

HIGH-THROUGHPUT X-RAY TECHNIQUES AND DRUG DISCOVERY

HARREN JHOTI

Astex Technology Ltd, 250 Cambridge Science Park, Cambridge CB4 0WE, UK

E-Mail: h.jhoti@astex-technology.com

Received: 18th June 2002 / Published: 15th May 2003

BACKGROUND

In the past two decades the promise of structure-based drug design has continued to attract significant interest from the pharmaceutical industry. The initial wave of enthusiasm in the late eighties resulted in some notable successes, for example, the crystal structures of HIV protease and influenza neuraminidase were used to design Viracept and Relenza, both drugs currently used in anti-viral therapy (1, 2). However, although structure-based design methods continued to be developed, the approach became largely eclipsed in the early nineties by other technologies such as combinatorial chemistry and high-throughput screening (HTS) which seemed to offer a more effective approach for drug discovery. The goal of obtaining a crystal structure of the target protein, particularly in complex with lead compounds was regarded as a resource-intensive, unpredictable and slow process. During that period it was clear that protein crystallography was unable to keep pace with the other drug discovery technologies being performed in a high-throughput mode.

More recently, there has been resurgence in interest for using structure-based approaches driven largely by major technology developments in protein crystallography that have resulted in crystal structures for many of today's therapeutic targets. Furthermore, the ability to rapidly obtain crystal structures of a target protein in complex with small molecules is driving a new wave of structure-based drug design. In this chapter I will briefly describe some of these technology developments and focus on how they have enabled high-throughput X-ray crystallography to be applied to drug discovery.

TECHNOLOGY ADVANCES

There are many areas in which new technologies and methods are being developed to enable high-throughput structure determination by X-ray crystallography (3, 4). The process from gene to crystal structure is clearly multidisciplinary and advances in molecular biology, biochemistry, crystallisation, X-ray data collection and computational analysis underpin high-throughput X-ray crystallography. Many of these advances are being made in the public-initiatives focused on structural genomics. The most progressed and well-funded initiatives are found in the US where the NIGMS (National Institute of General Medical Sciences) is planning to spend US\$ 150M and is currently funding nine structural genomics centres under its Protein Structure Initiative (5). Similar programs are underway in other countries, for example, the Protein Structure Factory in Germany is focusing on solving structures of human proteins in collaboration with the German Human Genome Project (DHGP) and the Japanese government is supporting the RIKEN Structural Genomics Initiative.

The main focus of these structural genomics initiatives is to automate all steps of the protein crystallographic process and apply the methods to determine structures of proteins for which no three-dimensional information exists (6). In addition to these publicly-funded centres, some specialist biotechnology companies have also been formed to pursue structural genomics programs. These include Structural GenomiX and Syrrx, both based in San Diego (US), who are developing significant automation to streamline the gene to crystal structure process (7).

CLONE TO CRYSTAL

Expression, purification and characterisation of a novel protein in a quantity and form that is suitable for crystallisation and X-ray analysis probably occupies over 80% of the time in most structural biology groups. Consequently, methods for high-throughput parallel expression and purification are now being developed in many laboratories (8). Typically, 10-50 mgs of protein is required to screen sufficient numbers of crystallisation conditions to obtain initial crystals. Traditionally, a handful of different DNA constructs would be generated, after analysis of the protein sequence, in an attempt to remove flexible regions of the protein that may hinder crystallisation. Each construct would then be tested for expression in the host cell, usually *Escherichia coli* or insect cells, and the level of functional protein analysed using bioassay and polyacrylamide gel electrophoresis (PAGE).

In the past these different constructs would be analysed sequentially, but recent developments in molecular biology, based on DNA recombination, now enable high-throughput approaches for cloning and expression where tens to hundreds of DNA constructs can be easily generated to test in parallel for high expression. Protein purification has also seen significant improvements owing to the development of affinity tags that allow proteins to be purified significantly faster and more efficiently (9). Automated methods based on affinity chromatography, such as a nickel-nitrilotriacetic acid (Ni-NTA) column, are now available which can process samples in parallel using a 96-well format.

