Thomas G. Davies

Marko Hyvönen   *Editors*

Fragment-Based Drug Discovery and X-Ray Crystallography

Springer

317

Topics in Current Chemistry

Topics in Current Chemistry
Recently Published and Forthcoming Volumes

Fragment-Based Drug Discovery and X-Ray Crystallography

Volume Editors: Thomas G. Davies · Marko Hyvönen

With Contributions by

E. Arnold · J.D. Bauman · P. Brough · T.G. Davies · H.L. Eaton ·
D.A. Erlanson · S. Greive · M. Hennig · W. Huber · R.E. Hubbard ·
M. Hyvönen · M. Marsh · A. Massey · D. Patel · A. Ruf · D. Rognan ·
S. Roughley · T. Sharpe · A.W. Stamford · C. Strickland · I.J. Tickle ·
E. Valkov · J.H. Voigt · Y.-S. Wang · L. Wright · D.F. Wyss · Z. Zhu

Springer

Editors
Dr. Thomas G. Davies
Astex Pharmaceuticals
436 Cambridge Science Park
Milton Road
Cambridge CB4 0QA
UK
tom.davies@astx.com

Dr. Marko Hyvönen
Department of Biochemistry
University of Cambridge
80 Tennis Court Road
Cambridge CB2 1GA
UK
marko@cryst.bioc.cam.ac.uk

Printed on acid-free paper

Topics in Current Chemistry
Also Available Electronically

Topics in Current Chemistry is included in Springer's eBook package *Chemistry and Materials Science*. If a library does not opt for the whole package the book series may be bought on a subscription basis. Also, all back volumes are available electronically.

For all customers with a print standing order we offer free access to the electronic volumes of the series published in the current year.

If you do not have access, you can still view the table of contents of each volume and the abstract of each article by going to the SpringerLink homepage, clicking on "Chemistry and Materials Science," under Subject Collection, then "Book Series," under Content Type and finally by selecting *Topics in Current Chemistry*.

You will find information about the

– Editorial Board
– Aims and Scope
– Instructions for Authors
– Sample Contribution

at springer.com using the search function by typing in *Topics in Current Chemistry*.

Color figures are published in full color in the electronic version on SpringerLink.

Aims and Scope

The series *Topics in Current Chemistry* presents critical reviews of the present and future trends in modern chemical research. The scope includes all areas of chemical science, including the interfaces with related disciplines such as biology, medicine, and materials science.

The objective of each thematic volume is to give the non-specialist reader, whether at the university or in industry, a comprehensive overview of an area where new insights of interest to a larger scientific audience are emerging.

Thus each review within the volume critically surveys one aspect of that topic and places it within the context of the volume as a whole. The most significant developments of the last 5–10 years are presented, using selected examples to illustrate the principles discussed. A description of the laboratory procedures involved is often useful to the reader. The coverage is not exhaustive in data, but rather conceptual, concentrating on the methodological thinking that will allow the non-specialist reader to understand the information presented.

Discussion of possible future research directions in the area is welcome.

Review articles for the individual volumes are invited by the volume editors.

In references *Topics in Current Chemistry* is abbreviated *Top Curr Chem* and is cited as a journal.

Impact Factor 2010: 2.067; Section "Chemistry, Multidisciplinary": Rank 44 of 144

Preface

The fragment-based approach has emerged in the last decade as a highly promising component of modern drug discovery. Despite its relatively short history, it has been the subject of many research articles, reviews and books, and is responsible for several compounds currently in clinical development. Its contribution is increasingly recognized by the medicinal chemistry community, and it now forms an important part of drug discovery efforts within the pharmaceutical industry.

Despite the exponential growth of interest in this field, fragment-based drug discovery (FBDD) represents a significant paradigm shift for drug discoverers, both philosophically, and in terms of methodology and work-flow. In particular, it has required a shift away from relatively potent, drug-like hits, readily identified by enzymatic high-throughput assays, to the more challenging detection of very weakly (but efficiently) binding compounds. As such, the development and application of robust and sensitive biophysical techniques to detect and characterise the binding of simple, low molecular compounds has been a key part of enabling this approach. X-ray crystallography was one of the earliest techniques demonstrated to be capable of detecting the binding of fragments, and its additional ability to provide precise three-dimensional detail on their binding modes, and hence guide their subsequent elaboration has led to it playing a central role in this approach.

In this volume we bring together chapters by a number of practitioners in the field, drawn from both the pharmaceutical industry and academia. Our aim has been to highlight the important roles that X-ray crystallography plays in the fragment-approach: as a sensitive technique for primary screening, its use in combination with other biophysical techniques to allow robust hit validation, and its importance in providing structural information to guide progression from hit to clinical candidate.

In the first chapter, Erlanson from Carmot Therapeutics provides an introduction to the FBDD field as a whole, highlighting some of the advantages of fragments and their means of detection, and giving examples of fragment-derived compounds which have already reached the clinic. Davies and Tickle from Astex Therapeutics then provide a review of the use of X-ray crystallography for fragment screening,

and describe some of the computational developments developed at Astex that have allowed the rapid generation of protein-ligand structural data required for this approach.

In chapter 3, Roughley and colleagues from Vernalis present the first of a number of personal case studies of FBDD – in this case, the application of the fragment-approach to the development of Hsp90 inhibitors, with emphasis on the role of *in silico* screening, and its interplay with experimental structural information. This is followed by a chapter from Wyss *et al* (Merck), who describe their work on the fragment-based development of BACE inhibitors and how complementary information from both NMR and X-ray crystallography were combined to successfully prosecute a drug discovery campaign against this important target. Continuing on the theme of combining biophysical techniques, Hennig and colleagues then describe the approach to FBDD taken at Roche, and in particular how Surface Plasmon Resonance and structural information are used together in an integrated approach.

Fragment-based approaches are increasingly being applied to challenging therapeutic targets, and in particular those for which conventional drug discovery methods have failed. In chapter 6, Valkov and colleagues (University of Cambridge) provide a review of small molecule inhibition of protein-protein interactions, and the application of fragment-based methods against this class of target. Bauman *et al* (Rutgers) then describe the use of X-ray crystallographic fragment screening to identify novel hits against HIV targets, and highlight the growing trend for academic-based FBDD. Indeed, the close association of biophysical and structural techniques, combined with the manageable size of screening libraries make fragment-based methods increasingly appealing and accessible to academic laboratories in addition to those in the pharmaceutical industry.

In the concluding chapter, and in a departure from the predominantly experimental methods discussed above, Rognan (University of Strasbourg) provides a computational perspective on the fragment-based approach, and discusses the application and development of *in silico* approaches which are increasingly being applied in this area.

We hope this book will provide a useful introduction to some of the key concepts and techniques in fragment-based drug discovery, highlighting the diverse set of targets it is applied to, as well as emphasizing the importance of structural information in this field. The application of X-ray crystallography to structure-based drug discovery is now a mature discipline, but one whose potential has sometimes been under-exploited. In driving various aspects of the fragment-based approach, it clearly plays a central role.

Cambridge, January 2012   Thomas Davies
                                                                       Marko Hyvönen

Contents

Introduction to Fragment-Based Drug Discovery

Daniel A. Erlanson

Abstract Fragment-based drug discovery (FBDD) has emerged in the past decade as a powerful tool for discovering drug leads. The approach first identifies starting points: very small molecules (fragments) that are about half the size of typical drugs. These fragments are then expanded or linked together to generate drug leads. Although the origins of the technique date back some 30 years, it was only in the mid-1990s that experimental techniques became sufficiently sensitive and rapid for the concept to be become practical. Since that time, the field has exploded: FBDD has played a role in discovery of at least 18 drugs that have entered the clinic, and practitioners of FBDD can be found throughout the world in both academia and industry. Literally dozens of reviews have been published on various aspects of FBDD or on the field as a whole, as have three books (Jahnke and Erlanson, Fragment-based approaches in drug discovery, 2006; Zartler and Shapiro, Fragment-based drug discovery: a practical approach, 2008; Kuo, Fragment based drug design: tools, practical approaches, and examples, 2011). However, this chapter will assume that the reader is approaching the field with little prior knowledge. It will introduce some of the key concepts, set the stage for the chapters to follow, and demonstrate how X-ray crystallography plays a central role in fragment identification and advancement.

Keywords Fragment-based drug discovery · Fragment-based lead discovery · Fragment-based screening · Kinase · Nuclear magnetic resonance spectroscopy · Structure-based drug design · X-ray crystallography

Contents

D.A. Erlanson
Carmot Therapeutics, Inc., 409 Illinois Street, San Francisco, CA 94158, USA
e-mail: derlanson@carmot.us

1 Why Fragments?

> Space is big. You just won't believe how vastly, hugely, mind-bogglingly big it is. I mean, you may think it's a long way down the road to the chemist's, but that's just peanuts to space.
>
> Douglas Adams

In this famous quote from The Hitchhiker's Guide to the Galaxy [1], Adams was referring to physical space, but he could just as accurately have been writing about chemical space. There have been several attempts to estimate the number of possible drug-like molecules, one of the most widely quoted being a footnote in a review on structure-based drug design which proposed the number 10^{63} [2]. Although this may be off by orders of magnitude in either direction, clearly the numbers in question are barely comprehensible, yet alone achievable.

