DDT Vol. 7, No. 9 May 2002

Structure-based screening of low-affinity compounds

Robin Carr and Harren Jhoti

Conventional bioassay-based screening remains a mainstream approach for lead discovery. However, its limitations have meant that other, more biophysical methods, such as X-ray crystallography and NMR, are now being developed as lead discovery tools. These methods are particularly effective at detecting the binding of low affinity, low molecular weight compounds and transforming them into novel potent leads using structure-guided chemistry. Here, we describe some of the technologies and approaches that are being developed in structure-based screening using X-ray crystallography, which promise to have a major impact on lead discovery.

Robin Carr and Harren Jhoti*
Astex Technology
250 Cambridge Science Park
Cambridge
UK CB4 0WE
*tel: +44 1223 226 200
fax: +44 1223 226 201
e-mail: h.jhoti@astex-technology.com

▼ Today, the process of lead discovery remains a high-risk endeavour. Indeed, major pharmaceutical companies acknowledge that they are currently successful in identifying a quality lead for a druggable protein target in only 25% of projects [1]. Thus, although chemistry and screening technologies have massively increased throughput over the last decade, lead discovery productivity has not improved accordingly [2]. This inability to routinely identify multiple high quality lead compounds against drug targets is a major issue facing the drug discovery industry.

Size matters

One aspect of this problem resides in the nature of 'hits' identified from traditional bioassay-based HTS. The average MW of successful drugs in the World Drug Index is in the low 300s [3], which is similar to that in current corporate collections; that is, corporate collections have evolved to be broadly 'drug like' with respect to MW and other features. However, recent publications conclude that hits from HTS should have a lower MW than drugs, therefore, screening drug-like compounds might not be the most effective way to find leads [4]. This conclusion is based on the expected increase in molecular weight, of ~80, during the lead optimization process. Therefore, an HTS hit from a corporate screening set with μM affinity towards the target could well already have an 'average drug MW'. However, it is probable that the MW will increase significantly during the lead optimization process, leading to significantly poorer drug-like properties with respect to solubility, absorption and clearance [5].

To address this issue, several groups have been developing methods to identify low molecular weight screen hits (MW 100–250), which could be optimized efficiently into novel lead compounds that possess excellent drug-like properties [6–11]. These 'molecular fragments' would, by definition, have limited functionality and will therefore typically exhibit weaker affinity (typically in the 50 μm–1 mM range). This affinity range is outside of the normal HTS range and, as such, can not routinely be identified in standard bioassays because of the high concentration of compound that would be required, interfering with the assay and leading to significant false positives. Rather than trying to push bioassays into this affinity range, people are increasingly turning to biophysical methods, such as NMR and X-ray crystallography, for fragment-based screening approaches. For example, Fesik and colleagues have pioneered methods in which NMR is used to screen libraries of molecular fragments [7,8]. In determining SARs by NMR, 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.

X-ray crystallography

X-ray crystallography has the advantage of defining the ligand-binding sites with more

PII: S1359-6446(02)02245-6

certainty and the binding orientations of the molecular fragments have a crucial role in guiding efficient lead optimization programs [12]. More generally, crystal structures of target proteins are routinely being used in conventional structure-based drug discovery where compounds are optimized using 3D information. However, X-ray crystallography has traditionally been regarded as a resource-intensive technique, which has restricted its use primarily to the lead optimization phase where it is used to study a small number of high-value compounds. This perception is now changing as a result of major technological advances in both software and hardware, which are now enabling the determination of protein structures at an unprecedented rate. In this article we focus on the impact that these technological advances are having and illustrate how these advances are resulting in X-ray crystallography being used, for the first time in the lead discovery phase, by its application to structure-based screening of low-affinity compounds.