Crystallisation is often regarded as a slow, resource-intensive step with low success rates in obtaining good quality crystals. However, much of the failure during this step can be attributed to poor quality protein samples that often have some level of chemical or conformational heterogeneity.

The use of biophysical methods, such as dynamic light scattering, to rigorously characterise the protein sample is a key step before performing crystallisation experiments. Significant advances in automation have also improved the process of crystallisation with the new generation of robots able to efficiently sample the multidimensional space by varying precipitant concentration, buffers and pH - all variables known to affect crystallisation. Video systems are being developed that allow the user to monitor the crystallisation experiment using image recognition techniques (10).

CRYSTAL TO STRUCTURE

Once X-ray quality crystals have been grown, data collection using several wavelengths or derivatives is required in order to obtain the protein structure. X-ray data collection has been revolutionised in the last decade by both better X-ray sources and detectors. Third generation synchrotrons are now available across the world which provide high intensity X-ray beams allowing the data collection time to be significantly reduced (11). Synchrotron radiation coupled with charged-coupled device (CCD) detectors have allowed complete X-ray datasets for a crystal to be collected and processed within hours instead of days. High-throughput X-ray data collection has required the development of robotic systems that store and mount crystals sequentially while maintaining the samples at liquid-nitrogen temperatures (12, 13).

Phase determination has also become dramatically easier by the application of synchrotron radiation to single and multi-wavelength anomalous diffraction techniques, known as SAD and

MAD, respectively. Finally, new methods of electron density interpretation and model-building have allowed rapid and automated construction of protein models without the need for significant manual intervention (14).

STRUCTURE-BASED LEAD DISCOVERY

All these technology advances have resulted in an exponential increase in the number of crystal structures being deposited into the Protein Data Bank (PDB) in recent years (15). Currently, the PDB holds nearly 18,000 protein structures, most of which have been determined using X-ray crystallography (Fig1).