Such numbers notwithstanding, one of the dominant methods of drug discovery in recent decades has been high-throughput screening (HTS), in which tens of thousands to millions of compounds are collected and screened against a target of interest. If chemical space was of a manageable size, one could be certain that a screen of, say, a million compounds would cover a good swath of it. But since chemical space is so vast, any collection of molecules assembled covers an insignificant portion of diversity space. A few years ago, the worldwide collection of isolated small molecules was estimated to be around 100 million [3], so even screening all of them would not begin to sample chemical space.

About half of all HTS campaigns fail, often because there are no good small-molecule starting points in the collection [4]. Failure is more common for newer targets or classes of targets for which there may not be many historical compounds, such as protein–protein interactions [5, 6]. Moreover, HTS is expensive: purchasing, maintaining, and screening a set of hundreds of thousands or millions of compounds can tax the resources of smaller companies and academic centers.

The fact that HTS does not always result in viable hits, coupled with the recognition of the vastness of chemical space, led to the concept of fragment-based drug discovery (FBDD). The basic premise is that, instead of searching huge collections of drug-sized molecules, one could search smaller collections of smaller molecules (or fragments), and then either grow a fragment or combine two

fragments to achieve the kind of potency one expects from HTS. From a practical standpoint, the smaller the molecule, the fewer the possibilities, so it is possible to search chemical space for fragments more efficiently. For example, computational enumeration of all possible molecules containing up to 11 carbon, nitrogen, oxygen, and fluorine atoms yields just over 100 million [7].

The late William Jencks of Brandeis University first proposed the theory behind FBDD 30 years ago [8]:

> It can be useful to describe the Gibbs free energy changes for the binding to a protein of a molecule, A–B, and of its component parts, A and B, in terms of the "intrinsic binding energies" of A and B, ΔG_A^i and ΔG_B^i, and a "connection Gibbs energy," ΔG^s that is derived largely from changes in translational and rotational entropy.

These ideas can be represented graphically as shown in Fig. 1. The top panel is a simplistic representation of a high-throughput screen: multiple compounds are screened against a target (most likely a protein) to identify a hit that binds – albeit imperfectly. This is subsequently optimized through medicinal chemistry. The middle panel represents the fragment linking as proposed by Jencks: two fragments that bind in nearby sites are chemically linked together. Just as with HTS, subsequent medicinal chemistry is necessary to further improve the molecule.

The linking concept was reduced to practice in a high profile *Science* paper from Abbott Laboratories in 1996 [9]. Since then, however, many groups have found that linking is much more challenging than might be expected (see below). Part of the
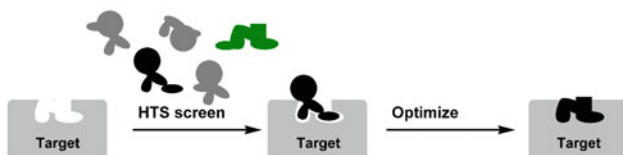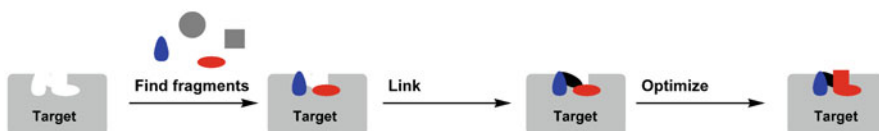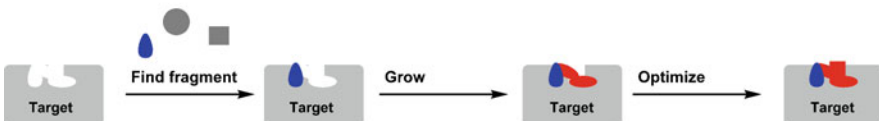
Fig. 1 Comparison of high-throughput screening (HTS, *top*) with fragment linking (*middle*) and fragment growing (*bottom*)

difficulty is that chemical bonds have strict length and geometric requirements, so if the two fragments are not perfectly positioned much of the potency gain expected will be lost due to strain in the linker [10, 11]. Therefore, a frequent alternative to fragment linking is fragment growing, as shown in the bottom panel of Fig. 1. In this approach, a single fragment is progressively grown to make further interactions with the protein.

It can be useful pedagogically to describe projects in terms of "linking" or "growing," but in the real world this distinction may be less clear. For example, part of one fragment may be merged with another in a process sometimes called "fragment merging" [11]. Medicinal chemists are adept at borrowing a portion from one chemical series and appending it onto a different chemical series to generate novel molecules; the same practices can be applied in FBDD. The increasing use of fragment-based approaches throughout all phases of a project has caused some people to refer to "fragment-assisted drug discovery," in which information from fragments is applied to more traditional drug discovery programs [12].

In addition to covering chemical space more efficiently, the hit rate for screening smaller compounds should in theory be higher than for larger compounds. This is because as molecules grow larger they grow more complex, and each additional moiety has an increasing probability of interfering with binding. This was demonstrated computationally a decade ago [13], but can be understood intuitively by examining the top panel of Fig. 1. The HTS molecule in the upper right-hand corner is perfectly complementary to the protein binding site save for a small appendage, which would prevent it from binding. In contrast, a fragment with high complementarity to the target will bind very efficiently (see Sect. 3.2), which will provide more scope for size increases during lead optimization.

So, the advantages of fragment screening are that it should allow one to explore chemical space more efficiently and achieve a higher hit rate than HTS. However, it took 15 years after Jencks' publication before the technique really demonstrated its utility, and several more years before it became widespread. This is because of two challenges: finding fragments, and figuring out what to do with them. Biophysical techniques such as X-ray crystallography play a key role in addressing both of these challenges today, but the high-throughput methods that researchers take for granted are a relatively recent innovation. Section 2 will discuss some of the challenges in finding fragments, and how to overcome them. This will be followed by a brief section on how to evaluate fragments. Finally, fragment-based programs that have produced clinical candidates will be discussed, with special attention given to the role crystallography played.

2 Finding Fragments

Traditionally (i.e., more than a couple of decades ago) active molecules were often found simply by testing them in a biological assay, often in cells or even in animals. As our understanding of biology and our ability to isolate proteins improved, it

became possible to take a more reductionist approach and test molecules against isolated enzymes or proteins in functional assays; this has become standard practice in HTS. In principle it should be possible to do this with fragments, but several pitfalls can arise: solubility and reactivity of molecules, and aggregation.

2.1 Down the Rabbit Hole: Pitfalls When Dealing with Low-Affinity Binders

2.1.1 Solubility

The first challenge when trying to find fragments is solubility: many fragments bind to proteins with dissociation constants of 1 mM or even higher, but many organic molecules are not soluble at these concentrations. Thus, it is imperative to check solubility of fragments in the appropriate biological buffer before screening. Though the need for this precaution may seem obvious, it is often overlooked, particularly when researchers are setting up fragment screening for the first time.

2.1.2 Reactive Molecules

Reactive molecules are another concern – not just the fragments themselves, but low level impurities. For example, if a compound in a high-throughput screen conducted at 1 μM concentration is contaminated with 1% of a reactive intermediate, this will be present at a mere 10 nM concentration and may not be problematic. However, if the same molecule is tested in a fragment screen at 1 mM concentration, the reactive intermediate will be present at 10 μM (probably a higher concentration than the target protein itself), and could thus cause a false positive signal by reacting with and inactivating the protein.

Many types of reactive molecules are well known to medicinal chemists: acyl halides, aldehydes, aliphatic esters, aliphatic ketones, alkyl halides, anhydrides, alpha-halocarbonyl compounds, aziridines, 1,2-dicarbonyl compounds, epoxides, halopyrimidines, heteroatom–heteroatom single bonds, imines, Michael acceptors and β-heterosubstituted carbonyl compounds, perhalo ketones, phosphonate esters, thioesters, sulfonate esters, and sulfonyl halides, to name a few [14]. This is not to say that these functionalities are not useful – some even appear in approved drugs – but all of these can react covalently with proteins, and thus should be regarded with suspicion. However, molecules can react covalently with proteins even if they do not contain functionalities that raise alarm. Jonathan Baell has referred to these as pan assay interference compounds, or PAINS, and has published a list of moieties to watch out for, as well as strategies to detect them [15, 16].

Even less obvious are molecules that may not react with proteins directly but that act as oxidizers, for example by generating hydrogen peroxide, which can in turn

Fig. 2 Examples of molecules demonstrated to generate hydrogen peroxide under standard biochemical assay conditions (**1–3**) and a similar molecule (**4**) reported without testing for redox activity
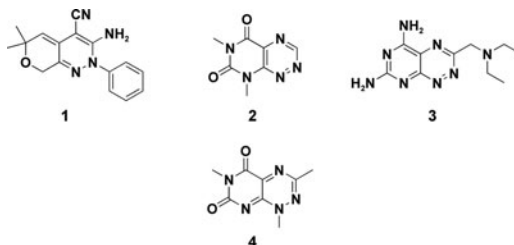
inactivate proteins. Examples of these types of compounds are shown in Fig. 2: all of them are small, fragment-like molecules. Molecule **1** and molecules **2** and **3** were all reported to inhibit PTP1B by generating hydrogen peroxide in the presence of buffers containing reducing agents, a common and generally wise practice to keep proteins in the reduced state [17, 18]. The problem is that compounds **1–3** can be reduced and subsequently reoxidized by ambient oxygen, generating hydrogen peroxide in the process. Unfortunately, this type of mechanism can be challenging to track down. For example, when compound **4** was reported as a novel protein–protein interaction inhibitor [19], no attempt was made to rule out hydrogen peroxide generation despite the close similarity between compounds **2** and **4**, and the fact that the buffers used contained reducing agents. In fact, compound **4** and several analogs do generate hydrogen peroxide, which is likely to be responsible for the activity observed [20–22]. As new chemical classes of molecules are added to screening collections it is essential to be vigilant for such problems.