Advances in X-ray crystallography

There are several useful reviews on the efforts to automate the process from gene to crystal structure for novel protein-structure-determination [13–16]. Technology advances are being made in parallel cloning, optimizing expression, nanolitre crystallization, automated phasing and model building [17–19]. Several structural genomics initiatives are underway in the USA, Germany and Japan in which these new technologies are being applied to achieve the goal of obtaining crystal structures of novel gene families [20–22]. Currently, there are 15,000 3D structures in the Protein Data Bank, although this represents only ~5000 different wild-type proteins [23]. Interestingly, the pharmaceutical industry has only developed drugs for ~500 proteins to date, 30% for which there is some 3D structural information [24]. The structural genomics efforts are expected to significantly increase these numbers in the next 5–10 years, with 3D structures available for all (non-membrane bound) classes of protein. A consequence of this explosion in new protein structures for the pharmaceutical industry is that there will be a much-increased chance of knowing the 3D structure of a new therapeutic target at the beginning of the lead discovery process. Even if the structure of the actual protein target has not been determined, structures of its close homologues are increasingly likely to be known.

Protein–ligand crystallography

Once the 3D structure of a target protein has been determined it is then important to identify the active site and key binding interactions. The most reliable approach is to determine the structure of a protein–ligand complex, either by co-crystallization or by soaking the ligand into the preformed crystal. However, when X-ray crystallography is used as a method for screening, the soaking option is much preferred. After collecting the X-ray data from a protein crystal exposed to a ligand, the next step is to analyze and interpret the resulting electron density. This step is often time-consuming and requires a crystallographer to spend several days assessing the data from a single protein–ligand experiment. This is a key bottleneck for using X-ray crystallography as a method for screening compounds.

Technology advances have now been made to automate and accelerate this step. Software tools such as Quanta from Accelrys (San Diego, CA, USA) and AutoSolve® from Astex (Cambridge, UK) assist the crystallographer in the analysis and interpretation steps [6]. In fact, AutoSolve® requires no human intervention if the quality of electron density is high, and can identify the correct compound bound at the active site from an experiment where the crystal has been exposed to a cocktail of compounds.

X-ray data collection

Advances have also been made in the field of X-ray data collection from protein crystals. The use of synchrotron radiation sources, such as the European Synchrotron Radiation Facility (ESRF) at Grenoble (France) and the Advanced Photon Source (APS) at Argonne National Laboratory (Chicago, IL, USA), to generate high-intensity X-rays have enabled many novel structures to be determined at unprecedented rates [25]. The unique characteristics of X-rays from synchrotron sources have been effectively exploited for structure determination of novel proteins by using methods such as Multi-wavelength Anomalous Dispersion (MAD). However, once the structure of a protein target is known, the challenge becomes that of determining the structure of a protein–ligand complex and the advantages of synchrotron radiation are unclear. As a result, the majority of protein–ligand crystallography is performed on laboratory X-ray sources. In this setting, recent developments of charged-couple device (CCD) detectors are significantly increasing the throughput of X-ray data collection [26].

To further maximize the efficiency of X-ray data collection, sample-changing robots have now been developed that mount crystals sequentially while maintaining liquid-nitrogen temperature, automatically align the crystal in the beam, collect complete X-ray datasets and return the crystals to storage [27]. These robots offer the possibility of around-the-clock X-ray data collection without the need for manual intervention and, together with developments in software for automated data analysis, will enable X-ray crystallography to be used as a screening method.

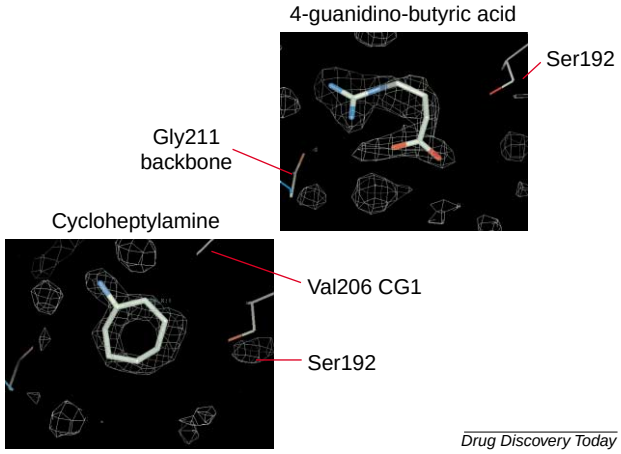

Figure 1. Molecular fragments bound into a pocket of trypsin. Electron density representing 4-guanidino-butyric acid (top) and cycloheptylamine (bottom) was detected and interpreted with models being automatically generated and fitted using AutoSolve® (Astex Technology, Cambridge, UK). The electron density maps are contoured at 3σ, and density caused by protein and solvent has been removed for clarity.

