989) X-ray analysis of HIV-1 proteinase at 2.7 Å resolution confirms structural homology among retroviral enzymes. *Nature* 342, 299–302
- 10 Varghese, J.N. (1999) Development of neuraminidase inhibitors as anti-influenza virus drugs. *Drug Dev. Res.* 46, 176–196
- 11 Hardy, L.W. and Malikayil, A. (2003) The impact of structure-guided drug design on clinical agents. *Curr. Drug Discov.* December, 15–20
- 12 Lombardino, J.G. and Lowe, J.A. (2004) A guide to drug discovery: the role of the medicinal chemist in drug discovery – then and now. *Nat. Rev. Drug Discov.* 3, 853–862
- 13 Wong, S. and Witte, O.N. (2004) The BCR-ABL story: bench to bedside and back. *Annu. Rev. Immunol.* 22, 247–306
- 14 Blundell, T. *et al.* (1988) 18th Sir Hans Krebs Lecture. Knowledge-based protein modelling and design. *Eur. J. Biochem.* 172, 513–520
- 15 Pearl, L.H. and Taylor, W.R. (1987) A structural model for the retroviral proteases. *Nature* 329, 351–354
- 16 Bowie, J.U. *et al.* (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253, 164–170
- 17 Rice, D.W. and Eisenberg, D. (1997) A 3D–1D substitution matrix for protein fold recognition that includes predicted secondary structure of the sequence. *J. Mol. Biol.* 267, 1026–1038
- 18 Overington, J. *et al.* (1990) Tertiary structural constraints on protein evolutionary diversity: templates, key residues and structure prediction. *Proc. Biol. Sci.* 241, 132–145
- 19 Johnson, M.S. *et al.* (1993) Alignment and searching for common protein folds using a data bank of structural templates. *J. Mol. Biol.* 231, 735–752
- 20 Shi, J. *et al.* (2001) Fugue: sequence–structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.* 310, 243–257
- 21 Jones, D.T. *et al.* (1992) A new approach to protein fold recognition. *Nature* 358, 86–89
- 22 Lindahl, E. and Elofsson, A. (2000) Identification of related proteins on family, superfamily and fold level. *J. Mol. Biol.* 295, 613–625
- 23 Jones, D.T. (1999) GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J. Mol. Biol.* 287, 797–815
- 24 Sutcliffe, M.J. *et al.* (1987) Knowledge-based modelling of homologous proteins, part I: three-dimensional frameworks derived from the simultaneous superposition of multiple structures. *Protein Eng.* 1, 377–384
- 25 Guex, N. (1999) Protein modelling for all. *Trends Biochem. Sci.* 24, 364–367
- 26 Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815
- 27 Venclovas, C. *et al.* (2003) Assessment of progress over the CASP experiments. *Proteins* 53 (Suppl. 6), 585–595
- 28 Diller, D.J. and Li, R. (2003) Kinases, homology models, and high-throughput docking. *J. Med. Chem.* 46, 4638–4647
- 29 Chuaqui, C. *et al.* (2005) Interaction profiles of protein kinase–inhibitor complexes and their application to virtual screening. *J. Med. Chem.* 48, 121–133
- 30 Heinemann, U. *et al.* (2003) Facilities and methods for the high-throughput crystal structural analysis of human proteins. *Acc. Chem. Res.* 36, 157–163
- 31 Service, R.F. (2002) Structural genomics. Tapping DNA for structures produces a trickle. *Science* 298, 948–950
- 32 Rupp, B. (2003) High-throughput crystallography at an affordable cost: the TB Structural Genomics consortium crystallization facility. *Acc. Chem. Res.* 36, 173–181
- 33 Lesley, S.A. *et al.* (2002) Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc. Natl. Acad. Sci. U. S. A.* 99, 11664–11669
- 34 Chambers, S.P. (2002) High-throughput protein expression for the post-genomic era. *Drug Discov. Today* 7, 759–765
