

Fragment Screening Using X-Ray Crystallography

Thomas G. Davies and Ian J. Tickle

Abstract The fragment-based approach is now well established as an important component of modern drug discovery. A key part in establishing its position as a viable technique has been the development of a range of biophysical methodologies with sufficient sensitivity to detect the binding of very weakly binding molecules. X-ray crystallography was one of the first techniques demonstrated to be capable of detecting such weak binding, but historically its potential for screening was underappreciated and impractical due to its relatively low throughput. In this chapter we discuss the various benefits associated with fragment-screening by X-ray crystallography, and describe the technical developments we have implemented to allow its routine use in drug discovery. We emphasize how this approach has allowed a much greater exploitation of crystallography than has traditionally been the case within the pharmaceutical industry, with the rapid and timely provision of structural information having maximum impact on project direction.

Keywords Fragment-based drug discovery · Structure-based drug design · X-ray crystallography

Contents

1	Introduction	34
2	The Pyramid Process for Fragment-Screening	35
2.1	Introduction to Pyramid	35
2.2	Fragment Libraries	37
2.3	Fragment Screening	41
3	Examples of Fragment Screening	52
3.1	Fragment–Protein Interactions	52
3.2	Hits-to-Leads Case Studies	53
4	Conclusions	55
	References	56

1 Introduction

The last 10 years have seen increasing acceptance of the fragment-based approach as an important part of modern drug discovery [1–3]. As reviewed by Erlanson [4], fragment-based approaches, which involve the detection and elaboration of simple, low molecular weight chemical start-points, offer a number of advantages over conventional HTS-driven paradigms [5, 6]. These include a more efficient sampling of chemical space [7, 8], a higher hit-rate due to lower molecular complexity [9], and a greater “efficiency” in binding, giving greater scope for controlling important compound properties (e.g., molecular weight and lipophilicity) during hit and lead optimization [10–13].

Historically, the key technical challenge for this approach was the detection of fragment hits, largely due to the fact that conventional bioassay-based methods are often unsuitable for screening such weakly binding compounds. Over the past decade, this issue has been successfully addressed using a variety of biophysical methods for detection [14], of which NMR [15–20] and surface plasmon resonance (SPR) [21–23] have perhaps been the most widely adopted. Indeed, many researchers pinpoint the start of fragment-based approaches to the use of protein-observed NMR to detect fragment binding by researchers at Abbott [24].

Arguably, the use of X-ray crystallography to detect the binding of small, low molecular ligands pre-dates this, with the seminal work of Ringe [25, 26] and others [27, 28], who highlighted the ability of organic solvents to map energetically important hot-spots on protein surfaces. In addition, Hol et al. published results from some of the earliest fragment-soaking experiments against crystals of the anti-parasitic target triose-phosphate isomerase from *Trypanosoma brucei* [29, 30]. During the early 2000s, interest in fragment-based approaches increased and X-ray screening was established in several industrial laboratories, including Astex [31–34], Abbott [35] and SGX (now part of Eli Lilly) [36, 37]. However, a shift away from its use as a primary screen has been evident in recent years, and it is now more usually used in conjunction with other techniques, and typically downstream of a biophysical pre-filter [38]. Indeed, a combination of multiple, “orthogonal” techniques has important advantages, and this approach is discussed in more detail by Wyss et al. [39] and Hennig et al. [40]. Despite this, X-ray crystallography remains one of the most sensitive of the biophysical techniques within the practical constraints of a typical fragment-screening experiment [41, 42]. In principle, there is no theoretical lower limit on the affinity of fragments detectable, with the main practical limitations being compound solubility and crystal robustness. In practice, with careful choice of fragment library (see Sect. 2.2), this allows reliable detection of compounds with a dissociation constant (K_d) > 5 mM, a regime that may not be accessible for all targets using other methods. For this reason, at Astex we have maintained X-ray screening as an important component of our fragment-based approach, albeit alongside full integration with other biophysical screening techniques such as NMR and thermal shift [3].

In addition to its sensitivity, the use of crystallography as a screening technique has a number of other advantages over alternative methods. Of key importance is the provision of precise structural information on the interaction between fragment hit and target at the earliest possible stage in a screening cascade. Thus, the technique not only provides an efficient means to detect weak binders, but also allows for the most rapid and efficient assessment of hits in terms of their medicinal chemistry tractability and utility, particularly in terms of synthetic vectors that are likely to yield to optimization by structure-based design techniques. In many ways, it is the most “natural” technique for an approach in which the downstream use of structural information (e.g. during fragment elaboration) has been shown to be so important. In addition, crystallography does not suffer from the problem of false positives, which are intrinsic to most other screening techniques. Potential disadvantages of fragment-screening by X-ray crystallography include the possibility of missing potential hits (false negatives), either due to occlusion of binding sites by crystal contacts, or because ligand binding requires protein conformational changes that are not tolerated within the crystalline environment. Nevertheless, in our experience, these issues have not been limiting, and can often be addressed through the use of alternative protein constructs and/or crystal forms.

A second perceived disadvantage has been relatively low throughput of X-ray crystallography as a technique compared to other methods such as NMR [41]. In this review we describe how we have successfully addressed this issue, allowing the power of X-ray based screening to be realized as a highly viable component of drug-discovery in a process which we call “Pyramid” [43–45]. We present a discussion of the issues involved in using crystallography as a high-throughput screening technique, the technology developed to address these, and case studies of fragment hits which have been successfully developed into clinical compounds. Where possible, we place the procedures and developments made at Astex in the context of progress made by the field of high-throughput crystallography as a whole. A further perspective on the use of fragment-screening by X-ray crystallography is provided by Bauman et al. [46], as applied to HIV therapeutic targets.

2 The Pyramid Process for Fragment-Screening

2.1 Introduction to Pyramid

Protein crystallography has historically been a relatively “low throughput” technique, and its use and impact within the pharmaceutical industry has generally been limited to the lead optimization phase. The key issue to be addressed in transforming it into a technique suitable for screening has been to decrease the time taken to generate structural information on protein–ligand complexes, as well as the implementation of a work-flow and informatics infrastructure to facilitate the handling of the resulting structural information. Although the following sections discuss the typical work flow

in the context of direct X-ray screening, it should be emphasized that many of the issues addressed here (e.g. speed and effective dissemination of structural information) also have relevance to expediting alternative screening cascades in which hits from a biophysical pre-filter (e.g. NMR) are subsequently examined by crystallography. As discussed in Sect. 1, we typically carry out fragment screening using a number of other biophysical techniques in addition to direct protein–ligand X-ray crystallography. This allows us the greatest degree of flexibility in screening, but also recognizes that the relative sensitivity of a particular technique is frequently target-specific. Nevertheless, at Astex, we do not consider a fragment hit to be “validated” and suitable as a starting point for medicinal chemistry until it has been observed to bind by crystallography. Again, this recognizes the important role that crystallography can play in filtering possible “false-positives” detected by other biophysical techniques, as well as highlighting the key role that structural information plays in guiding hit progression.

A flow-chart for a typical crystallographic fragment-screening experiment is shown in Fig. 1. Briefly, it involves the soaking of crystals with fragments of interest, followed by X-ray data collection and processing, placement of water molecules in the electron density, and refinement of the ligand-free complex to potentially reveal the difference electron density associated with the bound ligand. The electron density is then interpreted, fitted, and the complex further refined to give the final protein–ligand structure. The Pyramid approach to fragment-based

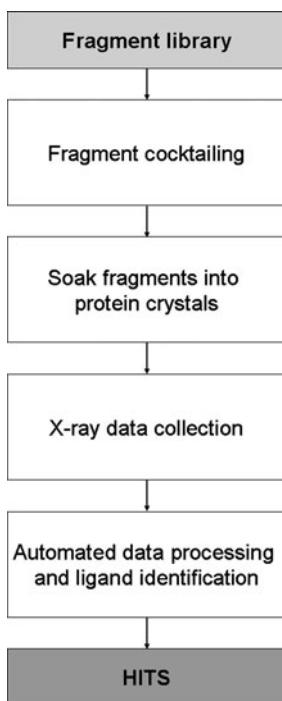


Fig. 1 Work-flow for a typical crystallographic fragment screen

discovery at Astex has streamlined many of the steps involved in the above procedure. In particular, it has relied on the development of high quality fragment libraries, and automated protocols for rapid X-ray data collection, processing and structure solution. The development of the various steps in our Pyramid approach are explained in more detail below.

2.2 *Fragment Libraries*

2.2.1 Overview

The composition of the compound libraries to be screened is a crucial part of fragment-based drug discovery. There are two complementary approaches that might be taken in their design and assembly. The first attempts to provide a general purpose library, with diverse coverage of chemical space, and hence is suitable for screening against any target. The second, a targeted or focussed library, provides a set of compounds that are tailored for a particular target. In practice, this latter approach relies on some kind of prior knowledge as to the sort of chemical moieties and interactions likely to provide affinity for the protein of interest, but can be very helpful for expansion around initial fragment hits, or for cases where hit rates from a general library are particularly low. For both types of library, the aim is to produce a set of screening compounds that are as small and simple as possible, to maximize the chance of a binding event.

Examples of both approaches towards library design have been described in the literature [35, 47–50], and commercially available fragment libraries are also now available as described by Bauman et al. [46]. We next review the approach taken towards fragment-library generation at Astex.