2.1.3 Aggregators

Solubility and reactive molecules are both serious problems, but an even more insidious pitfall is the phenomenon of aggregation. Many small molecules can form aggregates in aqueous solution at relatively high concentration, and these aggregates can nonspecifically inhibit proteins and interfere with biochemical assays [23]. The effect appears to be concentration dependent. Thus, aggregation becomes increasingly likely as higher concentrations are needed to detect low affinity binders. Sometimes molecules that aggregate are long, extended, planar "ugly" molecules, but even small fragment-sized molecules and approved drugs can aggregate. Figure 3 shows an example of two fragment-sized drugs (**5** and **6**) that fall into this category [24].

The degree to which this is a problem can be appreciated by a screen of 70,563 molecules to discover inhibitors of the enzyme AmpC β-lactamase [25]. Of the 1,274 hits, 1,213 turned out to be aggregators – more than 95%! Even worse, these compounds often display structure–activity relationships (SAR), and the effect can persist even at fairly low concentrations. Recently, a series of cruzain inhibitors with IC_{50} values as low as 200 nM were reported, but follow-up studies determined

Fig. 3 Two approved drugs that can form aggregates at high concentrations

5: benzyl benzoate
16 heavy atoms
MW 212

6: mefenamic acid
18 heavy atoms
MW 241
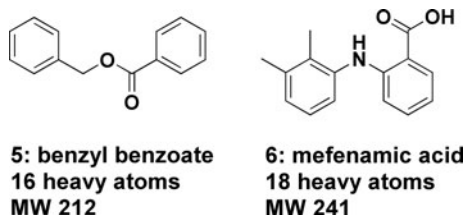
that they were aggregators, and that the medicinal chemistry effort had inadvertently optimized for aggregation [26].

It is hard to understate how serious this problem can be. Most large pharmaceutical companies are now aware of it and take steps to prevent it, but academic laboratories and smaller companies may not be so stringent. Fortunately, it is usually possible to prevent aggregate formation simply by adding small amounts of nonionic detergent to the assay buffer [27]. Other steps include increasing the protein concentration; this should usually not affect the measured IC_{50} values. Centrifuging samples can remove aggregates, and flow cytometry or dynamic light scattering can also reveal the presence of aggregators. Finally, unusually steep dose–response curves can be a tell-tale sign of aggregators [28].

Perhaps one reason that fragment-based approaches were slow to take off is because of all these problems. Aggregation in particular was not really appreciated until the early part of this century. In the absence of a clear understanding of some of these pitfalls, medicinal chemists who tried to optimize lead series starting from weak hits could quickly and unknowingly find themselves optimizing for aggregation. The resulting molecules would be unlikely to show cellular activity and ultimately reach a potency limit in the high nanomolar or low micromolar range. One or two programs such as this would be enough to dissuade chemists from pursuing low affinity hits. Happily, we now have sufficiently advanced tools, and an improved understanding of what can go wrong, to pursue fragments successfully.

2.2 Methods for Finding Fragments

Given the pitfalls described in the preceding section, it is not surprising that biophysical methods have dominated FBDD, and in fact the increasing sensitivity and throughput of biophysical techniques are in large part responsible for the success of the approach. However, non-biophysical methods are also coming into their own. In this section, methods for finding fragments are considered briefly; each has been reviewed in more detail elsewhere, and references to these reviews are provided.

2.2.1 Nuclear Magnetic Resonance

Although this book focuses on X-ray crystallography, it is appropriate to begin a discussion of fragment-finding approaches with nuclear magnetic resonance (NMR) because "SAR by NMR" was the technique that robustly demonstrated that fragment-based approaches were practical [9]. In this approach, two-dimensional NMR spectra are acquired of the protein in the presence and absence of fragments. Changes in protein chemical shifts in the presence of a fragment indicate binding, and if the chemical shifts have been assigned to specific protein residues the location of binding can be determined. This is an example of "protein-detected" NMR, which relies on changes in the NMR signal of the protein.

SAR by NMR is a powerful approach and has resulted in clinical compounds (see for example the Sects. 5.3.1 and 5.3.2 on ABT-518 and ABT-263, respectively). However, because it relies on changes in protein chemical shifts, it is limited to relatively small proteins (around 30–40 kD). Moreover, the approach requires large quantities of protein; the original paper suggested more than 200 mg [9], although miniaturization has decreased this requirement somewhat. As a result, several research groups have developed "ligand-detected" NMR techniques, in which changes in the NMR properties of the fragments, rather than the protein target, are detected. There are a number of techniques in use [29]: one of the most popular is saturation transfer difference (STD), which relies on the differences in relaxation between small molecules and large macromolecules [30]. This requires considerably less protein than SAR by NMR and is amenable to larger proteins, although it does not provide information on the site of binding.

An interesting ligand-detected approach that relies on interligand nuclear Overhauser effects (SAR by ILOE) detects two ligands that bind in close proximity to each other on the protein surface, facilitating linking [31, 32], although one needs to be cautious to avoid false positives due to aggregation of compounds [33]. Another interesting ligand-detected method is called target-immobilized NMR screening, or TINS, which relies on ligands binding to a protein that has been immobilized onto resin [34]. Appealingly, this method seems to be applicable to membrane proteins, which are generally challenging in NMR, as recently demonstrated by researchers from ZoBio [35].

Abbott Laboratories was the first company to report NMR for fragment screening, but the technique is now widely used, particularly ligand-detected methods. Companies known to use NMR include Abbott Laboratories, Astex Therapeutics, Evotec, Schering-Plough (now Merck), and Vernalis. NMR approaches have been extensively reviewed [29, 36–45], and are also covered in more depth by Wyss and coworkers [46].

2.2.2 X-Ray Crystallography

X-ray crystallography is covered in detail by Bauman et al. [47], Davies and Tickle [48], and Hennig et al. [49] and will thus be only briefly discussed here.

Crystallography and protein-detected NMR are unique in providing detailed empirical information on how ligands bind to proteins. Unlike NMR, crystallography can be applied to large proteins and can provide very high-resolution data. Fragment-based drug discovery owes much to the rapid increase in throughput of crystallography over the past 15 years. Most companies using FBDD now use X-ray crystallography. Some companies use crystallography as their primary screening technique, and several only pursue fragments that can be characterized crystallographically. Contract research organizations such as Emerald Biostructures provide access to crystallography for smaller companies that may not have these capabilities in-house.

Still, it is important to remember that a crystallographic model is just that – a model – and can be misleading. For example, particularly in the case of lower resolution structures, it is possible to misassign the position or conformation of a ligand. In severe cases the structure of the ligand itself could be incorrect, or the ligand may in fact be entirely absent [50]. More frequently, ligands can be affected by so-called crystal contacts: interactions that occur only when the protein is in the crystalline state and not in solution. A recent analysis suggests that this could apply to as many as a third of structures in a widely used database [51]. Finally, a crystal structure provides very limited information on binding affinity, and thus crystallographic data must be correlated with other experimental techniques in order to understand whether ligands have functional activity.

The use of X-ray crystallography in FBDD has been extensively reviewed [41, 43, 45, 52, 53].

2.2.3 Surface Plasmon Resonance

The use of surface plasmon resonance (SPR) to characterize fragment binding dates back a decade, but only recently has it become popular as a primary screening technique. In most cases, a protein is immobilized onto a metal-coated chip and ligands are allowed to flow past. Ligands that bind to the protein cause changes in the reflectivity properties of the metal that are related to the mass of the ligand and the mass of the protein. In some cases, association and dissociation rates can be directly determined, though in the case of fragments these are usually too rapid to be measured.

SPR experiments are relatively rapid and straightforward to set up, and they take less training to run than NMR or X-ray crystallography. However, this apparent simplicity can be dangerous because there are many ways to set up an experiment incorrectly or be misled by artifacts. A review of the 1,413 SPR articles published in 2008 stated rather pointedly that "less than 30% would pass the requirements for high-school chemistry" [54]. When done properly, SPR can be a very useful tool: not only can it provide dissociation constants, it can also provide stoichiometry [55–57].

SPR has rapidly become a dominant technique throughout industry, with Biacore instruments (now owned by GE Healthcare) becoming standard equipment. Roche (and Genentech) make extensive use of the technology, as do Vernalis,

Beactica, Kinetic Discovery, and other companies. There has also been considerable effort to automate the data collection and analysis, both by SPR instrument providers as well as by end users [58, 59].

Finally, it is worth noting that although the protein is usually immobilized, it is also possible to immobilize the ligands themselves and assess binding of the protein [60], an approach taken by Graffinity Pharmaceuticals. SPR approaches are discussed in more detail by Hennig et al. [49].

2.2.4 Other Biophysical Methods

NMR, X-ray crystallography, and SPR are the best-known biophysical methods for FBDD today, but several other approaches can also be used [45].

Interferometry

As in the case of SPR, interferometry relies on a shift in light, in this case caused by a change in both the refractive index and the physical thickness of a layer of protein upon binding to small molecules [56, 61]. Commercially available instruments (such as those from FortéBio) were introduced a few years later than SPR instruments, but the technique seems to be attracting increased interest.

Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) measures the heat released when a ligand binds to a protein; from this the enthalpy and entropy of binding can be calculated [62]. There is some evidence that selecting fragments that bind largely via enthalpic interactions will lead to superior molecules [63], although the data are limited. ITC also has a lower throughput and, in general, a higher protein requirement than other techniques and is thus probably better suited as a secondary rather than a primary screening method.