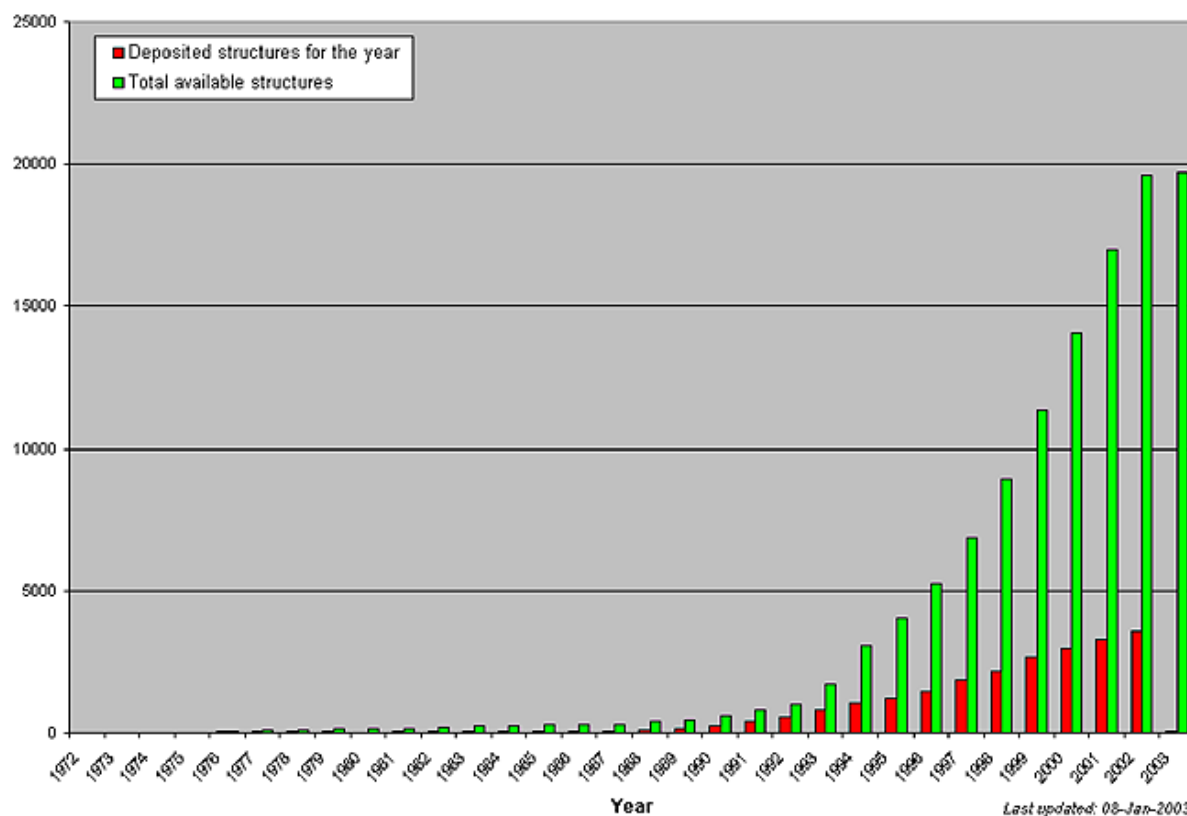


Figure 1. Growth in the Protein Data Bank. For many years the number of protein structures being determined and deposited into the PDB was linear, however, with the advent of major technology advances over the last decade the deposition rate has become exponential. (Source: The Protein Data Bank at www.rcsb.org; Berman *et al. Nucleic Acids Research*, **28** 235-242, 2000).

Due to this growing wealth of protein structure data, it is increasingly likely that the three-dimensional structure of a therapeutic target of interest to drug discovery scientists will already have been determined. Furthermore, it is expected that within the next five years, crystal structures of a large majority of the non-membrane protein targets of interest to the pharmaceutical industry will be available.

Although the structure of the native target protein is a useful start to guide a lead discovery program, the maximum value is derived only from structures of the protein in complex to potential lead compounds. This is due to the fact that many proteins undergo some level of conformational movement on ligand binding which has proved very difficult to predict from the native structure alone. Furthermore, water molecules often play a key role in the interactions between small molecules and proteins and their positions need to be established experimentally. The ability to rapidly determine crystal structures of protein-ligand complexes is required to effectively guide the lead optimisation phase, but may also allow X-ray crystallography to be applied to drug discovery in a new way: as a screening tool (4).

The most reliable approach to determine the structure of a protein-ligand complex, is either by co-crystallisation or by soaking the ligand into the preformed crystal. However, when X-ray crystallography is used as a method for ligand 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 analyse 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 the use of 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 Inc. (San Diego, CA, USA) and AutoSolve® from Astex (Cambridge, UK) can assist the crystallographer in the analysis and interpretation steps.

FRAGMENT-BASED LEAD DISCOVERY

There is growing interest in the use of molecular fragments for lead discovery. One reason for this interest is due to a problem that is evident in the nature of ‘hits’ identified from traditional bioassay-based High Throughput Screens (HTS). The average MW of successful drugs in the World Drug Index is in the low 300s, which is similar to the average MW in current corporate collections (16). This implies that corporate compound collections have evolved to be broadly “drug like” with respect to MW and other features. However, recent publications conclude that hits from a HTS should have a lower molecular weight than drugs, that is screening drug-like compounds may not be the most effective way to find good lead compounds (17). This conclusion is based on the expected increase in molecular weight, of about 80, during the lead optimisation process. Therefore, a HTS hit from a corporate compound collection with μM affinity towards the target may well already have an “average drug MW” yet it is likely that the

MW will increase very significantly during the lead optimisation process, leading to significantly poorer drug like properties with respect to solubility, absorption and clearance (18).

In order to address this issue several groups have been developing methods to identify low MW fragments (MW 100-250) that could be efficiently optimised into novel lead compounds possessing good drug like properties. These molecular fragments would by definition have limited functionality and would therefore exhibit weaker affinity (typically in the 50 μm -mM range). This affinity range is outside of the normal HTS sensitivity range and as such cannot routinely be identified in standard bioassays due to the high concentration of compound that would be required, interfering with the assay and leading to significant false positives. Rather than trying to push bio-assays into this affinity range, people are turning increasingly 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 (19, 20). In determining structure-activity relationships (SAR) 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. Once molecular fragments bound to the target protein have been identified they can then be linked together or 'grown' using structure-based chemical synthesis to improve the affinity for the target protein (Fig. 2).