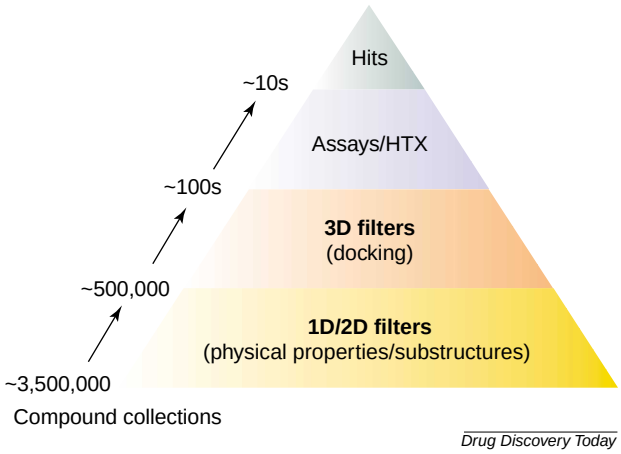

Figure 2. Process for the selection of compounds. A variety of computational filters are applied to select and generate libraries of molecular fragments for screening using X-ray crystallography. Chemical databases containing millions of compounds are sampled to produce fragment libraries containing 500–1000 compounds. Abbreviation: HTX, high-throughput crystallography.

Fragment-based screening using X-ray crystallography

Different sets of molecular fragments can be used to target a particular protein. For example, in a screen of fragments against trypsin, a 'focussed set' was selected based on known binders, such as benzamidine, 4-aminopyridine and cyclohexylamine [6]. These were then dissolved in an organic solvent such as dimethylsulphoxide (DMSO) and added to a single protein crystal, then left to soak for one hour to give the molecule time to penetrate into the active site. The concentration of the molecular fragment is typically >20 mM, reflecting the low-affinity that is expected. Fragment libraries can be screened as singlets or in cocktails using X-ray crystallography. As the output from an X-ray experiment is a visual description of the bound compound (its electron density) it is possible to screen cocktails of compounds without the need to deconvolute. An optimum cocktail size is typically between four and eight and is defined by the tolerance of the protein crystals to organic solvents and the concentration at which you wish to screen each fragment. For example, if the maximum tolerated solvent concentration is 240 mM then you can screen eight compounds each at a concentration of 30 mM.

Identifying hits

Some of the first experiments in which X-ray crystallography was used as a 'screening tool' were reported by Verlinde and colleagues who exposed crystals of trypanosomal triosephosphate isomerase to cocktails of compounds in their search for inhibitors [28]. More recently, Greer and colleagues have described a method for screening using X-ray crystallography that focusses on soaking the target crystals with cocktails of compounds with differing shapes that can be easily distinguished by visual inspection of the electron density [10]. However, to fully exploit X-ray crystallography as a screening approach it is desirable to implement an objective and automated process to address the key bottleneck of data interpretation and analysis. AutoSolve® enables rapid and automated analysis of electron density from fragment soaking experiments using singlets and cocktails of compounds. Examples of electron density that were unambiguously interpreted by AutoSolve® are shown in Fig. 1. In each case, the binding mode of the small-molecule fragment is clearly defined by the electron density, which indicates that, although the affinity could be in the millimolar range, the binding is ordered with key interactions made between the compound and the protein.

Another key advantage of using molecular fragments for screening is the significant amount of chemical space that is sampled using a relatively small library of compounds. For example, if the binding of several heterocycles is probed against specific binding pockets in a protein, the discrimination between a binding event and a non-binding event is solely dependent on the molecular complementarity and is not constrained or modulated by the heterocycle being part of a larger molecule. This is a far more comprehensive and elegant way to probe for new interactions than having the fragments attached to a rigid template, as might be

derived from a conventional combinatorial chemistry approach.