2.2.2 Astex Core Fragment Set

Astex's Core Fragment Set (CFS) is a general purpose library of approximately 1,000 fragments, which aims to effectively cover chemical space and be suitable for screening against a diverse range of targets. The assembly and refinement of Astex's fragment libraries has been an ongoing process, and the current CFS has evolved in part from Astex's original Drug Fragment Set (DFS) [43], in addition to a number of other fragment libraries. The DFS was a general-purpose library based on the idea that "drug-fragment space" can be effectively sampled with a relatively small number of compounds based on scaffolds and functional groups commonly found in drug molecules [15, 51, 52]. Since Astex's inception, the fragment libraries have undergone several iterations and improvements, and we now provide an overview of our approach.

The first stage in constructing the original DFS was to identify a set of frequently occurring simple organic rings systems found in known drugs. Several studies have shown that drugs contain only a relatively small number of such scaffolds, and their

selection as a basis for a fragment library may confer the advantage of a lower likelihood of toxicity, as well as being more amenable to medicinal chemistry. These ring systems were also complemented with a further set of simple carbocyclic and heterocyclic fragments to provide increased coverage of chemical space (see Fig. 2a, b).

A virtual library, from which the DFS was selected, was then generated by combining the ring systems described above, with a set of desirable side-chains (Fig. 2c, d). These included a set of side-chains found in existing drugs, as well as additional hydrophobic and nitrogen-containing substituents which were designed to pick up specific interactions within protein active sites. Enumeration of the virtual library was then carried out by substituting the side-chains onto the ring systems. Each ring carbon atom was substituted with side-chains found in known drugs and by the lipophilic side chains, whilst ring nitrogens were substituted by side-chains from the nitrogen-substituent group. With the exception of benzene and imidazole, each ring system was substituted at only one position at a time. This resulting virtual library consisted of 4,513 fragments, of which 401 were commercially available. Removal of insoluble compounds and known toxophores resulted in the original DFS of 327 compounds.

A second version of the DFS was constructed in a similar way to the first, but with a revised and enlarged set of scaffolds and side-chains from known drugs and leads, and more stringent control of physicochemical properties of its members. In particular, a retrospective analysis of hits against various in-house targets had shown that the most useful fragments have physical properties that lie within a limited range. These criteria are shown below, and we term these properties the “rule-of-three” [53], by analogy with Lipinski’s rule-of-five for orally available drug-like compounds:

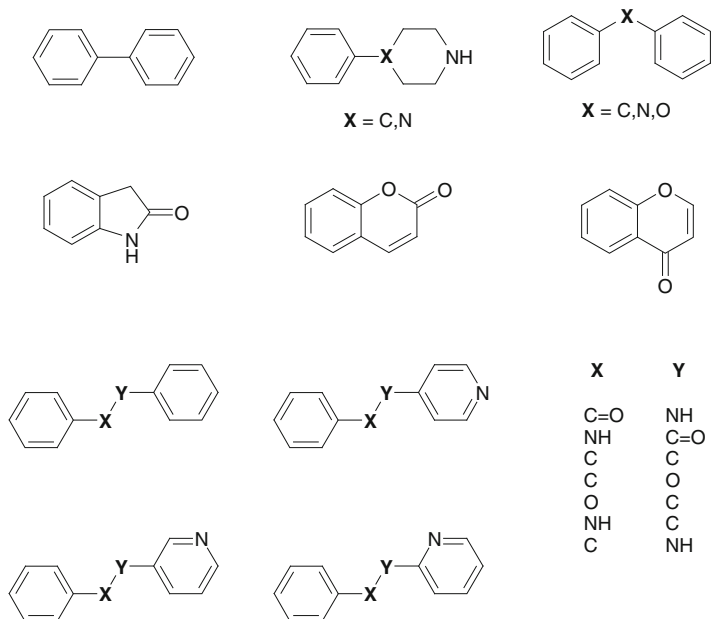
- Molecular weight ≤ 300
- Number of hydrogen-bond donors ≤ 3
- Number of hydrogen-bond acceptors ≤ 3
- clogP (computed partition coefficient) ≤ 3.0

Other criteria identified include polar surface area (PSA) $< 60 \text{ \AA}^2$, and the number of rotatable bonds ≤ 3 . These rules have since been adopted widely by the fragment-based community in general.

The rule-of-three was used to filter an enlarged virtual library to give approximately 3,000 compounds. Compounds were selected from this new set if they were commercially available, or easily synthesized by simple functional group interconversion from available analogues. In order to maximize our coverage of chemical and interaction fragment space, the compounds were then clustered using topological fingerprints [54]. By comparison with the initial DFS, this process allowed an examination of areas of chemical space that were under- or over-represented, and cherry-picking by experienced medicinal chemists and modellers yielded a revised set with improved properties.

Astex’s fragment libraries have continued to evolve, and have now been consolidated to give the current CFS. An important part of this has been a thorough

a



b

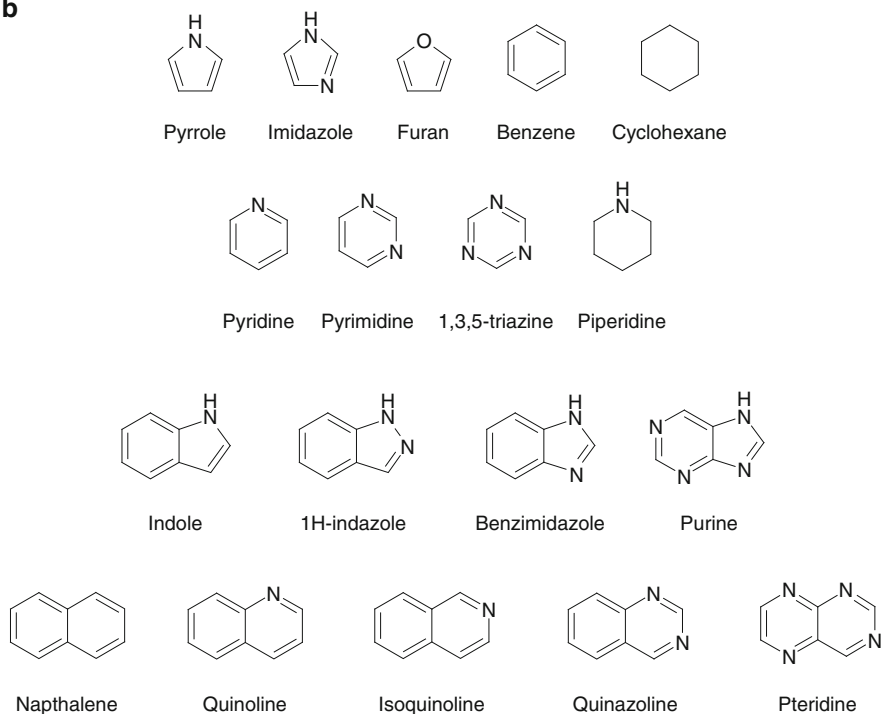


Fig. 2 (continued)

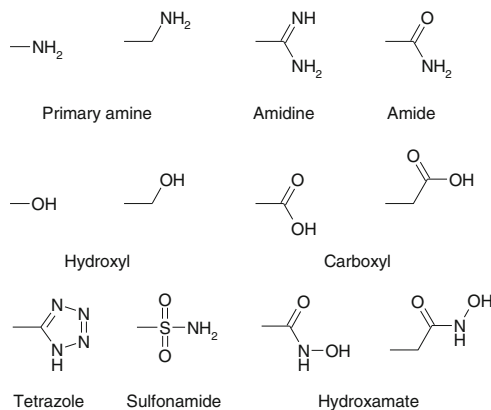
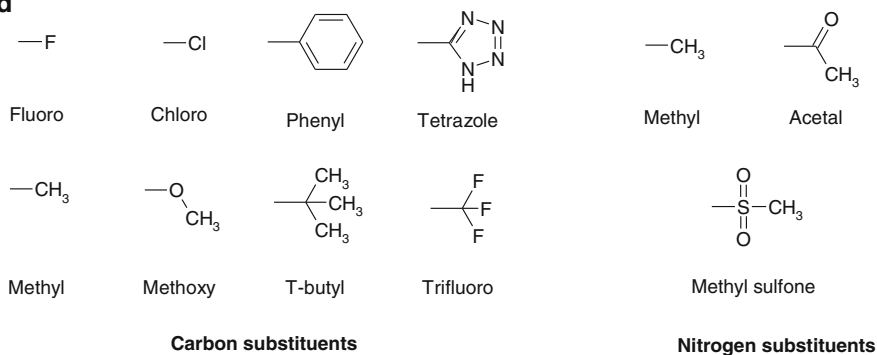
c**d**

Fig. 2 (a–d) Commonly occurring ring systems and side-chains used in the construction of the general purpose DFS

review of fragment performance against a range of target classes to ensure that the CFS provides the most efficient coverage of chemical and interaction space. Its composition has been chosen in the light of previous screening hit rates, and the range of compounds has been increased to encompass a greater proportion of non-commercially available molecules. Coverage of chemical space has been further improved by increasing the number of fragments that possess a greater degree of three-dimensional shape, and by introducing fragments with the potential for enhanced binding to protein–protein interaction targets.

The current CFS has a mean molecular weight of approximately 170, a mean heavy atom count of 12 and mean clogP of 0.9. Approximately 45% of the set has been previously observed to bind by X-ray crystallography, and components of oral drugs, natural product scaffolds and chiral building blocks are all well represented. In addition, the set has been through stringent quality control procedures to ensure that fragments are 90% pure, and meet minimum stability and solubility requirements, both in DMSO and in aqueous solution.