Mass-Spectrometry

Mass-spectrometry can be used to detect fragments that bind to a protein either covalently or non-covalently. In covalent approaches, such as Tethering [64], developed by researchers at Sunesis Pharmaceuticals, a reactive functionality such as a cysteine is introduced into a protein and used to capture fragments that bind in the vicinity, thus providing some information on the binding site. It is also possible to measure fragments binding to proteins via noncovalent interactions, an approach being pursued by NovAliX [65].

2.2.5 High Concentration Screening

Given the warnings about artifacts in the preceding section, the casual reader may perhaps be surprised that high concentration screening is used at all, but as long as appropriate precautions are taken, biochemical or fluorescent-based screens can be effective and rapid approaches for identifying fragments [66, 67]. For example, Evotec has screened fragments at low milimolar concentration and used confocal fluorescence spectroscopy to detect displacement of a fluorescent probe from a target protein or cleavage of a peptide labeled with a fluorescent probe [68, 69]. Plexxikon has also used high-concentration (100 or 200 µM) functional screening to look for inhibitors or activators of enzyme activity (see Sect. 5.1.3). These efforts ultimately led to two different clinical compounds, PLX4032 and indeglitazar [70, 71].

2.2.6 Computational Methods

Computational methods have a venerable history in FBDD [72]. Computing power continues to increase, though our understanding of molecular interactions is less quantitative than would be necessary to supplant experiments, particularly where proteins are flexible. Nonetheless, there are now many successful examples [73], and computational approaches are likely to play an increasingly important role in the field [74, 75]. Many companies are using computational methods for FBDD, a few of which include Ansaris, BioLeap, BioSolveIT, and MEDIT.

2.2.7 Summary

What should be apparent from this brief tour of methods is that there are many ways to successfully find and characterize fragments, each with its own set of strengths and weaknesses. Which techniques to use will depend as much on institutional resources and expertise as on scientific considerations. The best approach is to forego a single approach: several orthogonal methods should be used in combination. For example, high-concentration screening or computational methods could be used to screen a large set of fragments, the hits could be characterized by SPR, and those that confirm could be further examined by crystallography. This type of workflow is most likely to identify productive fragments while avoiding artifacts [41, 76].

3 Evaluating Fragments

The discussion so far has centered on the theoretical underpinnings of FBDD and how to find – and trust – fragments using a variety of methods. Before turning to some examples, it is important to actually define what constitutes a fragment as well as how to evaluate fragments.

3.1 What Is a Fragment?

FBDD is predicated on the notion that a small fragment can be identified and then either grown, merged, or linked with another fragment to improve potency. Therefore, the fragment should be small enough to avoid creating molecules that are too large to be useful as drugs. Taking Chris Lipinski's Rule of 5 as a springboard, researchers at Astex proposed the Rule of 3 [77]:

- Molecular weight < 300 Da
- Number of hydrogen bond donors ≤ 3
- Number of hydrogen bond acceptors ≤ 3
- ClogP (computed partition coefficient of a compound) ≤ 3

  Additionally, they proposed that:

- Number of rotatable bonds ≤ 3
- Polar surface area (PSA) ≤ 60 Å^2

These are of course only guidelines, and different organizations use different parameters. For example, some groups assembling fragment libraries set an upper limit on molecular weight of 250 or less, whereas others go up to 350, and some do not consider hydrogen bond donors or acceptors.

3.2 Weak Versus Low Affinity: The Importance of Ligand Efficiency

Is an ant weak? Anyone who has casually squashed one that has invaded their picnic will probably say yes. However, if you watch an ant escaping with a crumb, the answer is not so obvious: ants can carry at least ten times their own body weight. This is akin to the situation with fragments: they may have low absolute affinities, but often bind tightly for their size. The question is how to properly measure binding affinity in light of molecular weight.

Probably the most widely used metric is called ligand efficiency, or LE. It was first proposed as a brief letter in *Drug Discovery Today* by Andrew Hopkins and coworkers [78]:

LE = (free energy of ligand binding)/(number of heavy atoms)

The "free energy of ligand binding" is normally expressed in kilocalories per mole and the number of heavy atoms refers to the number of non-hydrogen atoms in the ligand. Of course, the free energy of ligand binding, ΔG_{bind}, is equal to $-RT\ln K_{\text{d}}$, where R is the ideal gas constant, T is temperature, and K_{d} is the dissociation constant. It is also very common for researchers to use IC_{50} values instead of true dissociation constants. Although this shortcut makes it difficult to compare LE values across programs, it is useful for following the progress of a series of compounds within a program.

The beauty of ligand efficiency is its simplicity: it is both intuitive and easy to calculate. Moreover, it gives a useful indication of how drug-like the affinity is with respect to the size of the molecule. For example, a drug with a K_d of 10 nM and a molecular weight of 500 Da (about 38 non-hydrogen atoms) would have an LE of 0.29 kcal/mol/heavy atom. Thus, many researchers look for fragments that have ligand efficiencies of 0.3 kcal/mol/heavy atom or better. Interestingly, a retrospective analysis of lead optimization programs at Abbott revealed that, as the compounds grew in size, each additional heavy atom added 0.3 kcal/mol of binding energy, suggesting that maintaining ligand efficiency at this level is within the realm of standard medicinal chemistry [79]. There are also cases in which ligand efficiency is improved during optimization, but this is something that cannot be assumed, so fragments with higher ligand efficiencies are usually prioritized over fragments with lower ligand efficiencies, all other factors being equal.

What is the upper limit for LE? In 1999 Kuntz and colleagues published a paper called "The maximal affinity of ligands," in which they analyzed the binding data of about 150 natural and synthetic ligands to a number of proteins [80]. By plotting the binding energy against the number of heavy atoms in the ligand, they found a roughly linear relation for the smallest fragments, with a slope of roughly 1.5 kcal/mol/heavy atom [80]. However, as this list includes metal ions and other unusual functionalities, this number represents an unreachable upper limit for molecules that will typically be encountered in a medicinal chemistry program. In practice ligand efficiency values vary considerably based on the target: for some proteins (for example Hsp90 and many kinases) it is not uncommon for inhibitors to have ligand efficiencies well above 0.5 kcal/mol/heavy atom, whereas for more challenging targets, such as most protein–protein interactions, ligand efficiencies may fall significantly below 0.3 kcal/mol/heavy atom [5].

The simplicity of LE has its drawbacks, and in recent years a number of additional metrics for evaluating fragments have been proposed. These include the closely related binding efficiency index (BEI), which has molecular weight in the denominator and the negative log of the inhibition constant in the numerator. This metric was developed at Abbott Laboratories [81] and, in recognition of the need to minimize polar surface area (PSA), the same group also described the surface-binding efficiency index, where the denominator is PSA. A related metric is ligand-efficiency-dependent lipophilicity (LELP), which is simply logP/LE [4]. Finally, in recognition of the fact that, empirically, ligand efficiencies tend to drop as molecular weight increases, two groups have proposed metrics that scale depending on the size of the molecule [82, 83].

4 What Is Fragment-Based Drug Discovery?

Section 1 discussed, somewhat theoretically, fragment linking and fragment growing, and acknowledged that these are but two ends of a continuum. The examples below demonstrate how these simple categories apply in practice, as well as how they break down.

However, before proceeding further, it is useful to ask exactly how to define FBDD. The challenge lies in the fact that more than 400 approved drugs could be defined as fragments using the Rule of 3 described above, yet few if any were discovered using fragment-based approaches.

To get around this difficulty, FBDD can be defined as the discovery of drugs using fragments – or information derived from fragments – that were not discovered using traditional methods. Thus, a lead development program that began with a fragment-sized molecule identified as a nanomolar inhibitor in a high-throughput biochemical screen would not be considered FBDD, but the same fragment identified in a crystallographic or SPR screen could be. Obviously this distinction is somewhat arbitrary, and as fragment techniques and concepts continue to gain ground, the boundaries become increasingly blurry. Difficult categorizations reflect the advancement of fragment-based discovery into wider applications: drug discovery is such a difficult task that it behooves the practitioner to draw upon any and every tool to increase the odds of success.

5 Success Stories in Fragment-Based Drug Discovery: Compounds in the Clinic

When the first comprehensive reviews of FBDD were published in 2004 [84, 85], it was possible to include just about every example of advancing fragments to potent leads that had been published up to that time. In the years since there have been so many examples reported that trying to compile them all is beyond the scope of a single chapter. Two books devoted to FBDD have been published [86, 87], as have numerous general reviews [6, 88–96]. There are also two blogs devoted to the topic (http://practicalfragments.blogspot.com/and http://fbdd-lit.blogspot.com/).

Given that the ultimate goal of any FBDD campaign is a drug, perhaps the best way to triage examples of how crystallography has enabled FBDD is to examine drugs that have entered the clinic starting from fragments [97, 98]; a list of these is shown in Table 1.

Details on how FBDD contributed to development candidates has only been reported for a subset of these compounds. This section will consider all examples of drugs that have entered the clinic from fragment-based efforts where sufficient details have been published in the literature to understand the fragment origins. This reduces the number of examples from dozens to just eight programs. For other (non-clinical) examples, a 2010 review by Christopher Murray and Tom Blundell briefly discusses a score of examples in which structural biology played a major role [53].