- 35 Endo, Y. and Sawasaki, T. (2003) High-throughput, genome-scale protein production method based on the wheat germ cell-free expression system. *Biotechnol. Adv.* 21, 695–713
- 36 Lanio, T. *et al.* (2003) High-throughput purification of polyHis-tagged recombinant fusion proteins. *Methods Mol. Biol.* 205, 199–203
- 37 Scheich, C. *et al.* (2003) An automated method for high-throughput protein purification applied to a comparison of His-tag and GST-tag affinity chromatography. *BMC Biotechnol.* 3, 12–17
- 38 Hui, R. and Edwards, A. (2003) High-throughput protein crystallization. *J. Struct. Biol.* 142, 154–161
- 39 Chayen, N.E. and Saridakis, E. (2002) Protein crystallization for genomics: towards high-throughput optimization techniques. *Acta Crystallogr. D Biol. Crystallogr.* 58, 921–927
- 40 Stewart, L. *et al.* (2002) High-throughput crystallization and structure determination in drug discovery. *Drug Discov. Today* 7, 187–196
- 41 Bodenstaff, E.R. *et al.* (2002) The prospects of protein nanocrystallography. *Acta Crystallogr. D Biol. Crystallogr.* 58, 1901–1906
- 42 Muchmore, S.W. (2000) Automated crystal mounting and data collection for protein crystallography. *Structure Fold. Des.* 8, R243–R246
- 43 Sharff, A. (2003) High-throughput crystallography on an in-house source using ACTOR. *The Rigaku Journal* 19, 5–10
- 44 Sheldrick, G.M. (1997) Patterson superposition and *ab initio* phasing. *Methods Enzymol.* 276, 628–641
- 45 Weeks, C.M. and Miller, R. (1999) Optimizing shake and bake for proteins. *Acta Crystallogr. D Biol. Crystallogr.* 55, 492–500
- 46 Lamzin, V.S. and Perrakis, A. (2000) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* 7, 978–981
- 47 DePristo, M.A. *et al.* (2004) Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. *Structure* 12, 831–838
- 48 Williams, P.A. *et al.* (2000) Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol. Cell* 5, 121–131
- 49 Williams, P.A. *et al.* (2004) Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science* 305, 683–686
- 50 Yano, J.K. *et al.* (2004) The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-Å resolution. *J. Biol. Chem.* 279, 38091–38094
- 51 Hofmann, K. *et al.* (1999) The PROSITE database, its status in 1999. *Nucleic Acids Res.* 27, 215–219
- 52 Kasuya, A. and Thornton, J.M. (1999) Three-dimensional structure analysis of PROSITE patterns. *J. Mol. Biol.* 286, 1673–1691
- 53 Herzberg, O. and Moulton, J. (1991) Analysis of

- the steric strain in the polypeptide backbone of protein molecules. *Proteins* 11, 223–229
- 54 Heringa, J. and Argos, P. (1999) Strain in protein structures as viewed through nonrotameric side chains: II. Effects upon ligand binding. *Proteins* 37, 44–55
- 55 Laskowski, R. *et al.* (1996) Protein clefts in molecular recognition and function. *Protein Sci.* 5, 2438–2452
- 56 Zvelebil, M.J. *et al.* (1987) Prediction of protein secondary structure and active sites using the alignment of homologous sequences. *J. Mol. Biol.* 195, 957–961
- 57 McPhalen, C.A. *et al.* (1992) Domain closure in mitochondrial aspartate aminotransferase. *J. Mol. Biol.* 227, 197–223
- 58 Irving, J.A. *et al.* (2001) Protein structural alignments and functional genomics. *Proteins* 42, 378–382
- 59 Lichtarge, O. and Sowa, M.A. (2002) Evolutionary predictions of binding surfaces and interactions. *Curr. Opin. Struct. Biol.* 12, 21–27