In silico *screening of molecular fragments*

Large compound collections are currently 1×10^6 in scale; however, an estimate of drug-like space is perhaps of the order 1×10^{60}. Hence, even ultra-HTS screening is sampling such a small percentage of all options that the choice of compound library is crucial. In addition, the high cost of screening large compound collections against more and more targets has focussed attention on virtual approaches to pre-filter compounds [29]. For example, 1D and 2D filters can be used to identify compounds with specific physico-chemical properties, such as specific molecular weights, undesirable functional groups and numbers of hydrogen bond donors and acceptors. These types of computational approaches can also be applied to generate libraries of molecular fragments (Fig. 2). Sub-libraries of molecular fragments can then be filtered further using either 'knowledge-based' or 'diversity-based' approaches. In a knowledge-based approach, a 3D pharmacophore defined using the shape and chemical nature of the active site derived from the protein structure can be used to select compounds that exhibit compatible features. Compounds can also be docked into the target binding-site using a protein–ligand docking program, such as GOLD (Cambridge Crystallographic Data Centre, Cambridge, UK) and ranked using scoring functions such as GOLD score or ChemScore [30–32].

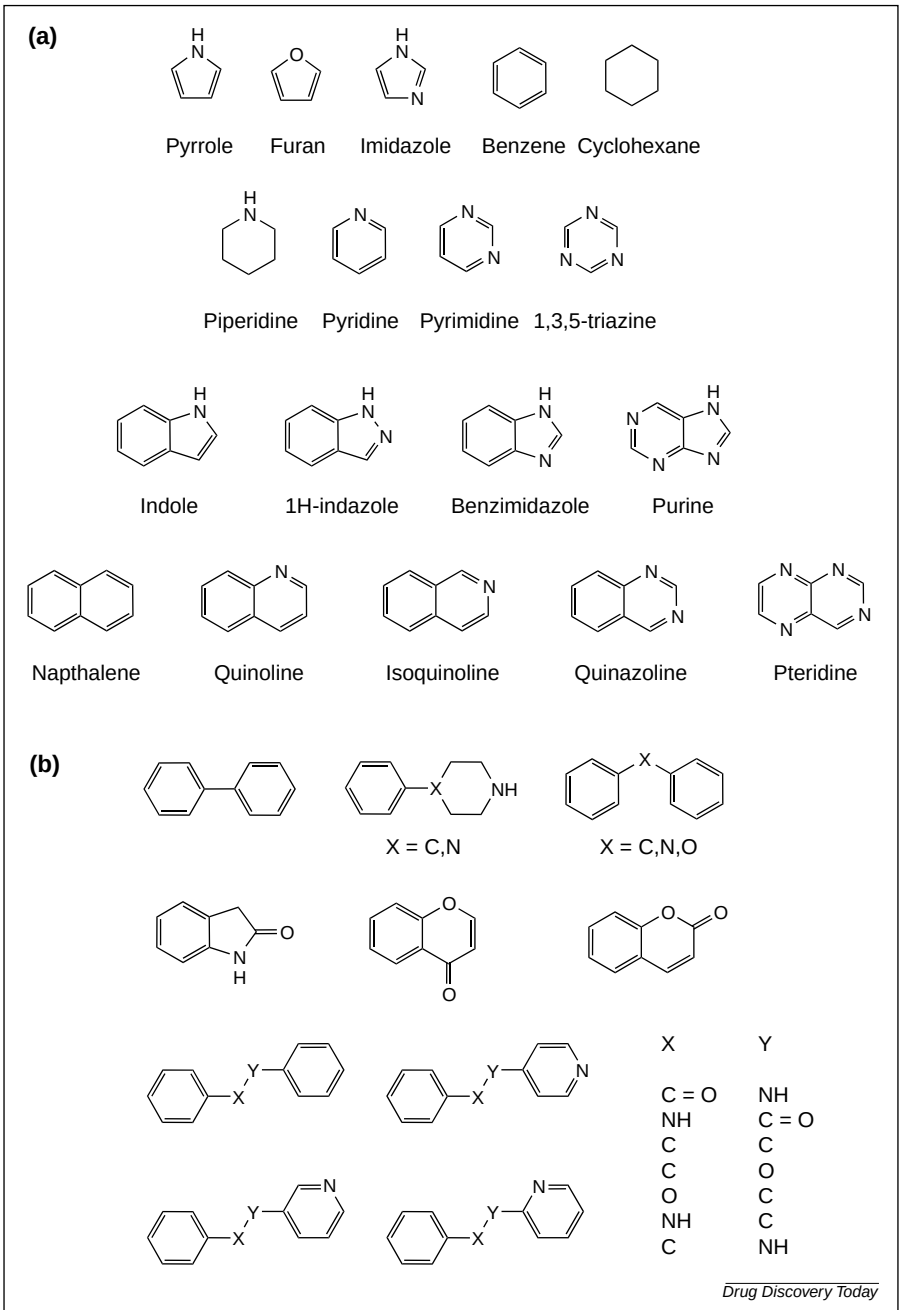

Figure 3. A molecular fragment library. A 'universal' fragment library can be constructed by enumerating simple scaffolds with commonly found side-chains as shown here. **(a)** Simple carbocyclic and heterocyclic frameworks. **(b)** Drug scaffolds [28].