5.1 Fragment Growing: Kinase Targets

Almost half of the fragment-derived drugs that have entered the clinic target protein kinases. This reflects both the recent popularity of this class of targets [99] as well

Table 1 Drugs from FBDD efforts that have reached clinical trials

Drug and latest reported development	Company	Target
Phase 3		
PLX-4032	Plexxikon	B-Raf V600E
Phase 2		
ABT 263	Abbott	Bcl-2/Bcl-xL
ABT 869	Abbott	VEGF and PDGFR
AT9283	Astex	Aurora
AT7519	Astex	CDKs 1,2,4,5
LY-517717	Lilly/Protherics	FXa
Indeglitazar	Plexxikon	PPAR agonist
VER-52296/NVP-AUY-922	Vernalis/Novartis	Hsp90
Phase 1		
ABT-518	Abbott	MMP-2 and MMP-9
ABT-737	Abbott	Bcl-2/Bcl-xL
AT13387	Astex	Hsp90
DG-051	deCODE/Emerald	LTA4H
IC-776	Lilly/ICOS	LFA-1
LP-261	Locus	Tubulin
PLX-5568	Plexxikon	Kinase
SGX-523	SGX	Met
SNS-314	Sunesis	Aurora

Some of these drugs have been discontinued (e.g., SGX-523), and for others no development has been reported for some time (e.g., ABT-518). Drugs highlighted in bold are discussed in the text

as the fact that it is relatively straightforward to identify small but highly ligand-efficient fragments that bind to the purine-binding site of kinases [100]. In this section, three examples of clinical kinase inhibitors will be described; each example illustrates clearly one or more aspects of FBDD.

5.1.1 AT7519

One of the clearest examples of the importance of crystallography to FBDD was published in 2008 by researchers from Astex Therapeutics [101]. They were interested in developing inhibitors of the anti-cancer target cyclin-dependent kinase 2 (CDK2), and started by soaking crystals of CDK2 with a library of just a few hundred fragments. More than 30 hits were identified that bound in the active site of the kinase and that made at least one hydrogen bond interaction with the so-called "hinge" region, analogous to the purine moiety of ATP. Using structure-based design, two of these were optimized to low micromolar or mid-nanomolar inhibitors before being abandoned in favor of a series derived from a 1H-indazole fragment, shown in Fig. 4.

Although this fragment (compound **7**) inhibited CDK2 with a fairly low potency (IC$_{50}$ = 185 μM), due to its small size (only nine atoms and molecular weight of 118) it has a high ligand efficiency. Examination of the crystal structure of this fragment bound to CDK2 revealed that substituents off two vectors of the fragment
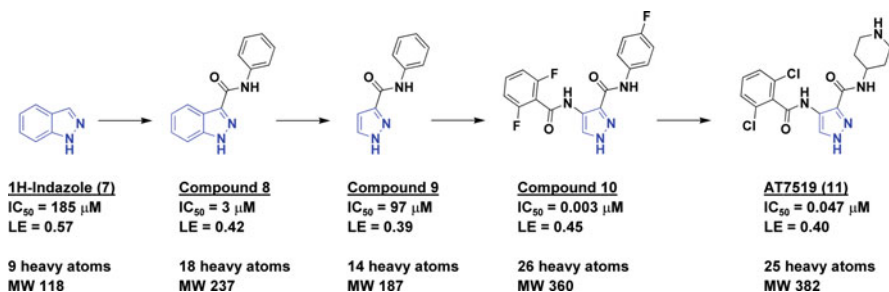
Fig. 4 Fragment growing to discover AT7519

could fill adjacent pockets. Adding a phenyl amide produced a gratifying boost in affinity by providing an additional hydrogen bond to the protein as well as hydrophobic interactions. Trimming back the indazole to a pyrazole led to a significant drop in potency, but only a modest drop in ligand efficiency, and crystallography revealed that the binding mode of compound **9** did not change. This is an important point: going from a 3 μM inhibitor to something with an IC_{50} of 97 μM appears to be a giant step backwards, and a more conservative or less experienced medicinal chemistry team might have abandoned this line of inquiry. Instead, by building off the pyrazole ring, the team was able to regain ligand efficiency and improve potency to low nanomolar levels (compound **10**). Replacing one of the aromatic rings with a piperidine improved cell activity, ultimately leading to AT7519, which as of late 2010 was in Phase II clinical trials for multiple myeloma.

This is a classic example of the "growing" approach for FBDD: a weak fragment was iteratively improved by adding appendages in two directions and tweaking the core fragment. It is also a poster child for the utility of crystallography, which was used every step of the way, from fragment identification through lead optimization. The development of AT7519 is also discussed in greater detail by Davies and Tickle [48].

5.1.2 AT9283

It is common in traditional drug discovery to test compounds made for one project in other projects, and this is also true for FBDD. During the course of the CDK program at Astex, some of the compounds were found to have activity against another kinase implicated in cancer, Aurora A. In particular, compound **12** (Fig. 5) has nanomolar activity for Aurora A, despite its small size. Initial structure-based design using CDK2 led to molecules such as compound **13**, with low nanomolar activity against Aurora A. Subsequently crystals were obtained of compounds **12** and **13** bound to Aurora A kinase itself, revealing that the benzimidazole moiety binds in a cleft that is somewhat more hydrophobic than the corresponding region of CDK2. In order to improve cell-based potency a basic functionality was

Compound 12
CDK2 IC$_{50}$ = 28 µM
CDK2 LE = 0.44

Aurora A IC$_{50}$ = 0.91 µM
Aurora A LE = 0.59

14 heavy atoms
MW 184

Compound 13
CDK2 IC$_{50}$ = 0.052 µM
CDK2 LE = 0.43

Aurora A IC$_{50}$ = 0.0059 µM
Aurora A LE = 0.49

23 heavy atoms
MW 303

Compound 14
CDK2 IC$_{50}$ = 0.14 µM
CDK2 LE = 0.31

Aurora A IC$_{50}$ = 0.0035 µM
Aurora A LE = 0.38

30 heavy atoms
MW 402

AT9283 (15)
CDK2 IC$_{50}$ = 0.51 µM
CDK2 LE = 0.16

Aurora A IC$_{50}$ ~ 0.003 µM
Aurora A LE = 0.41

28 heavy atoms
MW 381

Fig. 5 Fragment growing to discover AT9283
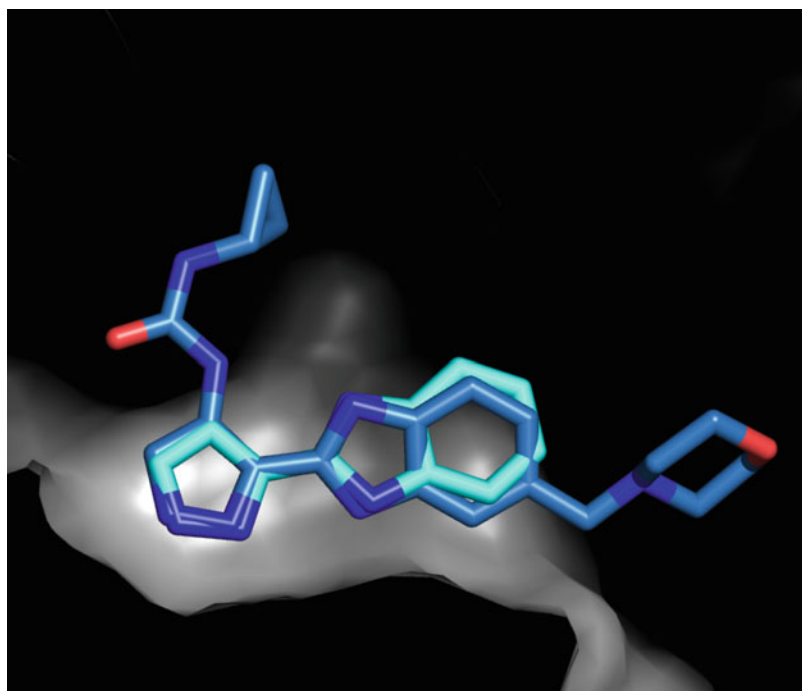
Fig. 6 Superposition of initial fragment (compound **12**, *light blue*) and final compound AT9283 (*dark blue*) bound to Aurora A

introduced, and replacement of the phenyl amide with a cyclopropyl urea moiety led to AT9283 [102], which has advanced to Phase II clinical trials for cancer.

One of the interesting features of this program is the fact that a crystal structure of the final compound, AT9283, was obtained and found to bind in exactly the same manner as the initial fragment; this is shown in Fig. 6. This is a common feature of fragment growing programs, where the initial fragment maintains its position and

orientation while added appendages pick up additional contacts to the protein [103], though there are exceptions [104]. As in the case of AT7519, this program is also a clear indication of the importance of crystallography for advancing a fragment to an experimental drug.

5.1.3 PLX4032

If all goes well, PLX4032 could be the first approved drug to come from a fragment-based approach. The molecule, a selective inhibitor of the kinase B-Raf, is in Phase III trials for metastatic melanoma, where it has displayed impressive activity [105, 106].