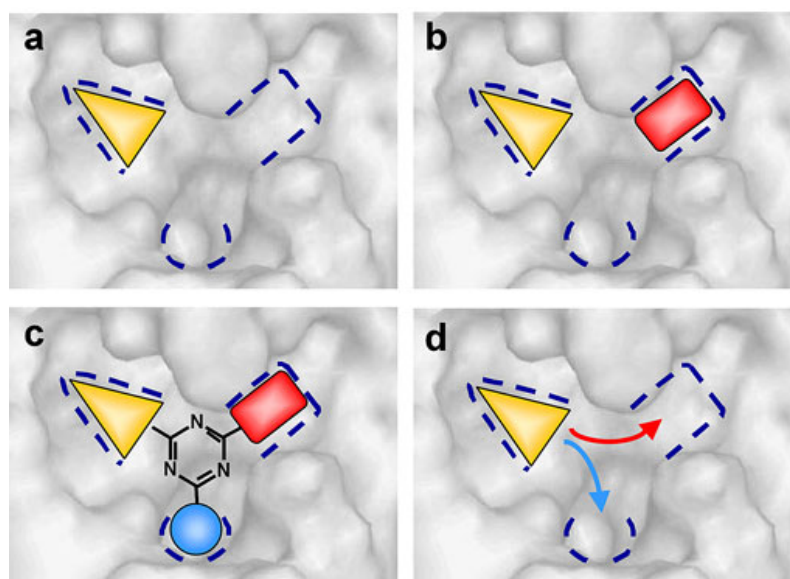


Figure 2. Once fragments have been identified bound into the active site they can be used as a start-point for iterative structure-driven chemistry resulting in a drug-size lead compound. If two fragments are bound in two different pockets (b) they could be used to decorate an appropriate scaffold (c). Alternatively, a single fragment could be rationally modified to occupy other neighbouring pockets (d).

FRAGMENT-BASED SCREENING USING X-RAY CRYSTALLOGRAPHY

X-ray crystallography has the advantage of defining the ligand-binding sites with more certainty than NMR and the binding orientations of the molecular fragments play a critical role in guiding efficient lead optimisation programs. Different sets of molecular fragments can be used to target a particular protein.

For example, in a screen of fragments against trypsin, a ‘focused set’ was selected based on known binders such as benzamidine, 4-aminopyridine and cyclohexylamine (21). These molecules were each used as starting points for similarity searches of chemical databases. Representatives from these searches were then purchased or synthesised and dissolved in an organic solvent (such as dimethylsulphoxide (DMSO)) added to a single protein crystal, and then left to soak for 1 hour to give the molecule time to penetrate into the active site.

The concentration of the molecular fragment is typically greater than 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 4-8 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 8 compounds each at a concentration of 30 mM.

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 (22). More recently, Greer and colleagues have described a method for screening using X-ray crystallography that focuses on soaking the target crystals with cocktails of compounds having differing shapes that can easily be distinguished by visual inspection of electron density (23). 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 (4). AutoSolve® allows 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 3.