2.2.3 Targeted Fragment Libraries and Virtual Screening

In addition to the CFS described above, smaller focussed sets are frequently generated for screening against a particular target. For example, a focussed kinase library might be constructed by simple substructure searching for fragments containing motifs that would be expected to satisfy the conserved set of hydrogen bonds that are frequently observed between kinase inhibitors and the protein hinge region. Structure-based virtual screening can then be used to refine this list of compounds by docking the compounds into the protein of interest. The docked protein-bound ligand is visualized to examine its putative fit and complementarity with the active site, its ability to form interactions known to be important to binding, and the availability of synthetically accessible vectors for further development.

The starting point at Astex for constructing a focussed set is typically through searching a database of more than 3.6 million unique commercially available compounds called ATLAS (Astex Technology Library of Available Substances) [43]. ATLAS can be queried using substructure filters and physico-chemical property filters (such as molecular weight, clogP, PSA, etc) to produce a list of commercially available fragments meeting specific user requirements. These compounds can then be automatically docked into the active site of the target of interest, using a proprietary version of GOLD [55, 56] with a choice of scoring functions [57, 58]. The results from virtual screening runs can subsequently be post-processed using a web-based interface, allowing the user to select subsets of compounds for visualization and purchase using various filters, including the presence of specified interactions between fragment and active site residues [59]. This approach has proved to be very powerful, although the scoring functions used to drive the docking have several limitations. For this reason, manual selection of docked compounds remains an important part of this process. A more extensive discussion of the use of fragment docking and virtual screening is given by Rognan [60].

2.3 *Fragment Screening*

2.3.1 Overview

The most resource-effective method of obtaining structures of a protein–ligand complex is by soaking the ligand of interest into apo protein crystals. This is usually achieved by placing a single crystal in a high-concentration solution of ligand for a suitable length of time, allowing the ligand to diffuse through the solvent channels in the crystal and bind at energetically favourable sites. When screening for fragments, high compound concentrations (50 mM or more) in the soak solution are typical, and reflect the thermodynamic requirements anticipated to achieve near full occupancy for low affinity ligands. For practical purposes, a ligand concentration tenfold greater than the IC_{50} or K_d (giving a theoretical occupancy of approximately 90%) is usually sufficient. Fragments are typically soaked in a solution based on the

chemical composition of the mother liquor, but frequently modified to increase crystal stability and longevity during the soak. Indeed, investigation of a variety of soaking conditions is an important part of optimization experiments, which are carried out before fragment-screening can take place. Ligand stocks are often formulated in DMSO, and therefore the final soak generally contains 1–10% organic solvent. Where such levels of solvent are found to have a detrimental effect on diffraction, it can be useful to add DMSO during the crystallization process, producing crystals that may be more tolerant of its presence during subsequent soaking. It is also advantageous to include a cryoprotectant in the compound soaking solution if possible, to avoid further manipulations at the crystal freezing stage.

An alternative procedure for obtaining structures of protein–ligand complexes is co-crystallization, in which the protein–ligand complex is prepared in the aqueous phase, and then crystallized with the ligand *in situ*. This method is less suitable for high-throughput fragment screening, because a separate crystallization experiment is effectively needed for each compound. This procedure can be further complicated if the presence of a ligand results in a change in the crystallization conditions. In addition, co-crystallization is not optimal for determination of weakly binding fragments because the high concentration of ligand needed to fully occupy the binding site can interfere with the crystallization process itself. It should be noted, however, that some proteins will not crystallize without the presence of a ligand, perhaps due to an ordering effect on mobile regions. In these cases, co-crystallization on a “per ligand” basis is the most likely alternative option, although it is sometimes possible to co-crystallize with a single, relatively weakly binding compound, and then “back-soak” or exchange with new ligands in the more usual soaking format. This approach was successfully used at Astex to generate structural information for inhibitors binding to the kinase Akt [61–63]. In addition, co-crystallization can be used in cases where fragment soaking causes crystals to crack, presumably by inducing conformational changes or binding at crystal contacts. Finally, we note that the testing of several protein constructs and/or crystal forms can sometimes be important in achieving a system suitable for robust and high-throughput protein–ligand crystallography.

2.3.2 Fragment Cocktailing

The efficiency of fragment screening can be increased substantially by pooling or cocktailing the compounds in the library [29, 35, 43]. Identification of the bound fragment at the end of the X-ray experiment then becomes a case of determining the best fragment-fit to the electron density. Assuming that compound binding occurs, one can imagine three potential outcomes of a cocktailed X-ray experiment [41]. In the first scenario, only one fragment binds to the protein, its identity being unambiguously determined from the electron density. In a second scenario, removal of the initially identified fragment from the cocktail reveals the binding of secondary or even tertiary binders, and in this case the soaking is effectively a competition experiment. A third situation occurs where the final difference electron density can

be explained by the simultaneous binding of more than one fragment with similar affinities. In these latter cases, rounds of “deconvolution” are necessary to extract all relevant information, which can partially negate the benefits of cocktailing.

The number of compounds per cocktail is a balance between the high concentrations required for sensitive detection, and total organic load. For these reasons, as well as ease of data deconvolution, cocktailing at Astex is usually performed in sets of four, with the selected components chosen to be as chemically diverse as possible within a particular cocktail. This diversity has the effect of reducing the number of hits per cocktail, as well as increasing the shape diversity, which expedites the automated interpretation of ligand electron density (see also Sect. 2.3.5 “Automated ligand fitting and refinement”). The Nienaber group at Abbott [35], and the Hol group at the University of Washington [64] have also described a similar use of fragment cocktailing using shape-diverse compounds.

At Astex, the initial partitioning of fragments into cocktails is achieved using a computational procedure that minimizes chemical similarity [43]. Fragments are described as feature vectors, which encode such properties as the number of donors/acceptors/non-hydrogen atoms, number of five- and six-membered rings and their substitution patterns. The chemical dissimilarity between two molecules, $d(i, j)$, is then calculated as the distance between the two vectors.

The number of unique ways that N compounds can be partitioned between n cocktails, each containing c compounds is given by:

$$\frac{N!}{n!(c!)^n}.$$

This number increases extremely rapidly with increasing library size, dictating an efficient algorithm to solve the problem. Our partitioning procedure [54] starts from a matrix that describes the dissimilarities between all compounds in the library of interest. Starting from an initially random assignment of compounds to cocktails, the cocktail score, S , is calculated as follows, where the first summation runs over all n cocktails, and the second over all compound pairs in a particular cocktail:

$$S = \sum_{c=1}^n \sum_{i,j \in c} d(i,j).$$

The score is then maximized using a procedure that swaps pairs of compounds in different cocktails. Swaps are accepted if the score remains the same or increases, with termination after 10,000,000 iterations, or 100 compound swaps that did not improve the score. A similar approach is also discussed by Bauman et al. [46].

2.3.3 X-Ray Data Collection

High-throughput screening of fragments using crystallography requires rapid and efficient X-ray data collection, either in-house or at a synchrotron radiation source.

Many of the recent developments in hardware have been driven by the need to streamline and improve data collection at synchrotron beamlines where new third-generation sources, producing brighter and better collimated X-ray beams, allow higher quality data to be collected more rapidly [65]. The rate-limiting step at third-generation synchrotrons is frequently the manual intervention required to change samples, where the time taken to mount and align crystals can easily exceed half that required to collect the data. As a result, most synchrotrons have now developed automatic sample changers and integrated them into their data collection systems. Their use has dramatically increased the throughput available, with typically around 100 protein–ligand datasets collected during a 24-h synchrotron trip. Increased synchrotron automation has also allowed the development of “service crystallography” such as MXpress (ESRF), freeing users from the more tedious aspects of routine data collection.

Commercially available sample changers such as ACTOR (Rigaku MSC), MARCSC (Marresearch) and BruNo (Bruker AXS) are also now readily available and increasingly utilized in the “home” laboratory setup where they have been a key step in the realization of high-throughput data collection in-house [66]. For example, at Astex we have reported collection of X-ray data from 53 crystals of protein tyrosine phosphatase 1B in approximately 80 h using ACTOR [67, 68], with near-continuous use on a range of projects.

Other developments in X-ray hardware have also had an important impact on the ability to collect rapid in-house diffraction data. The latest generation of high-intensity X-ray generators (such as Rigaku’s FR-E), coupled with steady improvements in X-ray optics, have revolutionized in-house X-ray equipment to the point where the beam intensity has become comparable to that obtainable at some synchrotrons. Parallel advances in X-ray detector design have resulted in a new generation of detectors based on charge-coupled devices (CCDs) such as the Quantum 315 Area Detector Systems Corporation (ADSC) and the PILATUS (SLS), which are larger, more sensitive and have a faster readout. In the case of the PILATUS, readout time has been reduced to a level where shutter-less data collection has become possible, giving a significant increase in data quality and speed. Coupled with stabilization of cost, the use of CCDs has increased, and combined with brighter rotating-anode generators they are an important component of a high-throughput setup in a commercial laboratory. At Astex, the high speed provided by Saturn and Jupiter CCDs, with FR-E+ source (Rigaku) is combined with two R-Axis HTC image plates (Rigaku) to give a flexible setup for routine high-throughput data collection.

Advances in the hardware involved in automating data collection demand a parallel development of software to control the system. The goal of many synchrotrons and/or hardware suppliers has been to develop “smart” systems that can encompass sample tracking, control of crystal mounting and aligning, evaluation of experimental strategy based on initial images, data collection, and finally integration, scaling and reduction of experimental intensities [69]. Aspects of these requirements have been incorporated into such synchrotron software as Blu-Ice [70], mxCuBE/DNA [71] (ESRF) and EDNA/XIA2 (Diamond), with the additional capability to allow full-remote collection of data over the Internet. At Astex, in-house

hardware control is achieved through the ACTOR-associated software Director (Rigaku MSC), coupled with the integration and scaling software d*TREK [72] as implemented in the CrystalClear package (Rigaku MSC). “Off-line” processing is also provided for with automated versions of the XDS [73] and Mosflm [74, 75] packages, as described further in Sect. 2.3.4.