Similar to the development of AT9283, the development of PLX4032 began as a crystallographic screening exercise against a kinase that was different from the one it was ultimately used against [70]. The researchers, from Plexxikon, started with a library of about 20,000 "scaffolds" ranging in size from 150 to 350 Da. These were screened in functional assays at 200 µM against several kinases to identify initial hits. The fragment 7-azaindole (compound **16**, Fig. 7) was selected and
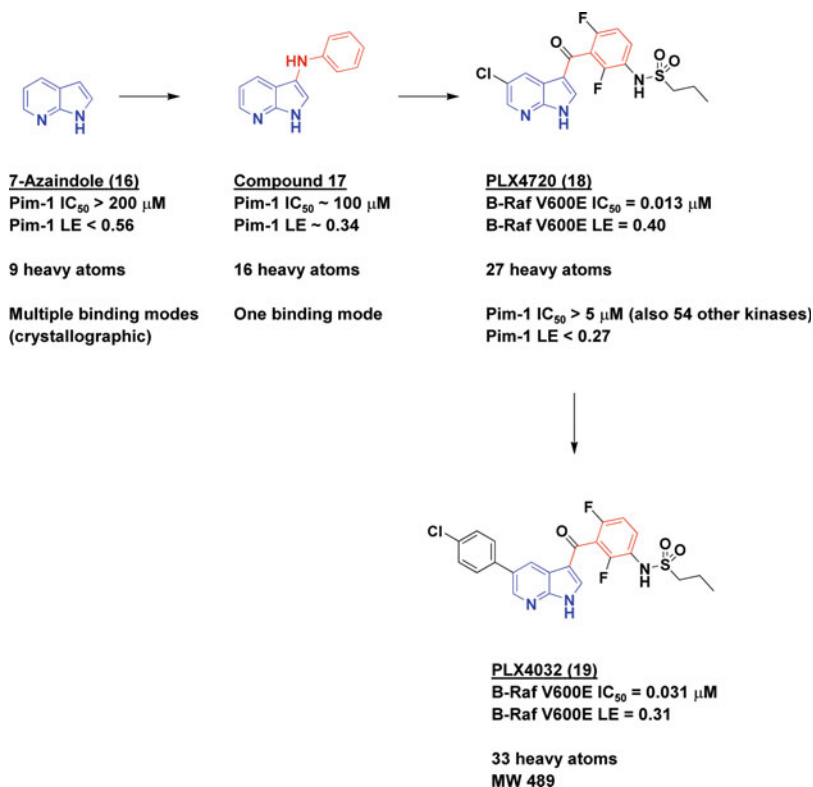
Fig. 7 Fragment growing to discover PLX4032

characterized crystallographically bound to the purine binding site of the kinase Pim-1. Unlike AT9283, this fragment seemed to bind in several orientations. However, elaborating 7-azaindole to give compound **17** boosted its potency and established a single binding mode. Further medicinal chemistry led to PLX4720 (compound **18**) and the related PLX4032 (compound **19**). One of the interesting features of both of these later molecules is that they are quite selective for Raf-family kinases, particularly the oncogenic V600E mutation of B-Raf, compared to many other kinases, including Pim-1, which was targeted by the initial fragment. The question often arises as to how selective a fragment should be, and this example illustrates that selectivity can be built in during the course of optimization [100].

5.2 Fragment Growing: Other Targets

5.2.1 Indeglitazar

Another program from Plexxikon that resulted in a compound in clinical trials is shown in Fig. 8 [71]. This is also a rare example of using FBDD to identify an agonist as opposed to an inhibitor.

The researchers were interested in developing an agonist that would activate all three peroxisome proliferator-activated receptors (PPARα, PPARγ, and PPARδ) as a treatment for Type 2 diabetes mellitus. As in the case of PLX4032, fragments with molecular weights between 150 and 350 Da were screened in a biochemical assay at 100 μM, and compounds that activated two or three of the PPARs were then selected for crystallography. Of the 170 compounds selected, just over a quarter gave structures. Molecular modeling was used to evaluate all previously reported PPAR agonists to try to determine which elements correlated with pan-activity. Compound **20** (see Fig. 8) was a weak agonist of all three PPARs, but the crystal structure in complex with PPARγ showed that it bound entirely in one pocket while leaving unoccupied an adjacent pocket (see Fig. 9). By adding a substituent to fill

Compound 20				Compound 21				Indeglitazar (22)			
	PPARα	PPARγ	PPARδ		PPARα	PPARγ	PPARδ		PPARα	PPARγ	PPARδ
EC$_{50}$	~100	~150	>200 μM	EC$_{50}$	1.3	1.3	10 μM	EC$_{50}$	0.51	0.37	2.7 μM
LE	0.34	0.33	<0.32	LE	0.32	0.32	0.27	LE	0.32	0.32	0.28

16 heavy atoms
MW 219

25 heavy atoms
MW 359

27 heavy atoms
MW 389
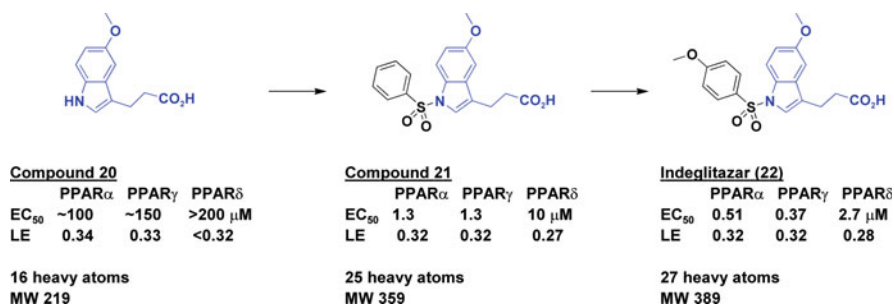
Fig. 8 Fragment growing to discover indeglitazar
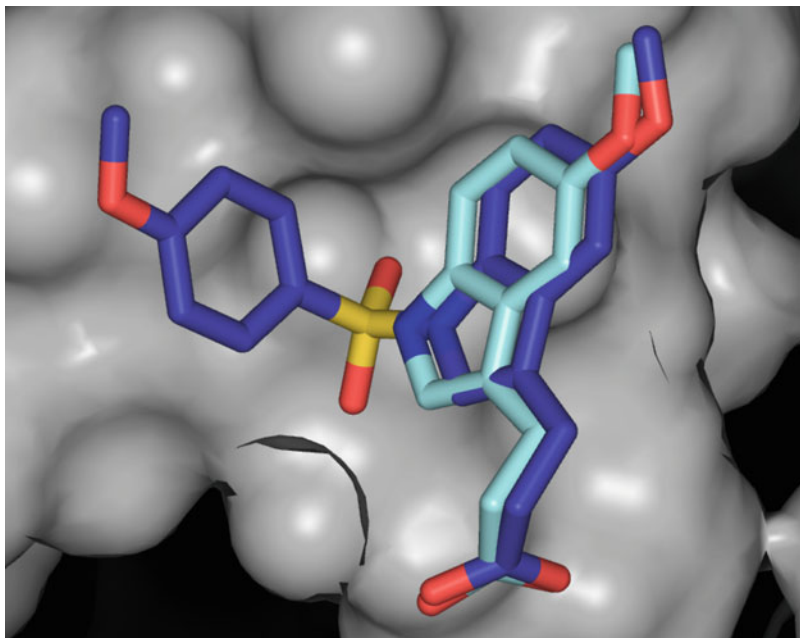
Fig. 9 Superposition of initial fragment (compound **20**, *light blue*) and final compound indeglitazar (*dark blue*) bound to PPARγ

this pocket the potency could be improved by two orders of magnitude, leading to indeglitazar (compound **22**), which progressed to Phase II clinical trials.

Crystallography was essential in this program at multiple points: for the initial compound triaging, for developing hypotheses about how the agonist was binding, and for actually growing the fragment.

5.2.2 AT13387

The protein Hsp90 is a popular anti-cancer target, and several potent molecules have been developed against it, including many from fragment-based efforts [11, 69, 107, 108]. An article by Hubbard and coworkers is devoted entirely to this target, including the discovery of VER-52296/NVP-AUY-922, in which crystallography played a key role [109]. Here the discussion will be confined to AT13387.

Unlike the two other molecules from Astex described in this chapter, screening started not with crystallography but with an NMR screen of 1,600 fragments [110]. Hsp90 uses ATP as a cofactor and also binds ADP. Compounds that were competitive with ADP were taken into crystallography experiments. Of the 125 fragments selected, 26 gave structures, including ethamivan (compound **23** in Fig. 10), which is itself an approved respiratory stimulant. ITC was used to determine the affinity of

Compound 23
$K_d = 790\ \mu M$
LE = 0.26

16 heavy atoms
MW 223

Compound 24
$K_d = 7\ \mu M$
LE = 0.41

17 heavy atoms
MW 235

Compound 25
$K_d = 0.068\ \mu M$
LE = 0.47
IC_{50} (cell)= 17 μM

21 heavy atoms
MW 281

Compound 26
$K_d = 0.00054\ \mu M$
LE = 0.57
IC_{50} (cell)= 0.031 μM

22 heavy atoms
MW 297

AT13387 (27)
$K_d = 0.00071\ \mu M$
LE = 0.42
IC_{50} (cell)= 0.048 μM

30 heavy atoms
MW 409
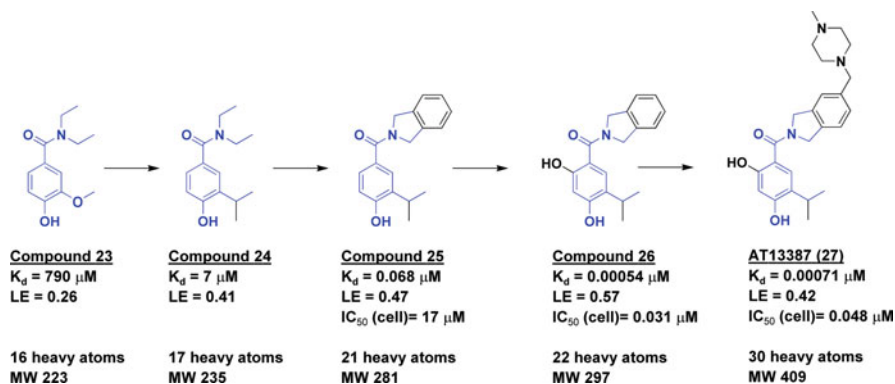
Fig. 10 Fragment growing to discover AT13387

the compounds, and although compound **23** did not stand out on the basis of affinity or ligand efficiency, the crystal structure revealed a similar binding mode to the potent natural product inhibitor radiciol and thus the opportunity for rapid improvements in potency. Replacing the methoxy group with the larger, more lipophilic isopropyl group gave a boost in affinity of roughly 100-fold along with a nice improvement in ligand efficiency. Crystallography confirmed that this change allowed the isopropyl moiety to better fit in a lipophilic pocket. Growing the molecule to compound **25** picked up additional lipophilic interactions in a flexible portion of the protein and resulted in a further improvement in potency of two orders of magnitude. Adding a second hydroxyl group, as found in radiciol, led to compound **26**, which binds more than a million-fold more tightly than the initial fragment. Finally, addition of a positively charged substituent to improve the physicochemical properties of the molecule led to the clinical compound [111].