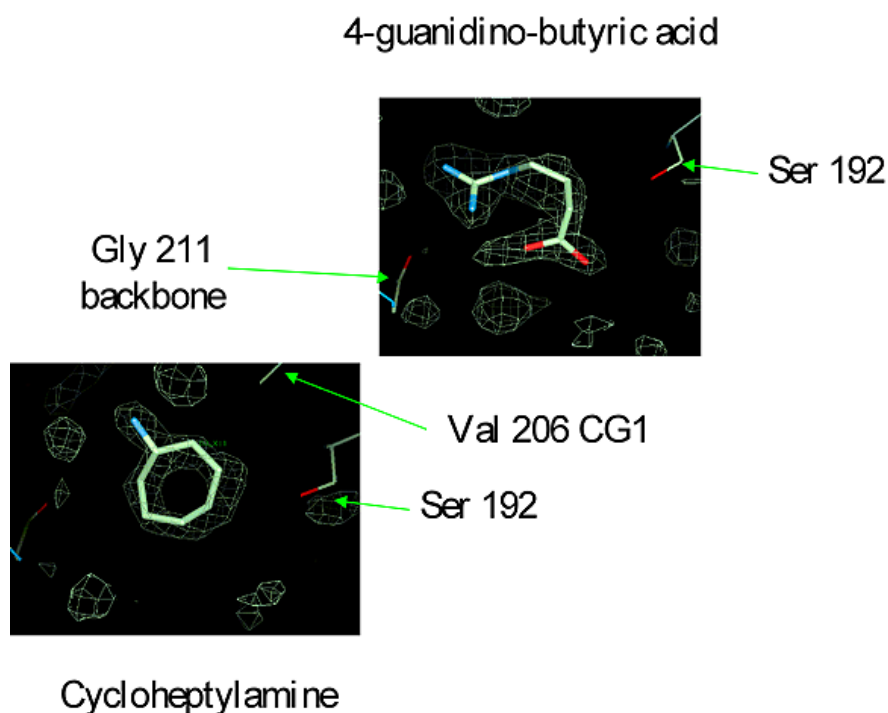


Figure 3. AutoSolve[®] interpretation of single compounds. Electron density can be automatically interpreted for small weak-binding fragments using AutoSolve[®]. Although the binding affinity is weak ($IC_{50} = 1$ mM for cyclohexylamine) the interactions with the protein are clearly defined.

In each case the binding mode of the small-molecule fragment is clearly defined by the electron density, which means that although the affinity may be in the millimolar range, the binding is ordered with key interactions being made between the compound and the protein. 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 (Fig 4).

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 and non-binding event depends solely 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 derive from a conventional combinatorial chemistry approach.

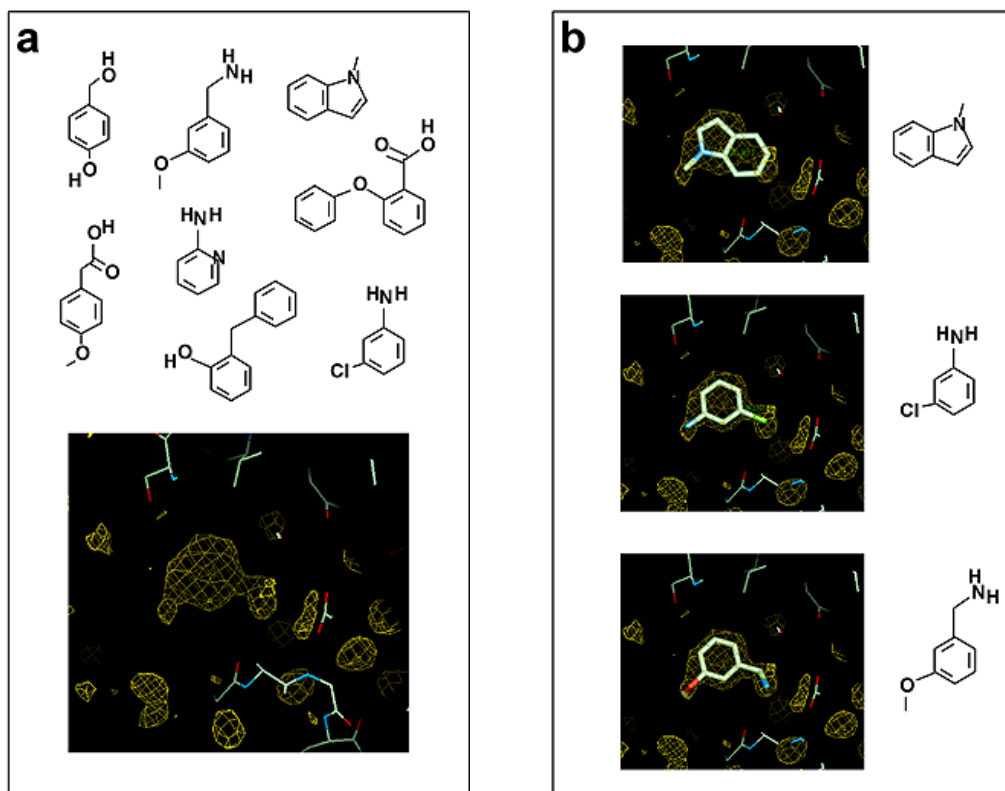


Figure 4. Analysing fragment cocktails using AutoSolve[®]. A crystal was exposed to a cocktail of 8 fragments and the resultant electron density is shown (A). Each of the eight molecules is fitted into the electron density by AutoSolve[®] and the optimal fit is identified by the program (B).