2.3.4 Automation of Data Processing

Data processing, structure solution, refinement and analysis have traditionally been a major bottleneck in the rapid use of X-ray data. Automation of these steps, combined with the full integration of the resulting information within an easily queried database environment has perhaps been the single most important factor in the application of crystallography as a primary screening technique at Astex. The various stages involved in our automated data-processing procedure are shown in Fig. 3 and will be briefly described below. Implicit in this approach is the

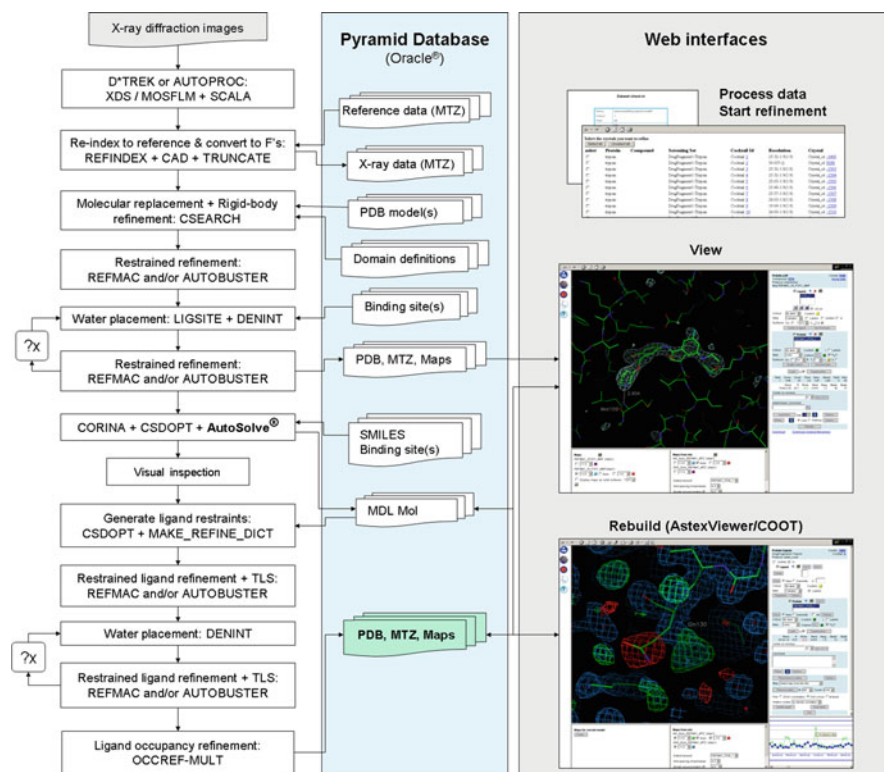


Fig. 3 Flow-diagram summarizing the AutoSolve platform and its automated data processing, refinement and ligand placement procedures. All data handling is carried out within an Oracle database, and the process is driven from a series of web-based interfaces

availability of a suitable protein starting model for phasing, in the same space group and isomorphous to (or nearly so) the protein–ligand complex crystal.

We have used commercially available software components wherever possible, for example programs in the CCP4 [76] and the Global Phasing suites. However, at the time our processing pipeline and database management system were developed, no suitable crystallographic software was available for a number of functions, which were additionally required to be run in batch mode with a high degree of reliability. Consequently, software to implement auto-re-indexing, limited search molecular replacement, multiple structure superposition, automated model selection, automated water-placing, binding-site cavity detection, ligand geometry optimization, automated ligand fitting into electron density, ligand restraint-dictionary generation and ligand-occupancy refinement all had to be developed in-house. We note that more recently, a number of commercially available ligand-fitting programs have become available, including Rhofit (Global Phasing), PrimeX (Schrodinger) and Afitt (OpenEye) [77], as well as within the Phenix suite [78, 79], ARP/wARP [80] and Coot [81, 82].

Automated data processing at Astex typically starts with the integration of in-house or synchrotron-collected data using the AutoPROC script from Global Phasing. This provides a “wrapper” for either Mosflm or XDS, followed by the data-scaling and merging program Scala (CCP4), and in the majority of cases provides high quality integrated data with no intervention from the user. Recently, there has been a move towards provision of initial data-processing capability at synchrotron beam-lines using computers with fast parallel processors, and we have found that this relieves much of the burden of processing large quantities of synchrotron data in-house. The pre-processed data, or data from AutoPROC, are passed to a batch-mode script responsible for handling re-indexing to a common reference frame and conversion of experimental intensities to amplitudes (implemented by the CCP4 programs Refindex, Sortmtz, CAD and Truncate), for all the datasets collected.

The initial data processing is followed by a limited-search 6D molecular replacement, i.e. combining the traditional 3D rotation and translation functions into a single six-parameter search for each protomer in the asymmetric unit of the crystal, but only considering orientations and positions close to that of the starting model. This limited-search protocol is both faster and more reliable than the traditional separate full-search rotation and translation functions as implemented in programs such as AMoRe [83] or Phaser [84]; however, it is reliant on the data having been re-indexed to a common reference frame. Additionally, it completely avoids the common problem of the final model being shifted to an alternate origin and/or asymmetric unit, which is a frequent issue with the full-search protocol. We provide the option to use more than one protein starting model in molecular replacement, which are usually obtained from previous protein–ligand refinements of other complexes of the same crystal form of the target protein.

Molecular replacement is followed by rigid-body refinement of each model, where individual domains have been specified. After a preliminary short restrained refinement of each protein model, the best model to carry forward to subsequent

processing is then selected by analysis of the local electron density correlation in the regions (usually the flexible loops) where the models differ most. Taken together, these initial steps effectively handle the small changes in isomorphism and loop/side-chain movements that can occur when protein crystals are soaked with small molecule ligands. The molecular replacement/model selection step is followed by cycles of restrained refinement interspersed with automated placement of water molecules into $mF_o - DF_c$ electron density, except in one or more user-defined binding sites. The resulting $mF_o - DF_c$ difference Fourier in the binding site region (s) is then passed to AutoSolve for ligand identification and fitting.

2.3.5 Automatic Ligand Fitting and Refinement

AutoSolve is Astex's in-house developed software for electron-density analysis, interpretation and fitting, and has been one of the most important steps in reducing the time and effort required to generate protein–ligand structural data [45]. At the time AutoSolve was developed, existing ligand-fitting programs [85, 86] aimed to fit to electron density only, which meant that there was a high probability of producing unreasonable geometries and interaction modes with the protein. In addition, they relied first on identification of an electron density peak corresponding to a ligand, and hence were very sensitive to the density threshold selected for analysis. AutoSolve overcomes the first of these issues by exploiting the similarities between protein–ligand docking and electron-density fitting. Ligands are placed using a docking program (GOLD), whilst poses are scored using the fit to the electron density as well as interactions with the protein using a modified form of the Chemscore [58, 87] scoring function. The score for the final ligand pose is given by:

$$Score = S_{density} + 0.15 S_{HB} + 0.3 S_{metal} - 0.1 S_{clash} - 0.2 S_{int-clash} - 0.1 S_{torsion},$$

where the various terms correspond to scores for fit to the electron density, protein–ligand hydrogen bonding, metal interaction, steric clashes (between protein and ligand and within the ligand itself) and a ligand torsional term. It is evident that although electron-density fit is the prime determinant of binding mode, the additional interaction terms will serve to give chemically plausible conformations and binding modes. For example, for the case of a pseudo-symmetric fragment bound to trypsin (Fig. 4), AutoSolve correctly orientates the compound in order to satisfy the hydrogen bonding between the fragment's amine functionality and the protein, despite the symmetrical density. In addition, the “flipped” binding mode is penalized by the torsional score, which would place the methoxy group out of plane. An additional benefit of the use of interaction information is that AutoSolve can automatically select the most likely tautomeric or protonation state of a compound where relevant.

The score provided by the program also allows for the automatic assessment of the likely binder(s) from a cocktail, which removes some of the subjectivity associated with this process. Some examples illustrating this are shown in Figs. 5 and 6 (adapted

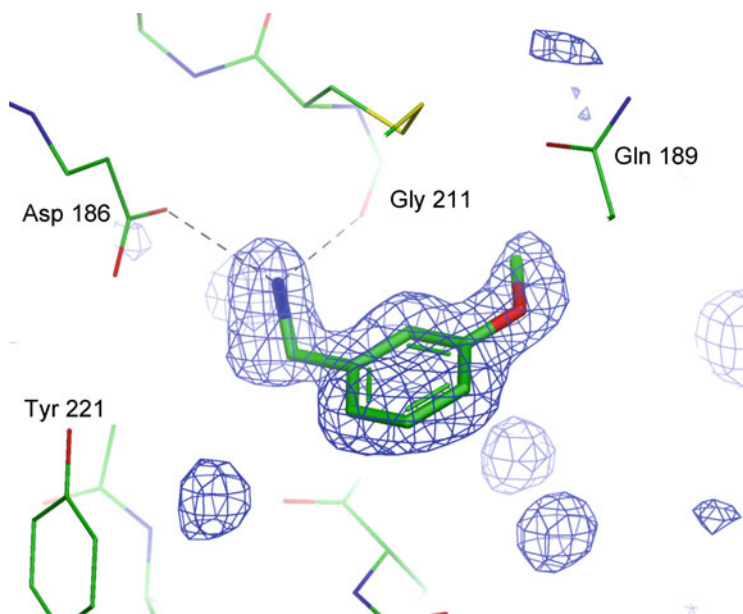


Fig. 4 AutoSolve solution for a fragment hit against trypsin. The initial $mF_o - DF_c$ difference Fourier contoured at 3σ is shown for the active site region. Despite the pseudosymmetric shape of the electron density, AutoSolve correctly orientates the ligand to satisfy the most likely hydrogen bonding pattern with the protein (denoted by *dashed lines*). Figure adapted from Mooij et al. [45]