Although crystallography was not used as an initial screen, it is clear that the technique played a pivotal role throughout the optimization of AT13387. This example is also remarkable for steadily improving not just potency but ligand efficiency throughout the lead optimization campaign, up to the final addition of the solubilizing group. As with previous examples, having very high ligand efficiency at the advanced lead stage meant that the molecule could be optimized for pharmacological parameters without becoming too large.

5.3 Fragment Linking

Fragment linking is perhaps the most conceptually appealing form of FBDD: there is something almost magical about linking two low-affinity fragments and obtaining a high-affinity binder. That said, linking is considerably more challenging than growing. First, two fragments that bind at an appropriate distance must be identified. If the binding sites are too close, it may be impossible to link the fragments;

if too far apart, the resulting molecule will be unacceptably large. Second, the linker needs to be introduced without disrupting any critical functionality from either fragment. Third, the linker must be able to bridge the fragments without causing either of them to change their orientation significantly. Fourth, the linker itself should be in a relaxed, unstrained state. Finally, the linker must not make any unfavorable interactions with the protein. Given these difficulties, it is understandable that fragment linking has resulted in fewer clinical candidates than other methods. However, there are a few, two of which are discussed here.

5.3.1 ABT-518

One of the earliest examples of FBDD was on the protein stromelysin, or matrix metalloproteinase 3 (MMP-3), which is implicated in arthritis and tumor metastasis. This was tackled by researchers at Abbott Laboratories using SAR by NMR. A previous high-throughput screen of 115,000 compounds had failed to identify any hits with potencies better than 10 μM.

One challenge of screening proteases is that they have a tendency to digest themselves, so the researchers added 500 mM acetohydroxamic acid (compound **28** in Fig. 11), which binds to the catalytic zinc and prevents autolytic degradation. With a dissociation constant of 17 mM, acetohydroxamic acid is also perhaps one of the lowest affinity fragments ever successfully advanced [112].

Previous SAR had suggested a preference for hydrophobic residues in peptide substrates of this enzyme, so a screen of hydrophobic compounds was conducted by NMR in the presence of acetohydroxamic acid. A number of weak hits were found, including several biphenyls; a few dozen analogs were synthesized to improve the potency, and compound **29** was found to be the most potent. NMR was used to solve the structure of stromelysin bound to acetohydroxamic acid and a highly-water-soluble biaryl compound. As expected, the acetohydroxamic acid chelates the

Acetohydroxamic acid (28)
K_d = 17,000 μM
LE = 0.48

5 heavy atoms
MW 75

Compound 29
K_d = 20 μM
LE = 0.43

15 heavy atoms
MW 195

Compound 30
IC_{50} = 0.025 μM
LE = 0.49

21 heavy atoms
MW 282

ABT-518 (31)
IC_{50} = 0.012 μM
LE = 0.32

34 heavy atoms
MW 505

n	IC_{50} (μM)	LE
0	0.26	0.45
1	0.025	0.49
2	3.4	0.34
3	3.5	0.32
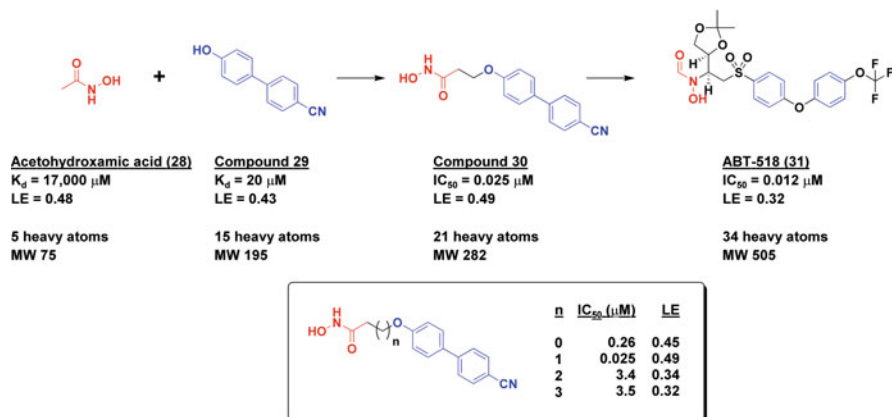
Fig. 11 Fragment linking to discover ABT-518

active site zinc while the biaryl binds in the S1′ pocket. The methyl group of acetohydroxamic acid was pointed towards the biaryl, suggesting that linking the fragments with a simple alkyl chain could improve potency. Gratifyingly, this was successful, with compound **30** showing an improvement in both potency and ligand efficiency. The structure of this compound bound to stromelysin was determined by NMR and found to be similar to the structure of the two fragments binding independently. Importantly the linker length was found to be critical (see Fig. 11): one carbon shorter or longer decreased the potency by one or two orders of magnitude. Subsequent medicinal chemistry to improve the in vivo stability led ultimately to ABT-518, which was tested in a Phase I trial for cancer [113, 114], though no development has been reported for some time. Still, this shows the power of a fragment-linking approach to yield clinical compounds.

5.3.2 ABT-737 and ABT-263

One of the most impressive success stories in FBDD concerns the Bcl-2 family of proteins. These are attractive anti-cancer targets but, because they form protein–protein interactions, it has been challenging to discover small molecule inhibitors. Using SAR by NMR, researchers from Abbot Laboratories identified fragments that bind at two nearby sites on Bcl-xL (Fig. 12) [115]. Fragment 32 was identified
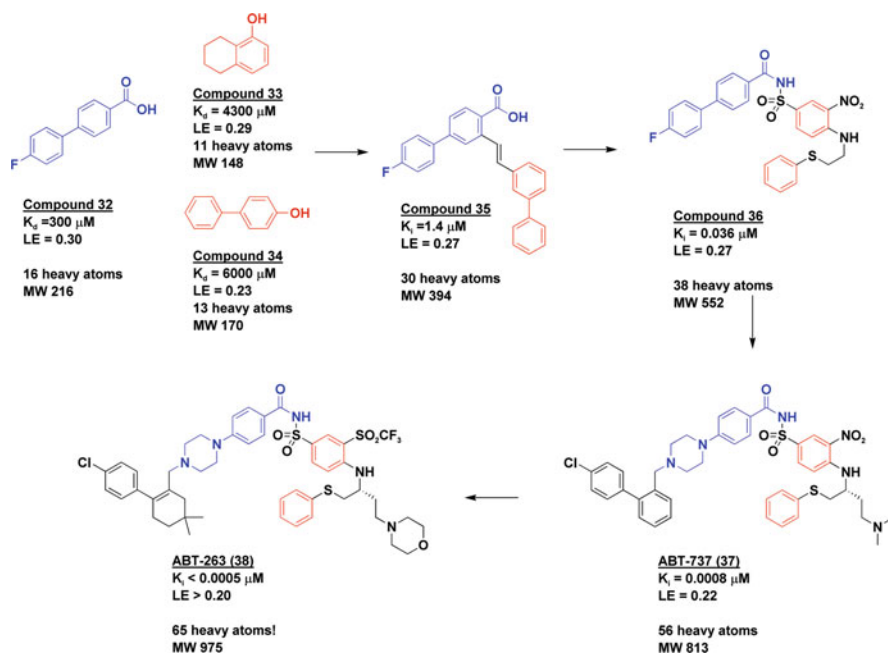
Fig. 12 Fragment linking to discover ABT-737 and ABT-263

from a screen of 10,000 fragments, and fragments 33 and 34 were found by screening 3,500 fragments in the presence of 2 mM of compound **32**. Parallel chemistry was used to link these compounds, ultimately leading to compound **35**. However, an NMR-based model of this compound bound to Bcl-xL suggested that the linker was not ideal: a phenylalanine residue on the protein blocked the right-hand fragment from binding deep in its pocket. This led to a redesign of the linker to replace the carboxylic acid with an acylsulfonamide, which is also negatively charged at physiological pH. This strategy led to a second set of compounds generated by parallel synthesis, ultimately leading to compound **36** [116].

Although compound **36** was potent, it was relatively insoluble and bound tightly to serum albumin. Further medicinal chemistry led to ABT-737, which although potent was not orally bioavailable [117, 118]. Additional medicinal chemistry finally yielded ABT-263, which is orally bioavailable and has improved pharmacodynamics [119].