The alternative approach of diversity-based screening can also be explored by generating molecular fragment libraries that contain common drug frameworks and sidechains (Fig. 3) [33,34]. This 'universal fragment set' is useful for the following reasons: (1) the compounds all have simple functional groups and, therefore, key interactions with the protein will be clear; (2) they are small and should be more soluble than larger compounds; and (3) they are free of arbitrary chemical functionality and thus it should be possible to make analogues through medicinal chemistry programs.

At Astex, as in other companies such as Abbott Laboratories (Abbott Park, IL, USA), our structure-based screening approach uses both knowledge-based and diversity-based strategies to generate libraries of molecular fragments to probe the active site of a protein. This process yields fragment libraries containing 100–1000s of compounds

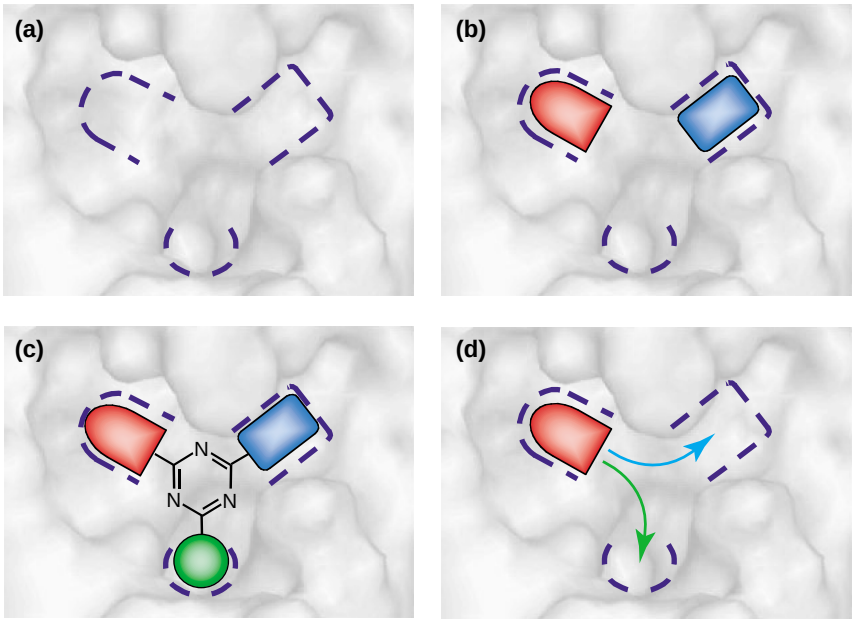

Figure 4. Structure-based fragment screening. **(a)** A protein with three different binding pockets. **(b)** Structure-based screening can identify molecular fragments that bind into one, two (shown) or all three pockets. **(c)** A lead compound can then be designed by arranging the fragments around a core template; or **(d)** growing out using iterative structure-based design from a single fragment.

that are either obtained from chemical suppliers or synthesized in-house. These compounds can then be cocktailed into groups of 4–10 so that each cocktail contains compounds with differing shapes and chemistries, before soaking into the protein crystal. Greer and colleagues have reported using this approach to help minimize the chances of multiple binders in one cocktail, and to make the visual analysis and interpretation more reliable [10]. Alternatively, computational methods can be employed to analyze the resulting electron density in an objective and automated manner. For example, AutoSolve® interprets electron density using an automated molecular-fragment matching and fitting process to rank candidate fragments in a cocktail.