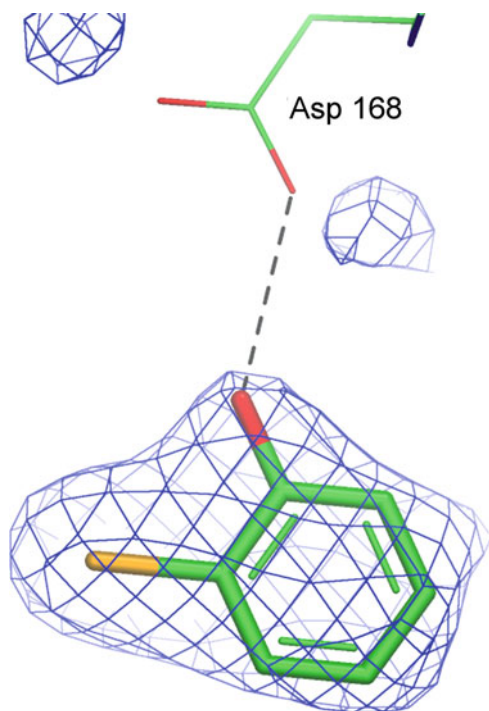


Fig. 5 Top ranked AutoSolve solution for a fragment-screening experiment against the kinase p38. The initial $mF_o - DF_c$ difference Fourier contoured at 3σ is shown for the active site region, and hydrogen bonds between protein and ligand are denoted by *dashed lines*. Figure adapted from Mooij et al. [45]

from [45]). In Fig. 5, AutoSolve correctly identifies the identity of a fragment bound to the kinase p38 as the top-scoring component of a cocktail of four. This is despite the resolution being lower (2.3 Å), and the density less distinct compared to the example given for trypsin above. Figure 6a shows the successful identification by AutoSolve of fragment hits in the less-common situation where more than one fragment binds simultaneously in the binding site. In this case, the program automatically identifies two compounds, which bind simultaneously from a cocktail of eight. Figure 6b shows the result from a confirmatory de-convolution experiment, in which the two compounds were subsequently soaked individually.

AutoSolve is normally run without the requirement to first search for peaks within the target active site: in other words it utilizes the electron density at all points within a cavity region (calculated from a user-defined “seed” atom), and without the necessity to define a particular threshold. This approach ensures that weakly bound ligands, perhaps with discontinuous electron density, will not be missed. Taken together, these approaches provide robust fitting to the electron density at a range of resolutions, and the ability of AutoSolve to reproduce known ligand-binding modes has been validated against a test set of 40 protein–ligand complexes from the RSCB Protein Data Bank (PDB) [45]. In 88% of cases, the top-ranked score reproduced the manually fitted binding mode to within 1.0 Å root mean square deviation (RMSD), and in 98% of cases a solution within 1.0 Å RMSD was found. In addition, this methodology exploits the full power of the genetic algorithm (GA) used by GOLD to place ligands within the active site, giving efficient sampling of conformational space and rapid fitting, even for cases of compounds with many torsional degrees of freedom.

In terms of a typical ligand-fitting run, initial ligand input is provided from the database as a set of SMILES strings, encoding the compound(s) for all relevant tautomers, protonation states and stereoisomers. These are converted to 3D geometries for ligand fitting using CORINA [88], which is used only to generate the

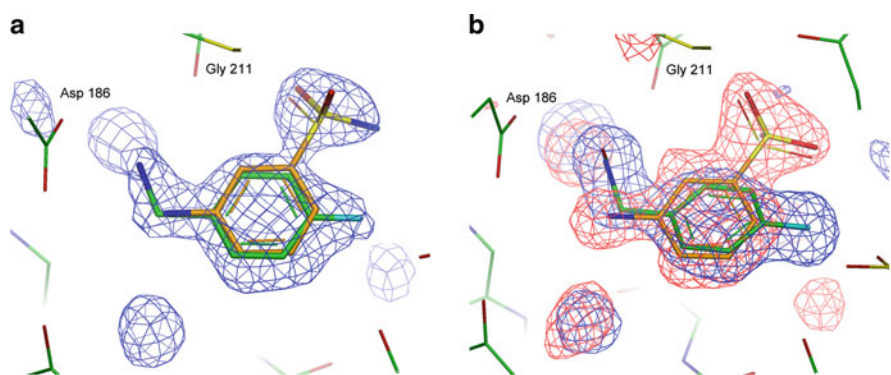


Fig. 6 (a) AutoSolve solutions for fragment-screening experiment against trypsin, with simultaneous binding of two compounds from a cocktail of eight. (b) Overlay of AutoSolve solutions and electron densities for subsequent deconvolution experiments in which compounds were individually soaked. Figure adapted from Mooij et al. [45]

connectivity, and then optimized using a CSD-derived force-field using the in-house developed software CSDOPT. Automated ligand fitting and inspection by a crystallographer (using the graphics program AstexViewer [89] or Coot [81, 82]) is then performed. This is followed by iterations of restrained refinement using TLS parameterization and automatically generated ligand restraints, further automated water-placing, and, where necessary, manual structure rebuilding. Finally the group ligand occupancies and B-factors are optimized, and standard quality-control checks on the final protein–ligand structure are performed before the structure is ready for release to project teams. The total process from initial integration of data, through AutoSolve and rebuilding, to the final fitted protein–ligand complex is driven entirely from a series of web-based interfaces, with the options for fully-automated running, or user intervention if required. All file storage and retrieval is performed by a company-wide Oracle database, which not only streamlines the whole process, and obviates the need for laborious file-management by the crystallographer, but also allows rapid tracking and querying of all information associated with the experiment.

2.3.6 Exploiting Structural Information

The full integration of structural information with other experimental data (e.g. cloning, purification, bioassay, chemical synthesis) is of key importance for the most effective and timely use of data. In addition to this valuable ability to query and cross-reference various aspects of each protein–ligand experiment, the seamless integration of all structural information within a database environment allows for the most efficient distribution of the resulting coordinates to project teams. Once identified as a “validated hit”, the protein–ligand structure becomes viewable to computational and medicinal chemists within a number of in-house chemo- and bioinformatic platforms and allows further cycles of ligand design. These tools allow a variety of queries to be performed, including searching for similar structures, for example, in terms of ligand substructure, protein sequence or protein–ligand interactions.

A key aspect of using the resulting structural information effectively has been the development of AstexViewer, which is a simple Java-based graphics program for viewing protein–ligand structures and electron density [89]. The design goal of AstexViewer was to produce a tool that could be used by scientists without a specialist background in crystallography or modelling. It is run as an applet in the Microsoft Internet Explorer web browser on a standard Windows PC, removing the need for specialist graphics workstations and unfamiliar operating systems, and is available to all members of the company on their desktop. It provides a simple interface that allows users to easily navigate the structure, measure molecular geometry, and permits a variety of protein and ligand representations and surfaces. It also allows easy display of electron density, and this has been important in encouraging modellers and medicinal chemists to look at the experimental maps in conjunction with fitted structures in their judgment of the structural information.

This ensures, for example, that undue time is not spent on design ideas for a part of the ligand that is disordered or mobile.

As discussed in the previous section, AstexViewer is used by crystallographers for visualization and rebuilding during the protein–ligand refinement process. It is also embedded within a number of other applications. For example, in order to maximize the impact of the structural information on the drug discovery process, we have developed a simple web-based interface that brings together the structural information available for a project [54]. We term these “project overlay pages”, and they provide a simple-to-use tool for use in project discussions and design. Project pages consist of a set of pre-superposed protein–ligand complexes, along with additional information such as bioassay results. The pages are typically built and maintained by the project modeller, and new structures can be added in a semi-automatic manner, with superposition being carried out relative to a previously defined reference. The pages themselves consist of a viewing pane, which contains AstexViewer, and a simple hierarchical tree of folders allowing structures be grouped according to certain criteria (Fig. 7). For example, a typical page might consist of folders for fragment hits (perhaps subdivided by different chemical classes), folders illustrating the hit-to-lead