- 60 Chelliah, V. *et al.* (2004) Distinguishing structural and functional restraints in evolution in order to identify interaction sites. *J. Mol. Biol.* 342, 1487–1504
- 61 Seneci, P. and Miertus, S. (2000) Combinatorial chemistry and high-throughput screening in drug discovery: different strategies and formats. *Mol. Divers.* 5, 75–89
- 62 Bailey, N. *et al.* (1997) Solution-phase combinatorial chemistry in lead discovery. *Chimia* 51, 832–837
- 63 Dolle, R.E. (2004) Comprehensive survey of combinatorial library synthesis: 2003. *J. Comb. Chem.* 6, 623–679
- 64 Spencer, R.W. (1998) High-throughput screening of historic collections: observations on file size, biological targets, and file diversity. *Bioeng.* 61, 61–67
- 65 Hopkins, A.L. and Groom, C.R. (2002) Opinion: the druggable genome. *Nat. Rev. Drug Discov.* 1, 727–730
- 66 Taylor, R.D. *et al.* (2002) A review of protein-small molecule docking methods. *J. Comput. Aided Mol. Des.* 16, 151–166
- 67 Nissink, J.W.M. *et al.* (2002) A new test set for validating predictions of protein-ligand interaction. *Proteins* 49, 457–471
- 68 Friesner, R.A. *et al.* (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *J. Med. Chem.* 47, 1739–1749
- 69 Kitchen, D.B. *et al.* (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* 3, 935–949
- 70 Shoichet, B.K. *et al.* (2002) Lead discovery using molecular docking. *Curr. Opin. Chem. Biol.* 6, 439–446
- 71 Bajorath, J. (2002) Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discov.* 1, 882–894
- 72 Lyne, P.D. (2002) Structure-based virtual screening: an overview. *Drug Discov. Today* 7, 1047–1052
- 73 Jain, A.N. (2004) Virtual screening in lead discovery and optimization. *Curr. Opin. Drug Discov. Devel.* 7, 396–403
- 74 Abagyan, R. and Totrov, M. (2001) High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* 5, 375–382
- 75 Boehm, H.-J. *et al.* (2000) 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
- 76 Vangrevelinghe, E. *et al.* (2003) Discovery of a potent and selective protein kinase CK2 inhibitor by high-throughput docking. *J. Med. Chem.* 46, 2656–2662
- 77 Ewing, T.J. *et al.* (2001) Dock 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des.* 15, 411–428
- 78 Knegtel, R.M.A. *et al.* (1997) Molecular docking to ensembles of protein structures. *J. Mol. Biol.* 266, 424–440
- 79 Gohlke, H. and Klebe, G. (2002) DrugScore meets CoMFA: adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein. *J. Med. Chem.* 45, 4153–4170
- 80 Wu, G. and Vieth, M. (2004) SDOCKER: a method utilizing existing X-ray structures to improve docking accuracy. *J. Med. Chem.* 47, 3142–3148
- 81 Hindle, S.A. *et al.* (2002) Flexible docking under pharmacophore type constraints. *J. Comput. Aided Mol. Des.* 16, 129–149
- 82 Ferrari, A.M. *et al.* (2004) Soft docking and multiple receptor conformations in virtual screening. *J. Med. Chem.* 47, 5076–5084
- 83 Kovacs, J.A. *et al.* (2004) Predictions of protein flexibility: first-order measures. *Proteins* 56, 661–668
- 84 Claussen, H. *et al.* (2001) FlexE: efficient molecular docking considering protein structure variations. *J. Mol. Biol.* 308, 377–395
- 85 Rarey, M. *et al.* (1999) The particle concept: placing discrete water molecules during protein-ligand docking predictions. *Proteins* 34, 17–28
- 86 Cohen, N.C. and Tschinke, N. (1995) Generation of new-lead structures in computer-aided drug design. *Prog. Drug Res.* 45, 205–243
- 87 Murcko, M.A. (1997) Recent advances in ligand design methods. *Rev. Comput. Chem.* 11, 1–66