Structure-based optimization of fragments

When the binding of one or more molecular fragments has been determined in the protein active-site, this provides a starting point for medicinal chemistry to optimize the interactions using a structure-based approach. The fragments can be combined onto a template or used as the starting point for 'growing out' an inhibitor into other pockets of the protein (Fig. 4). The potency of the original weakly binding fragment can be improved rapidly using iterative structure-based chemical synthesis. For example, in one of our lead discovery programs targeted against a protein kinase, we identified an initial fragment, AT464, which exhibited an IC_{50} value of 1 mM in an enzyme assay. Using the crystal structure of AT464 bound to the protein kinase we were able to improve the potency by >20-fold by synthesizing only 20 analogues. The resulting compound, AT660, had an IC_{50} value of 40μM (Murray *et al.*; unpublished data). Compounds from this novel lead series were then further optimized into nanomolar kinase inhibitors using rapid structure-based chemical synthesis. Using such a structure-based chemistry strategy, progressing from millimolar hits to nanomolar leads required the synthesis of <100 compounds.

Fragment-based screening can also be used to modify or replace substituents on drug-like lead compounds to improve pharmacokinetic properties. For example, over the past decade there has been much work on the design of thrombin and Factor Xa inhibitors. Early lead molecules often contained benzamidine in the S1 pocket and these usually displayed undesirable pharmacokinetic properties because of the high basicity and polarity of the amidine. Replacements for the benzamidine were sought and 1-aminoisoquinoline was identified as a suitable S1 binder. Several lead compounds now contain the 1-aminoisoquinolines, which have been claimed as a general replacement for benzamidines on the basis of their improved oral bioavailability [35]. Additionally, other S1 binders, such as anilines and aminopyridines, have been identified using structure-based fragment screening and these molecules are also known to be active replacements of amidines in thrombin inhibitors. Furthermore, Greer and colleagues have developed orally bioavailable inhibitors, containing a 2-aminoquinoline substituent, for urokinase using a fragment-based approach based on X-ray crystallography [10]. More recently, Fesik and co-workers described the design of inhibitors for adenosine kinase using NMR-based fragment screening [36].

The future

The pharmaceutical industry has struggled to improve its productivity in drug discovery over the past decade. This problem is predicted to worsen as novel types of therapeutic targets identified from the genome initiatives fail to yield to conventional lead discovery approaches. In this

scenario, alternative technologies for lead discovery are being developed and based on the success of early reports, structure-based fragment screening using NMR and/or X-ray crystallography shows great promise as a new approach for the discovery and optimization of lead compounds. Furthermore, these structure-based approaches for drug discovery will become even more powerful as more and more protein structures of key therapeutic targets are solved by the structural genomics initiatives around the world.

Acknowledgements

We wish to thank Mike Hartshorn and Ian Tickle, who developed the AutoSolve® software program, for providing data from fragment screening experiments. We also thank Chris Murray and Marcel Verdonk for contributions to the *in silico* screening methods and Anne Cleasby for data from X-ray crystallography experiments. And finally, we thank Emma Southern for assistance with the figures and compilation of this article.