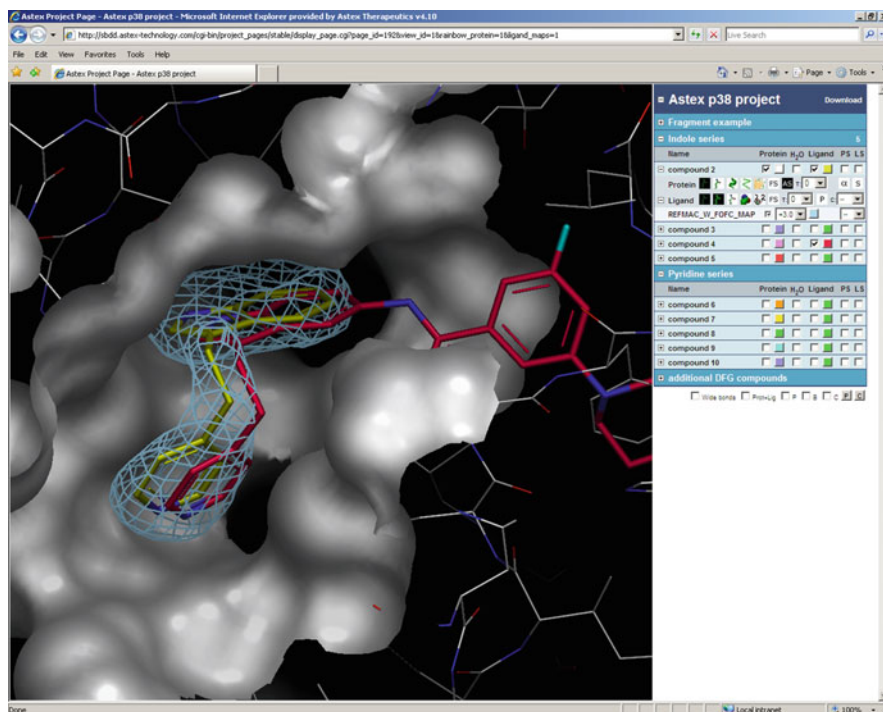


Fig. 7 Overlay page containing protein–ligand structures for the kinase p38. Structures are visualized within AstexViewer (*left-hand pane*), whilst the *right-hand pane* contains folders of display controls for sets of pre-superposed complexes

elaboration process, and folders for publically available structures from the PDB for comparison purposes. Each folder contains a set of Javascript controls, which drive functions such as loading protein and ligand, displaying molecular surfaces and determining the protein representation (colour, cartoon, sticks, spheres etc). They also have the ability to display experimental electron density and Superstar [90] maps if required.

3 Examples of Fragment Screening

3.1 Fragment–Protein Interactions

Over the last 10 years, we have carried out fragment screening campaigns against a wide range of targets including kinases, phosphatases, proteases and ATPases. Figure 8 shows examples of some hits we have observed during fragment-screening campaigns, and it can be seen that the approach is amenable to detection of binding driven by the full repertoire of non-covalent interactions. For example, Fig. 8 shows the binding mode for fragments forming neutral and non-classical CH \cdots O hydrogen bonds (Fig. 8a, CDK2) [43], lipophilic interactions (Fig. 8b, p38) [43] and charge–charge interactions (Fig. 8c, PTB1B [43]). It is notable that despite their weak potencies, all of the fragments exhibit clear electron densities indicative of unique binding modes. In addition, we have observed that even very weakly binding fragments can induce conformational changes: the PTB1B fragment hit shown in Fig. 8c induces a substantial movement of the enzyme’s “WPD” loop on binding. In Sect. 3.2 we present more detailed description for two case studies where we have successfully optimized fragment hits to potent inhibitors.

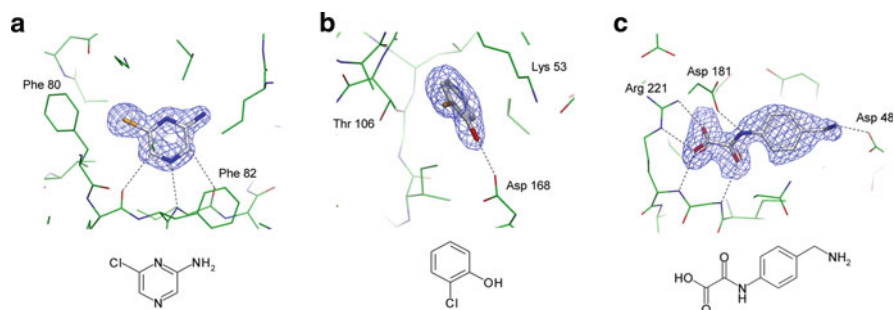


Fig. 8 Examples of fragment hits against selected targets, illustrating different aspects of molecular recognition. (a) CDK2 (neutral hydrogen bonding), (b) p38 (lipophilic interactions), (c) PTB1B (charge–charge interactions). Hydrogen bonds and electrostatic interactions are denoted by *dashed lines*, and the initial $mF_o - DF_c$ difference Fourier maps contoured at 3σ are shown for the ligands

3.2 Hits-to-Leads Case Studies

3.2.1 Development of CDK2 Inhibitor AT7519

The cyclin-dependent kinases (CDKs) are key regulators of cell-cycle progression and cellular proliferation. Aberrant control of the CDKs has been implicated in the molecular pathology of cancer, and it anticipated that their inhibition may provide an effective method for controlling tumour growth [91, 92].

We used X-ray crystallographic screening to identify fragments binding to CDK2 [93]. A library of approximately 500 fragments was soaked into crystals of CDK2 in cocktails of four, and more than 30 hits were observed to bind within the ATP cleft. Of these, indazole (**1**, Fig. 9), which exhibited a potency of 185 μM and an excellent ligand efficiency of 0.57, was selected for optimization using structure-based approaches. In order to increase the molecular weight of the compound, whilst still maintaining ligand efficiency, we initially sort to simplify the indazole to the pyrazole. The 3-substituted pyrazole, **2** (IC_{50} = 97 μM), forms an additional hydrogen-bonding interaction to the hinge region of the kinase, whilst adopting the same orientation as the starting fragment. This compound also places a phenyl ring near the backbone of Gln85, a region of the protein known to form energetically favourable interactions with aromatic groups, and a number of substitutions of this ring were investigated. The 4-fluoro analogue of **2** was then elaborated through addition of a second amide function at the pyrazole 4-position, allowing the formation of a water-mediated interaction with the backbone of Asp145 and giving a 100-fold increase in activity. Interestingly, this compound adopts a planar structure due to formation of an intramolecular hydrogen bond between the two amide functionalities, giving good shape-complementarity with the narrow ATP cleft. A small number of substitutions were explored from the second amide to probe

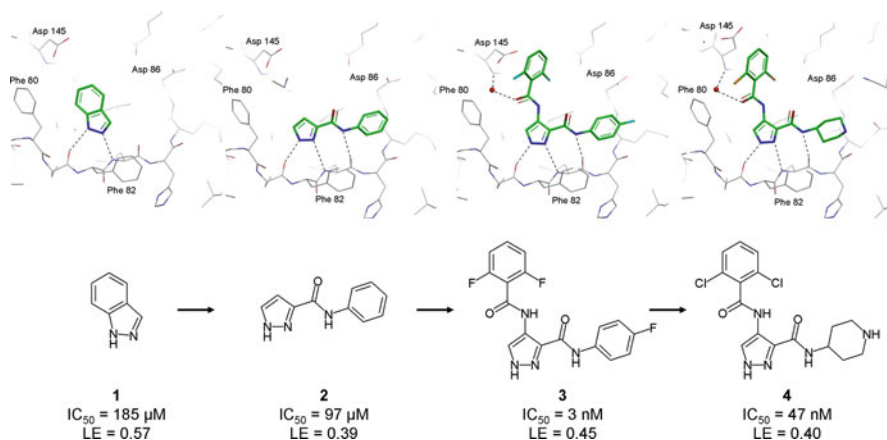


Fig. 9 Fragment evolution for the target CDK2 as described in the text. Key hydrogen bonding interactions with the protein are denoted by *dashed lines*

further the region near Asp145. In particular, the di-fluorophenyl, **3**, exhibited good kinase activity and ligand efficiency ($IC_{50} = 3$ nM). The crystal structure of the unsubstituted phenyl analogue had shown that the aromatic group binds with an energetically unfavourable twist relative to the amide, and diortho substitution was introduced to stabilize this conformation. Further optimization was then sought to improve cell-based potency and pharmacokinetic properties, and led to the replacement of the lipophilic 4-fluorophenyl group with the more polar piperidine. Substitution of the 2,5 difluoro by the dichloro finally led to AT7519, **4**, which exhibits good enzyme and cell-based potency (AT7519 $IC_{50} = 47$ nM; HCT116 $IC_{50} = 82$ nM), tumour regression in a number of xenograft models and is currently in clinical trials for the treatment of various cancers. The development of AT7519 is a successful example of the fragment-growth method, in which small changes are gradually introduced to increase potency. As is typical for this approach, the position and interactions of the initial fragment are maintained in the elaborated compound and, through careful use of structure-based design, ligand efficiency is maintained during the process.

3.2.2 Development of an Orally Bioavailable Inhibitor of Urokinase

Urokinase-type plasminogen activator (uPA) is a trypsin-like serine protease that catalyses the conversion of plasminogen to plasmin. Plasmin is associated with induction of cell-migration through degradation of the extracellular matrix, and uPA has been implicated in several disease states, including metastatic processes in cancer [94, 95]. The peptide binding site of uPA contains an acidic S_1 pocket, and a key challenge in the development of inhibitors against this target has been overcoming the low oral bioavailability associated with the highly basic arginine mimetics, which are typically required for potent binding.

A crystallographic screen was carried out against uPA, yielding more than 100 fragment hits [96]. From these, fragment **5** (Fig. 10), which is the known drug mexiletine, was selected for progression. Despite its weak binding ($IC_{50} > 1$ mM), it nevertheless exhibited a clear and unambiguous crystallographic binding mode, and as a known oral drug offered a promising starting point for further development.