STRUCTURE-BASED LEAD OPTIMISATION

Determination of the binding of one or more molecular fragments in the protein active site provides a starting point for medicinal chemistry to optimise 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. 2). The potency of the original weakly-binding fragment can be rapidly improved using iterative structure-based chemical synthesis. For example, in one of our lead discovery programs targeted against p38 kinase, we identified an initial fragment, AT464 (MW=X), which exhibited an IC_{50} of 1 mM in an enzyme assay.

Using the crystal structure of AT464 bound to the protein kinase we were able to improve potency more than 20-fold by synthesising only 20 analogues. The resulting compound, AT660, had an IC_{50} of 40 μ M (unpublished results). Compounds from this novel lead series were further optimized to improve potency using rapid structure-based chemical synthesis. This resulted in the current lead compound, AT1731, which has an IC_{50} of 100 nM against the enzyme and is

active in inhibiting TNF release in LPS-stimulated cells. This improvement in affinity is produced by iteratively increasing the number of interactions between the protein and the compound (Fig. 5).

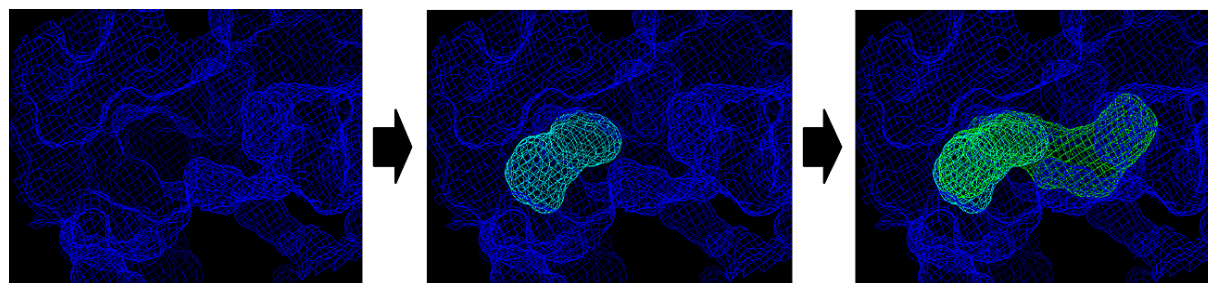


Figure 5. Optimisation of initial low affinity fragment into potent lead compound. The initial molecular fragment is used as a starting point from which extra protein/ligand interactions are built, guided by the 3-D structure of the protein. This can be seen in the increasing volume of occupation within the protein active site.

Using such a structure-based chemistry strategy, progressing from millimolar hits to nanomolar leads for our first lead series required the synthesis of <250 compounds. More recently, we have identified a second lead series for p38 kinase with a structurally distinct template, again by optimising a weakly-binding molecular fragment using structure-based synthesis.

CONCLUSIONS

The role of protein structure within the drug discovery process is likely to increase significantly over the coming years as more and more crystal structures become available for the therapeutic targets. This will no doubt fuel an increase in structure-based drug design programs which look to optimise lead compounds that were initially identified using traditional HTS campaigns. Recent technology advances in structure determination may also allow X-ray crystallography to be used as a method for ligand screening. This may have particular value for fragment-based lead discovery where the initial molecular fragments are likely to have an affinity too weak to enable detection using traditional bioassay-based methods. Initial data generated using X-ray crystallographic screening of molecular fragment libraries indicates that novel scaffolds can be identified and subsequently optimised using rapid structure-based synthesis to generate useful lead compounds. The potential of this fragment-based screening approach using X-ray crystallography may be significant, particularly against targets which have remained intractable using conventional screening methods.