- 88 Clark, D.E. *et al.* (1997) Current issues in *de novo* molecular design. *Rev. Comput. Chem.* 11, 67–125
- 89 Honma, T. (2003) Recent advances in *de novo* design strategy for practical lead identification. *Med. Res. Rev.* 23, 606–632
- 90 Dean, P.M. *et al.* (2004) *De novo* drug design: integration of structure-based and ligand-based methods. *Curr. Opin. Drug Discov. Devel.* 7, 347–353
- 91 Stahl, M. *et al.* (2002) A validation study on the practical use of automated *de novo* design. *J. Comput. Aided Mol. Des.* 16, 459–478
- 92 Wyss, P.C. *et al.* (2003) Novel dihydrofolate reductase inhibitors. Structure-based versus diversity-based library design and high-throughput synthesis and screening. *J. Med. Chem.* 46, 2304–2312
- 93 Doman, T.N. *et al.* (2002) Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1b. *J. Med. Chem.* 45, 2213–2221
- 94 Bohm, H.-J. and Stahl, M. (2000) Structure-based library design: molecular modeling merges with combinatorial chemistry. *Curr. Opin. Chem. Biol.* 4, 283–286
- 95 Laird, E.R. (2004) Structure-based generation of viable leads from small combinatorial libraries. *Curr. Opin. Drug Discov. Devel.* 7, 354–359
- 96 Liebeschuetz, J.W. *et al.* (2002) PROSELECT: combining structure-based drug design and array-based chemistry for rapid lead discovery. 2. The development of a series of highly potent and selective factor Xa inhibitors. *J. Med. Chem.* 45, 1221–1232
- 97 Adams, C. *et al.* (2003) Mapping the kinase domain of janus kinase 3. *Bioorg. Med. Chem. Lett.* 13, 3105–3110
- 98 Hopkins, A.L. *et al.* (2004) Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* 9, 430–431
- 99 Shuker, S.B. *et al.* (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274, 1531–1534
- 100 Fejzo, J. *et al.* (1999) The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chem. Biol.* 6, 755–769
- 101 Muchmore, S.W. and Hajduk, P.J. (2003) Crystallography, NMR and virtual screening: integrated tools for drug discovery. *Curr. Opin. Drug Discov. Devel.* 6, 544–549
- 102 Moore, J. *et al.* (2004) Leveraging structural approaches: applications of NMR screening and X-ray crystallography for inhibitor design. *J. Synchrotron Radiat.* 11, 97–110
- 103 Verlinde, C.L.M.J. *et al.* (1997) Antitrypanosomiasis drug development based on structures of glycolytic enzymes. In *Structure-Based Drug Design* (Veerapandian, P., ed.), pp. 365–394, Marcel Dekker
- 104 Nienaber, V.L. *et al.* (2000) Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nat. Biotechnol.* 18, 1105–1108
- 105 Lesuisse, D. *et al.* (2002) SAR and X-ray. A new approach combining fragment-based screening and rational drug design: Application to the discovery of nanomolar inhibitors of Src SH2. *J. Med. Chem.* 45, 2379–2387
- 106 Hirth, K.P. and Milburn, M.V. (2004) Methods for the design of molecular scaffolds and ligands. *US Patent Application Publication No. US2004/0171062*
- 107 Blundell, T.L. *et al.* (2002) High-Throughput X-ray Crystallography for Drug Discovery. In *Drug Design, Cutting Edge Approaches* (Flower, D., ed.), pp. 53–59, Royal Society Chemistry
- 108 Carr, R. and Jhotti, H. (2003) Structure-based screening of low affinity compounds. *Drug Discov. Today* 7, 522–527
- 109 Hartshorn, M.J. *et al.* (2005) Fragment-based lead discovery using X-ray crystallography. *J. Med. Chem.* 48, 403–413