Mexiletine binds in the S_1 pocket of uPA with its primary amine forming electrostatic and hydrogen-bonding interactions with the side-chain of Asp189 and the backbone carbonyls of Ser190 and Gly219. In addition, the ethanolamine spacer and the aromatic ring make several hydrophobic contacts with residues lining the pocket. The structure indicated that removal of the “angular” methyl group might be beneficial to binding by relief of unfavourable contacts, and previously published compounds suggested that substitution at the 4 position of the aromatic ring would also afford an increase in potency. The intermediate acid, **6**, exhibited an increase in potency to 40 μ M and, guided by virtual screening, a small number of aromatic amides were then prepared at this position. The crystal structure of **7** ($IC_{50} = 1.3$ μ M) revealed that it forms a number of aromatic contacts

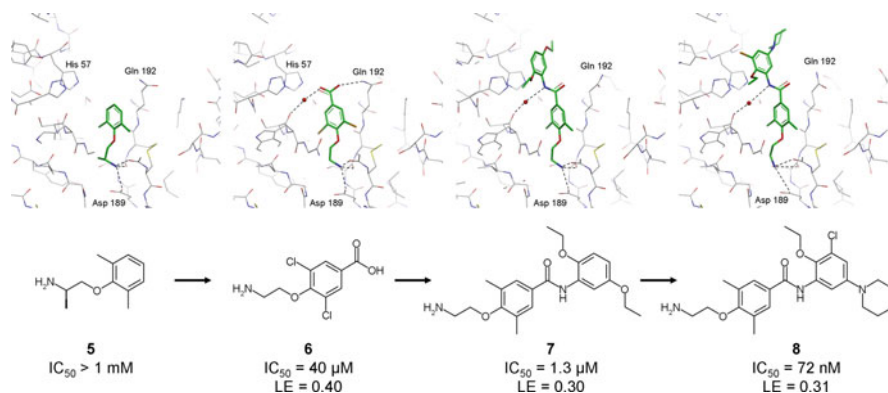


Fig. 10 Fragment evolution for the target urokinase as described in the text. Key hydrogen bonding interactions with the protein are denoted by *dashed lines*

between the newly added phenyl ring and the protein. In addition, a water-mediated hydrogen bond is observed between the amide nitrogen of **7** and the backbone carbonyl of Ser214. Further structure-guided optimization of the compounds (predominantly different space-filling decorations of the second aromatic ring) then led to lead compound **8**, which is a potent inhibitor of uPA ($IC_{50} = 72 \text{ nM}$). Of particular note is the relatively low pK_a of the basic amine, which is hypothesized to arise due to the effect of the *para*-amide functionality on the electron-withdrawing properties of the side-chain β -oxygen. As a result, the compound exhibits good pharmacokinetic properties, including high levels of oral bioavailability ($F_{\text{rat}} = 60\%$). With the exception of the highly related enzyme, trypsin, the compound also shows greater than 50-fold selectivity against a panel of proteases, and represents a promising lead-like compound with desirable pharmacokinetic properties.

4 Conclusions

The fragment-based approach is now firmly established as an important part of modern drug discovery. A range of biophysical and computational techniques are currently used for identifying fragment hits, and the combination of several methodologies in a typical screening cascade has shown to be a powerful approach for triaging possible binders and reducing false positives. The use of X-ray crystallography as a primary screen has a number of advantages, but traditionally was impractical due to low throughput, and in our view continues to be underexploited in drug discovery. We have described here how we approached this issue through compound cocktailing, streamlined data collection, automated data processing and ligand fitting. These techniques have allowed us to transform crystallography into a highly efficient technique that is suitable for rapid screening of fragment libraries,

and can provide timely structural information as a project progresses. Crystallography continues to form a central part of fragment screening at Astex, alongside full integration with other biophysical techniques. This approach has allowed us to apply the fragment-based method to the widest range of targets, with the most efficient combination of speed and sensitivity. Alongside the development of tools for the efficient dissemination and exploitation of crystallographic data by project teams, this has produced a highly efficient drug-discovery engine that has produced a pipeline of promising clinical candidates within a short time-frame.

Acknowledgments The authors gratefully acknowledge the support of many people at Astex Therapeutics who have contributed to aspects of the work presented here. We would also like to thank Dr Chris Murray and Dr Harren Jhoti for valuable suggestions regarding the content of this chapter.

References

1. Congreve M, Chessari G, Tisi D, Woodhead AJ (2008) *J Med Chem* 51:3661
2. Hajduk PJ, Greer J (2007) *Nat Rev Drug Discov* 6:211
3. Schulz MN, Hubbard RE (2009) *Curr Opin Pharmacol* 9:615
4. Erlanson DA (2011) Introduction to fragment-based drug discovery. *Top Curr Chem*. doi:10.1007/128_180
5. Rees DC, Congreve M, Murray CW, Carr R (2004) *Nat Rev Drug Discov* 3:660
6. Murray CW, Rees DC (2009) *Nat Chem* 1:187
7. Fink T, Bruggesser H, Reymond JL (2005) *Angew Chem Int Ed Engl* 44:1504
8. Fink T, Reymond JL (2007) *J Chem Inf Model* 47:342
9. Hann MM, Leach AR, Harper G (2001) *J Chem Inf Comput Sci* 41:856
10. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) *Adv Drug Deliv Rev* 46:3
11. Oprea TI, Davis AM, Teague SJ, Leeson PD (2001) *J Chem Inf Comput Sci* 41:1308
12. Hopkins AL, Groom CR, Alex A (2004) *Drug Discov Today* 9:430
13. Teague SJ, Davis AM, Leeson PD, Oprea T (1999) *Angew Chem Int Ed Engl* 38:3743
14. Erlanson DA, McDowell RS, O'Brien T (2004) *J Med Chem* 47:3463
15. Fejzo J, Lepre CA, Peng JW, Bemis GW, Ajay, Murcko MA, Moore JM (1999) *Chem Biol* 6:755
16. Fejzo J, Lepre C, Xie X (2003) *Curr Top Med Chem (Hilversum, Netherlands)* 3:81
17. Hajduk PJ, Gerfin T, Boehlen JM, Haberli M, Marek D, Fesik SW (1999) *J Med Chem* 42:2315
18. Hajduk PJ, Meadows RP, Fesik SW (1997) *Science* 278:497,499
19. Hajduk PJ (2006) *Mol Interv* 6:266
20. Siegal G, Hollander JG (2009) *Curr Top Med Chem* 9:1736
21. Kuglstatter A, Stahl M, Peters JU, Huber W, Stihle M, Schlatter D, Benz J, Ruf A, Roth D, Enderle T, Hennig M (2008) *Bioorg Med Chem Lett* 18:1304
22. Neumann T, Junker HD, Schmidt K, Sekul R (2007) *Curr Top Med Chem* 7:1630
23. Danielson UH (2009) *Curr Top Med Chem* 9:1725
24. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) *Science* 274:1531
25. Mattos C, Ringe D (1996) *Nat Biotechnol* 14:595
26. Fitzpatrick PA, Steinmetz AC, Ringe D, Klibanov AM (1993) *Proc Natl Acad Sci USA* 90:8653
27. English AC, Done SH, Caves LS, Groom CR, Hubbard RE (1999) *Proteins* 37:628
28. English AC, Groom CR, Hubbard RE (2001) *Protein Eng* 14:47

29. Verlinde CLMJ, Kim H, Bernstein BE, Mandé SC, Hol WGJ (1997) Antitrypanosomiasis drug development based on structures of glycolytic enzymes. In: Veerapandian P (ed) Structure-based drug design. Marcel Dekker, New York, p 365
30. Verlinde CLMJ, Rudenko G, Hol WG (1992) *J Comput Aided Mol Des* 6:131
31. Gill AL, Frederickson M, Cleasby A, Woodhead SJ, Carr MG, Woodhead AJ, Walker MT, Congreve MS, Devine LA, Tisi D, O'Reilly M, Seavers LC, Davis DJ, Curry J, Anthony R, Padova A, Murray CW, Carr RA, Jhota H (2005) *J Med Chem* 48:414
32. Murray CW, Carr MG, Callaghan O, Chessari G, Congreve M, Cowan S, Coyle JE, Downham R, Figueroa E, Frederickson M, Graham B, McMenamin R, O'Brien MA, Patel S, Phillips TR, Williams G, Woodhead AJ, Woolford AJ (2010) *J Med Chem* 53:5942
33. Murray CW, Callaghan O, Chessari G, Cleasby A, Congreve M, Frederickson M, Hartshorn MJ, McMenamin R, Patel S, Wallis N (2007) *J Med Chem* 50:1116
34. Woodhead AJ, Angove H, Carr MG, Chessari G, Congreve M, Coyle JE, Cosme J, Graham B, Day PJ, Downham R, Fazal L, Feltell R, Figueroa E, Frederickson M, Lewis J, McMenamin R, Murray CW, O'Brien MA, Parra L, Patel S, Phillips T, Rees DC, Rich S, Smith DM, Trewartha G, Vinkovic M, Williams B, Woolford AJ (2010) *J Med Chem* 53:5956
35. Nienaber VL, Richardson PL, Klighofer V, Bouska JJ, Giranda VL, Greer J (2000) *Nat Biotechnol* 18:1105
36. Antonysamy S, Hirst G, Park F, Sprengeler P, Stappenbeck F, Steensma R, Wilson M, Wong M (2009) *Bioorg Med Chem Lett* 19:279
37. Antonysamy SS, Aubol B, Blaney J, Browner MF, Giannetti AM, Harris SF, Hebert N, Hendle J, Hopkins S, Jefferson E, Kissinger C, Leveque V, Marciano D, McGee E, Najera I, Nolan B, Tomimoto M, Torres E, Wright T (2008) *Bioorg Med Chem Lett* 18:2990
38. Murray CW, Blundell TL (2010) *Curr Opin Struct Biol* 20:497
39. Wyss DF, Wang Y-S, Eaton HL, Strickland C, Voigt JH, Zhu Z, Stamford AW (2011) Combining NMR and X-ray crystallography in fragment-based drug discovery: discovery of highly potent and selective BACE-1 inhibitors. *Top Curr Chem*. doi:10.1007/128_183
40. Hennig M, Ruf A, Huber W (2011) Combining biophysical screening and X-ray crystallography for fragment-based drug discovery. *Top Curr Chem* doi:128_2011_225
41. Davies TG, van Montfort RLM, Williams G, Jhota H (2006) Pyramid: an integrated platform for fragment-based drug discovery. In: Jahnke W, Erlanson DA (eds) *Fragment-based approaches in drug discovery*. Wiley-VCH, Weinheim, p 193
42. Blaney J, Nienaber V, Burley SK (2006) Fragment-based lead discovery and optimization using X-ray crystallography, computational chemistry, and high-throughput organic synthesis. In: Jahnke W, Erlanson DA (eds) *Fragment-based approaches in drug discovery*. Wiley-VCH, Weinheim, p 215
43. Hartshorn MJ, Murray CW, Cleasby A, Frederickson M, Tickle IJ, Jhota H (2005) *J Med Chem* 48:403
44. Blundell TL, Abell C, Cleasby A, Hartshorn MJ, Tickle IJ, Parasini E, Jhota H (2002) High-throughput X-ray crystallography for drug discovery. In: Flower DR (ed) *Drug design: special publication*. Royal Society of Chemistry, Cambridge, UK, p 53
45. Mooij WT, Hartshorn MJ, Tickle IJ, Sharff AJ, Verdonk ML, Jhota H (2006) *ChemMedChem* 1:827
46. Bauman JD, Patel D, Arnold E (2011) Fragment screening and HIV therapeutics. *Top Curr Chem* doi:128_2011_232
47. Baurin N, Aboul-Ela F, Barril X, Davis B, Drysdale M, Dymock B, Finch H, Fromont C, Richardson C, Simmonite H, Hubbard RE (2004) *J Chem Inf Comput Sci* 44:2157
48. Siegal G, Ab E, Schultz J (2007) *Drug Discov Today* 12:1032
49. Chen IJ, Hubbard RE (2009) *J Comput Aided Mol Des* 23:603
50. Schuffenhauer A, Ruedisser S, Marzinzik AL, Jahnke W, Blommers M, Selzer P, Jacoby E (2005) *Curr Top Med Chem* 5:751
51. Bemis GW, Murcko MA (1996) *J Med Chem* 39:2887
52. Bemis GW, Murcko MA (1999) *J Med Chem* 42:5095