ACKNOWLEDGEMENTS.

I wish to thank Drs. Mike Hartshorn and Ian Tickle who developed AutoSolve[®] and Dr. Robin Carr for useful discussions and for reviewing the manuscript. I also appreciate the assistance of Dr. Emma Southern in the production of this manuscript.

This manuscript first published in: Ernst Schering Research Foundation Workshop, Series Volume 42: Waldmann/Koppitz: Small Molecule Protein Interaction, Springer Verlag 2003

REFERENCES

- [1] Kaldor S. W. et al. (1997). Viracept (Nelfinavir Mesylate, AG1343): A potent, orally bioavailable inhibitor of HIV-1 protease. *J. Med. Chem.* **40**:3979-3885.
 - [2] von Itzstein, M. et al. (1993). Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **363**:418-423.
 - [3] Heinemann U. et al. (2001). High-throughput three-dimensional protein structure determination. *Curr. Opin. Biotech.* **12**: 348-354.
 - [4] Blundell T. L. et al. (2002). High-throughput crystallography for lead discovery in drug design. *Nat. Rev. Drug Disc.* **1**:45-54.
 - [5] Norvell J. C. & Machalek A. Z. (2000). Structural genomics programs at the US National Institute of General Medical Sciences. *Nat. Struc. Biol.* **7**:931.
 - [6] Vitkup D. et al. (2001). Completeness in structural genomics. *Nat. Struct. Biol.* **8**:559-566.
 - [7] Dry S. et al. (2000). Structural genomics in the biotechnology sector. *Nat. Struc. Biol.* **7**:946-949.
 - [8] Lesley S. A. (2001). High throughput proteomics: protein expression and purification in the post-genomic world. *Protein Exp. Purif.* **22**:159-164.
 - [9] Crowe J. et al. (1994). 6xHis-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification. *Methods Mol. Biol.* **31**:371-387.
 - [10] Stewart L. et al (2002). High-throughput crystallisation and structure determination in drug discovery. *Drug Disc. Today* **7**:187-196.
 - [11] Hendrickson W. (2000). Synchrotron crystallography. *Trends. Biochem. Sci.* **25**:637-643.
 - [12] Abola E. et al. (2000). Automation of X-ray crystallography. *Nat. Struc. Biol.* **7**:973-977.
 - [13] Muchmore S. W. et al. (2000). Automated crystal mounting and data collection in protein crystallography. *Structure* **8**:R243-R246.
 - [14] Perrakis A. et al. (1999). Automated protein model building combined with iterative structure refinement. *Nat. Struc. Biol.* **6**:458-463.
-

-
- [15] Berman H. M. (2000). The Protein Data Bank and the challenge of structural genomics. *Nat. Struc. Biol.* **7**:957-959.
- [16] Oprea T. I. (2001). Is there a difference between Leads and Drugs? A Historical Perspective. *J. Chem. Inf. Comp. Sci.* **41**:1308-1315.
- [17] Hann M. et al. (2001). Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comp. Sci.* **41**:856-864.
- [18] Lipinski C. A. et al. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development. *Adv. Drug Delivery Rev.* **46**:3-26.
- [19] Shuker S. B. et al. (1996). Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**:1531-1534.
- [20] Hajduk P. J. et al. (1999). NMR-based screening in drug discovery. *Quart. Rev. Biophys.* **32**:211-240.
- [21] Blundell T. L. et al. High throughput X-ray crystallography for drug discovery. *Proceedings of the Royal Society of Chemistry meeting Cutting Edge Approaches to Drug Design*, March 2001 (Flower, D ed.) RSC Publications Dept, London, (in press).
- [22] Verlinde C. et al. (1997). Antitrypanosomiasis drug development based on structures of glycolytic enzymes. *Structure-based Drug Design* (ed. Veerapandian, P) 365-394 (Marcel Dekker, Inc, New York, NY).
- [23] Nienaber V. L. et al. (2000). Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nat. Biotech.* **18**:1105-1108.
-