- 110 Gill, A.L. *et al.* (2005) Identification of novel p38a MAP kinase inhibitors using fragment-based lead generation. *J. Med. Chem.* 48, 414–426
- 111 Rees, D.C. *et al.* (2004) Fragment-based lead discovery. *Nat. Rev. Drug Discov.* 3, 660–672
- 112 Erlanson, D.A. *et al.* (2004) Fragment-based drug discovery. *J. Med. Chem.* 47, 3463–3482
- 113 Parang, K. and Sun, G. (2004) Design strategies for protein kinase inhibitors. *Curr. Opin. Drug Discov. Devel.* 7, 617–629
- 114 Gill, A. (2004) New lead generation strategies for protein kinase inhibitors – fragment-based screening approaches. *Mini Rev. Med. Chem.* 4, 301–311
- 115 Fabian, M.A. *et al.* (2005) A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat. Biotechnol.* 23, 329–336
- 116 Mol, C.D. *et al.* (2004) Structural insights into the conformational selectivity of STI-571 and related kinase inhibitors. *Curr. Opin. Drug Discov. Devel.* 7, 639–648
- 117 Wood, E.R. *et al.* (2004) A unique structure for the epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among

- protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Res.* 64, 6652–6659
- 118 Cushman, D.W. *et al.* (1977) Design of potent inhibitors of ACE. *Biochemistry* 16, 5484–5491
- 119 Baldwin, J.J. *et al.* (1989) Thienothiopyran-2-sulfonamides: novel topically active carbonic anhydrase inhibitors for the treatment of glaucoma. *J. Med. Chem.* 32, 2510–2513
- 120 Kaldor, S.W. *et al.* (1997) Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *J. Med. Chem.* 40, 3979–3985
- 121 Kim, E.E. *et al.* (1995) Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. *J. Am. Chem. Soc.* 117, 1181–1182
- 122 Sham, H.L. *et al.* (1998) ABT-378, a highly potent inhibitor of the human immunodeficiency virus protease. *Antimicrob. Agents Chemother.* 42, 3218–3224
- 123 von Itzstein, M. *et al.* (1993) Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* 363, 418–423
- 124 Kim, C.U. *et al.* (1998) Structure–activity relationship studies of novel carbocyclic influenza neuraminidase inhibitors. *J. Med. Chem.* 41, 2451–2460
- 125 Zimmermann, J. *et al.* (1996) (Phenylamino)pyrimidine (PAP) derivatives: a new class of potent and highly selective PDGF-receptor autophosphorylation inhibitors. *Bioorg. Med. Chem. Lett.* 6, 1221–1226
- 126 Zimmermann, J. *et al.* (1997) Potent and selective inhibitors of the ABL-kinase: phenylaminopyrimidine (PAP) derivatives. *Bioorg. Med. Chem. Lett.* 7, 187–192
- 127 Pollack, V.A. *et al.* (1999) Inhibition of epidermal growth factor receptor-associated tyrosine phosphorylation in human carcinomas with CP-358774: dynamics of receptor inhibition *in situ* and antitumor effects in athymic mice. *J. Pharmacol. Exp. Ther.* 291, 739–748
- 128 Gustafsson, D. *et al.* (2001) The direct thrombin inhibitor melagatran and its oral prodrug H 376/95: intestinal absorption properties, biochemical and pharmacodynamic effects. *Thromb. Res.* 101, 171–181

The Feature article in this issue of *Drug Discovery Today* highlights the potential of microdosing in the drug discovery process. Ian Wilding from Pharmaceutical Profiles (www.pharmprofiles.com), one of the authors of the article, has just been awarded the 2005 'Career Achievement in Oral Drug Delivery Award' sponsored by the Controlled Release Society (CRS) and Eurand. At the forthcoming CRS meeting in Miami, USA (18–22 June 2005), Dr Wilding will present an update on using human absorption and scintigraphic studies to overcome problems with developing molecules for oral drug delivery.