53. Congreve M, Carr R, Murray C, Jhoti H (2003) *Drug Discov Today* 8:876
54. Murray CW, Hartshorn MJ (2007) New applications for structure-based drug design. In: Mason JS (ed) *Computer-assisted drug design*. Elsevier, Amsterdam, p 775
55. Verdonk ML, Cole JC, Hartshorn M, Murray CW, Taylor RD (2003) *Proteins* 52:609
56. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) *J Mol Biol* 267:727
57. Jones G, Willett P, Glen RC (1995) *J Mol Biol* 245:43
58. Baxter CA, Murray CW, Clark DE, Westhead DR, Eldridge MD (1998) *Proteins* 33:367
59. Watson P, Verdonk ML, Hartshorn MJ (2003) *J Mol Graph Model* 22:71
60. Rognan D (2011) Fragment-based approaches and computer-aided drug discovery. *Top Curr Chem*. doi:10.1007/128_182
61. Davies TG, Woodhead SJ, Collins I (2009) *Curr Top Med Chem* 9:1705
62. Davies TG, Verdonk ML, Graham B, Saalau-Bethell S, Hamlett CC, McHardy T, Collins I, Garrett MD, Workman P, Woodhead SJ, Jhoti H, Barford D (2007) *J Mol Biol* 367:882
63. Saxty G, Woodhead SJ, Berdini V, Davies TG, Verdonk ML, Wyatt PG, Boyle RG, Barford D, Downham R, Garrett MD, Carr RA (2007) *J Med Chem* 50:2293
64. Verlinde CL, Fan E, Shibata S, Zhang Z, Sun Z, Deng W, Ross J, Kim J, Xiao L, Arakaki TL, Bosch J, Caruthers JM, Larson ET, Letrong I, Napuli A, Kelly A, Mueller N, Zucker F, Van Voorhis WC, Merritt WC, Hol WG (2009) *Curr Top Med Chem* 9:1678
65. Blakely MP, Cianci M, Helliwell JR, Rizkallah PJ (2004) *Chem Soc Rev* 33:548
66. Muchmore SW, Olson J, Jones R, Pan J, Blum M, Greer J, Merrick SM, Magdalinos P, Nienaber VL (2000) *Struct Fold Des* 8:R243
67. Sharff AJ (2004) *The Rigaku Journal* 20:10
68. van Montfort RL, Congreve M, Tisi D, Carr R, Jhoti H (2003) *Nature* 423:773
69. Beteva A, Cipriani F, Cusack S, Delageniere S, Gabadinho J, Gordon EJ, Guijarro M, Hall DR, Larsen S, Launer L, Lavault CB, Leonard GA, Mairs T, McCarthy A, McCarthy J, Meyer J, Mitchell E, Monaco S, Nurizzo D, Pernot P, Pieritz R, Ravelli RG, Rey V, Shepard W, Spruce D, Stuart DI, Svensson O, Theveneau P, Thibault X, Turkenburg J, Walsh M, McSweeney SM (2006) *Acta Crystallogr D Biol Crystallogr* 62:1162
70. McPhillips TM, McPhillips SE, Chiu HJ, Cohen AE, Deacon AM, Ellis PJ, Garman E, Gonzalez A, Sauter NK, Phizackerley RP, Soltis SM, Kuhn P (2002) *J Synchrotron Radiat* 9:401
71. Gabadinho J, Beteva A, Guijarro M, Rey-Bakaikoa V, Spruce D, Bowler MW, Brockhauser S, Flot D, Gordon EJ, Hall DR, Lavault B, McCarthy AA, McCarthy J, Mitchell E, Monaco S, Mueller-Dieckmann C, Nurizzo D, Ravelli RB, Thibault X, Walsh MA, Leonard GA, McSweeney SM (2010) *J Synchrotron Radiat* 17:700
72. Pflugrath JW (1999) *Acta Crystallogr D* 55(Pt 10):1718
73. Kabsch W (2010) *Acta Crystallogr D Biol Crystallogr* 66:125
74. Leslie AGW (1992) *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, No. 26
75. Leslie AGW, Brick P, Wonacott A (2004) *Daresbury Lab Inf Quart Protein Crystallography* 18:33
76. Collaborative Computational Project N4 (1994) *Acta Crystallogr D Biol Crystallogr* 50:760
77. Wlodek S, Skillman AG, Nicholls A (2006) *Acta Crystallogr D Biol Crystallogr* 62:741
78. Terwilliger TC, Klei H, Adams PD, Moriarty NW, Cohn JD (2006) *Acta Crystallogr D Biol Crystallogr* 62:915
79. Terwilliger TC, Adams PD, Moriarty NW, Cohn JD (2007) *Acta Crystallogr D Biol Crystallogr* 63:101
80. Evrard GX, Langer GG, Perrakis A, Lamzin VS (2007) *Acta Crystallogr D Biol Crystallogr* 63:108
81. Emsley P, Cowtan K (2004) *Acta Crystallogr D Biol Crystallogr* 60:2126
82. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) *Acta Crystallogr D Biol Crystallogr* 66:486

83. Navaza J (2004) *Acta Crystallogr A* 50:157
84. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) *J Appl Crystallogr* 40:658
85. Oldfield TJ (2001) *Acta Crystallogr D Biol Crystallogr* 57:696
86. Zwart PH, Langer GG, Lamzin VS (2004) *Acta Crystallogr D Biol Crystallogr* 60:2230
87. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP (1997) *J Comput Aid Mol Des* 11:425
88. Gasteiger J, Rudolph C, Sadowski J (2004) *Tetrahedron Comput Methodol* 3:537
89. Hartshorn MJ (2002) *J Comput Aided Mol Des* 16:871
90. Verdonk ML, Cole JC, Watson P, Gillet V, Willett P (2001) *J Mol Biol* 307:841
91. Malumbres M, Carnero A (2003) *Prog Cell Cycle Res* 5:5
92. Sausville EA, Zaharevitz D, Gussio R, Meijer L, Louarn-Leost M, Kunick C, Schultz R, Lahusen T, Headlee D, Stinson S, Arbuck SG, Senderowicz A (1999) *Pharmacol Ther* 82:285
93. Wyatt PG, Woodhead AJ, Berdini V, Boulstridge JA, Carr MG, Cross DM, Davis DJ, Devine LA, Early TR, Feltell RE, Lewis EJ, McMenamin RL, Navarro EF, O'Brien MA, O'Reilly M, Reule M, Saxty G, Seavers LC, Smith DM, Squires MS, Trewartha G, Walker MT, Woolford AJ (2008) *J Med Chem* 51:4986
94. Schweinitz A, Steinmetzer T, Banke IJ, Arlt MJ, Sturzebecher A, Schuster O, Geissler A, Giersiefen H, Zeslowska E, Jacob U, Kruger A, Sturzebecher J (2004) *J Biol Chem* 279:33613
95. Almholt K, Lund LR, Rygaard J, Nielsen BS, Dano K, Romer J, Johnsen M (2005) *Int J Cancer* 113:525
96. Frederickson M, Callaghan O, Chessari G, Congreve M, Cowan SR, Matthews JE, McMenamin R, Smith DM, Vinkovic M, Wallis NG (2008) *J Med Chem* 51:183