References

1 Milne, G.M. (2001) Accelerating R&D productivity. Presented at *IBC's Drug Discovery Technology World Congress*, Boston, MA, USA, 13–15 August 2001
2 Campbell, S.F. (2000) Science, art and drug discovery: a personal perspective. *Clin. Sci.* 99, 255–260
3 Oprea, T.I. (2001) Is there a difference between leads and drugs? A Historical Perspective. *J. Chem. Inf. Comput. Sci.* 41, 1308–1315
4 Hann, M. *et al.* (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* 41, 856–864
5 Lipinski, C.A. *et al.* (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* 46, 3–26
6 Blundell, T.L. *et al.* High-throughput X-ray crystallography for drug discovery. *Proceedings of Royal Society of Chemistry* Meeting *Cutting Edge Approaches to Drug Design*, March 2001 (Flower, D., ed.) RSC Publications Dept, London (in press)
7 Shuker, S.B. *et al.* (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274, 1531–1534
8 Hajduk, P.J. *et al.* (1999) NMR-based screening in drug discovery. *Quart. Rev. Biophys.* 32, 211–240
9 Fejzo, J. *et al.* (1999) The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chem. Biol.* 6, 755–769
10 Nienaber, V.L. *et al.* (2000) Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nat. Biotechnol.* 18, 1105–1108
11 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
12 Blundell, T.L. (1996) Structure-based drug design. *Nature* 384, S23–S26
13 Abola, E. *et al.* (2000) Automation of X-ray crystallography. *Nat. Struc. Biol.* 7, 973–977
14 Jhoti, H. (2001) High-throughput structural proteomics using X-rays. *Trends Biotechnol.* 19, S67–S71
15 Heinemann, U. *et al.* (2001) High-throughput three-dimensional protein structure determination. *Curr. Opin. Biotech.* 12, 348–354
16 Blundell, T.L. *et al.* (2002) High-throughput crystallography for lead discovery in drug design. *Nature Rev. Drug Disc.* 1, 45–54
17 Lesley, S.A. (2001) High-throughput proteomics: protein expression and purification in the post-genomic world. *Prot. Exp. Purif.* 22, 159–164
18 Stevens, R.C. (2000) High–throughput protein crystallization. *Curr. Opin. Struc. Biol.* 10, 558–563
19 Adams, P.D. and Grosse-Kunstleve, R.W. (2000) Recent developments in software for the automation of crystallographic macromolecular structure determination. *Curr. Opin. Struc. Biol.* 10, 564–568
20 Norvell, J.C. and Machalek, A.Z. (2000) Structural genomics programs at the US National Institute of General Medical Sciences. *Nat. Struct. Biol.* 7, S931
21 Dry, S. *et al.* (2000) Structural genomics in the biotechnology sector. *Nat. Struct. Biol.* 7, S946–S949
22 Stevens, R. *et al.* (2001) Global efforts in structural genomics. *Science* 294, 89–92
23 Berman, H.M. *et al.* (2000) The Protein Data Bank and the challenge of structural genomics. *Nat. Struct. Biol.* 7, 957–959
24 Drews, J. (2000) Drug discovery: a historical perspective. *Science* 287, 1960–1964
25 Hendrickson, W.A. (2000) Synchrotron crystallography. *Trends Biochem. Sci.* 25, 637–643
26 Muchmore, S.W. (1999) Experiences with CCD detectors on a home X-ray source. *Acta Cryst. D* 55, 1669–1671
27 Muchmore, S.W. *et al.* (2000) Automated crystal mounting and data collection for protein crystallography. *Struc. Fold Des.* 8, R243–R246
28 Verlinde, C. *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
29 Walters, P. *et al.* (1998) Virtual screening – an overview. *Drug Disc. Today* 3, 160–178
30 Jones, G. *et al.* (1997) Development and validation of a genetic algorithm for flexible docking, *J. Mol. Biol.* 267, 727–748
31 Eldridge, M. *et al.* (1997) Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* 11, 425–445
32 Bissantz, C. *et al.* (2000) Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* 43, 4759–4767
33 Bemis, G.W. and Murcko, M.A. (1996) The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* 39, 2887–2893
34 Bemis, G.W. and Murcko, M.A. (1999) Properties of known drugs. 2. Side chains. *J. Med. Chem.* 42, 5095–5099
35 Rewinkel, J.B. *et al.* (1999) 1-Aminoisoquinoline as benzamidine isoster in the design and synthesis of orally active thrombin inhibitors. *Bioorg. Med. Chem. Lett.* 9, 685–690
36 Hajduk, P.J. *et al.* (2000) Design of adenosine kinase inhibitors from the NMR-based screening of fragments. *J. Med. Chem.* 43, 4781–4786