It is worth considering the structure of ABT-263 for a moment (Fig. 12). This molecule, with a molecular weight approaching 1,000 Da, has traveled some distance from its fragment origins. Yet, prior to the publication of its structure, it is certain that nothing like it would have been found in a high-throughput screening collection. This illustrates the potential of fragment-based approaches to seek out and explore new regions of chemical space.

5.4 Fragment-Assisted Drug Discovery

Earlier in this chapter the notion of fragment-assisted drug discovery was discussed, in which fragment information is used to inform a medicinal chemistry program. The example given here describes how structural information provided from fragment screening contributed to the development of a clinical compound.

5.4.1 DG-051

The research described here was done at several sites then belonging to parent company deCODE, one of which has since regained independence as Emerald Biostructures. The scientists assembled a library of "fragments of life," consisting of just over 1,300 compounds derived from metabolites, derivatives and isosteres of metabolites, and biaryl compounds. A subset of these were then crystallographically screened against the cardiovascular and inflammatory target leukotriene A4 hydrolase (LTA4H), yielding 13 co-structures, including that of compound **39** (Fig. 13) [120]. In some of these, in addition to the fragment, an acetate was observed bound to the zinc ion in the enzyme.

At the same time, the researchers were aware of previous work in the field that had resulted in molecules such as compound **40** [121]. Appending the amine portion of this molecule onto fragment **39** gave a modest boost in potency, but crystallography
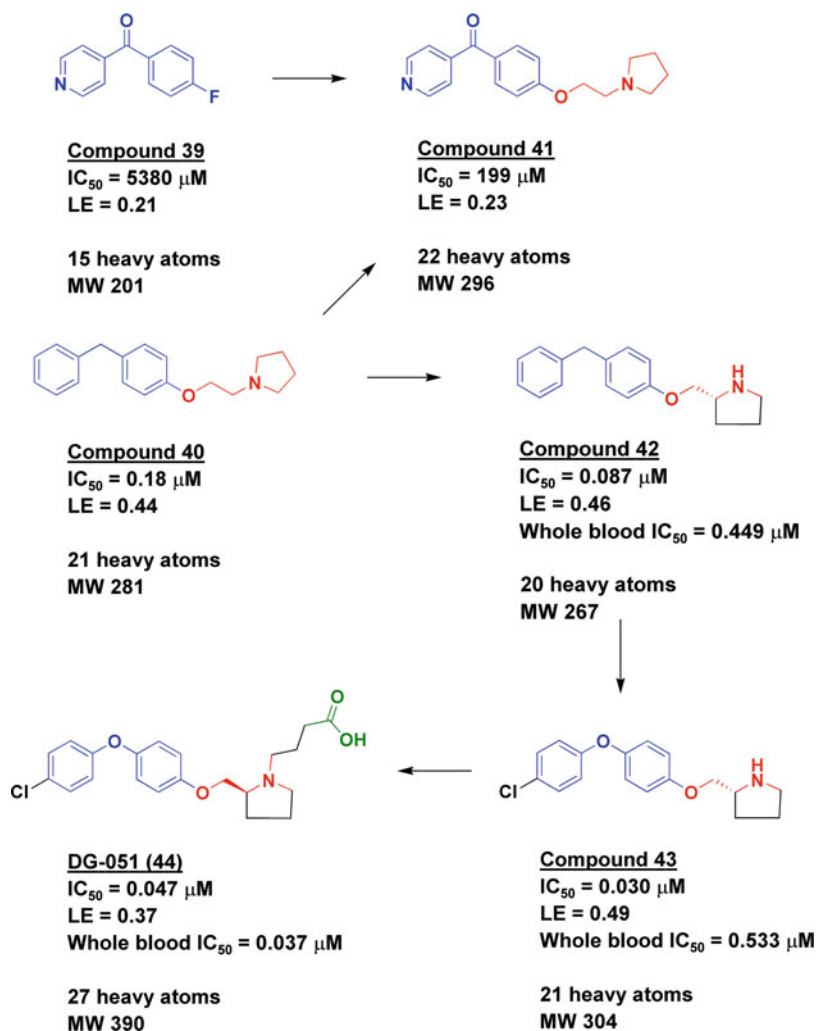
Fig. 13 Fragment-assisted drug discovery to discover DG-051

revealed unfavorable interactions, and compound **40** was pursued instead to generate compound **42**. Further modification to improve pharmacokinetic properties led to compound **43**, and adding a carboxylic acid functionality to try to mimic the acetate sometimes observed in crystal structures produced DG-051. Interestingly, this last step of "linking" the acetate to compound **43** did not improve biochemical potency, but it did improve biological activity as well as solubility [122].

Unlike some of the earlier examples such as AT9283 and PLX4032, the initial fragments themselves did not end up in the final molecule. In theory, it might have been possible to derive DG-051 solely by performing "fast follower" medicinal chemistry on compound **40**. As described, fragment-based approaches, guided by

extensive crystallographic data, played a pivotal role in generating a compound with sufficient novelty and pharmaceutical properties to make it to Phase II clinical trials. This illustrates the power of fragment-based approaches to make useful contributions to a drug program, even if they are not necessarily central.

6   Conclusion

Of the eight programs described above, X-ray crystallography played a direct role in six. The remaining two, both from Abbott, were empowered using SAR by NMR, and at least one of these projects dates back to the 1990s. Although NMR has historically been favored by researchers at Abbot, the company did publish one of the first papers describing crystallographic fragment screening [123], and crystallography is playing an increasingly important role there [124]. Clearly, structural data has played a crucial role in both discovering fragments and advancing them to the clinic. Both NMR and X-ray crystallography are capable of providing detailed structural data, but as X-ray crystallography tools have increased in speed and accessibility this technique seems to have overtaken NMR to assume a dominant position in industry.

This book contains many more detailed examples of how X-ray crystallography, inevitably in combination with other techniques, has played a profound role both in discovering and advancing fragments. But are structural data really essential for advancing fragments to drugs? There is at least one case of taking fragments to potent leads in the absence of structural data [125, 126]. Moreover, it is easy to forget that drug discovery just 30 years ago rarely had access to the kind of structural information taken for granted today, yet by some measures was more productive. Some classes of targets, such as most membrane proteins, are challenging to characterize using experimental structural methods. X-ray crystallography is a key tool for FBDD, but it is worth developing methods that will work in the absence of structural data.

That said, X-ray crystallography provides the most detailed information about protein–ligand interactions and has established itself as an essential tool for FBDD. In the next few years, many more of the drugs that enter the clinic will have started as fragments and advanced with the aid of crystallography. Hopefully, some of these will soon be approved and begin making a serious difference to human health.

References

1. Adams D (1980) The hitchhiker's guide to the galaxy, 1st American edn. Harmony Books, New York
2. Bohacek RS, McMartin C, Guida WC (1996) The art and practice of structure-based drug design: a molecular modeling perspective. Med Res Rev 16:3–50
3. Hann MM, Oprea TI (2004) Pursuing the leadlikeness concept in pharmaceutical research. Curr Opin Chem Biol 8:255–263

4. Keseru GM, Makara GM (2009) The influence of lead discovery strategies on the properties of drug candidates. Nat Rev Drug Discov 8:203–212

5. Wells JA, McClendon CL (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature 450:1001–1009

6. Hajduk PJ, Greer J (2007) A decade of fragment-based drug design: strategic advances and lessons learned. Nat Rev Drug Discov 6:211–219

7. Fink T, Reymond JL (2007) Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F: assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery. J Chem Inf Model 47:342–353

8. Jencks WP (1981) On the attribution and additivity of binding energies. Proc Nat Acad Sci USA 78:4046–4050

9. Shuker SB, Hajduk PJ, Meadows RP et al (1996) Discovering high-affinity ligands for proteins: SAR by NMR. Science 274:1531–1534

10. Chung S, Parker JB, Bianchet M et al (2009) Impact of linker strain and flexibility in the design of a fragment-based inhibitor. Nat Chem Biol 5:407–413

11. Huth JR, Park C, Petros AM et al (2007) Discovery and design of novel HSP90 inhibitors using multiple fragment-based design strategies. Chem Biol Drug Des 70:1–12

12. Whittaker M (2009) Picking up the pieces with FBDD or FADD: invest early for future success. Drug Discov Today 14:623–624

13. Hann MM, Leach AR, Harper G (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. J Chem Inf Comput Sci 41:856–864

14. Rishton GM (2003) Nonleadlikeness and leadlikeness in biochemical screening. Drug Discov Today 8:86–96

15. Baell JB, Holloway GA (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J Med Chem 53:2719–2740

16. Baell JB (2010) Observations on screening-based research and some concerning trends in the literature. Future Med Chem 2:1529–1546

17. Guertin KR, Setti L, Qi L et al (2003) Identification of a novel class of orally active pyrimido [5,4-3][1,2,4]triazine-5,7-diamine-based hypoglycemic agents with protein tyrosine phosphatase inhibitory activity. Bioorg Med Chem Lett 13:2895–2898

18. Tjernberg A, Hallen D, Schultz J et al (2004) Mechanism of action of pyridazine analogues on protein tyrosine phosphatase 1B (PTP1B). Bioorg Med Chem Lett 14:891–895

19. Yi F, Regan L (2008) A novel class of small molecule inhibitors of Hsp90. ACS Chem Biol 3:645–654

20. Lor LA, Schneck J, McNulty DE et al (2007) A simple assay for detection of small-molecule redox activity. J Biomol Screen 12:881–890

21. Johnston PA, Soares KM, Shinde SN et al (2008) Development of a 384-well colorimetric assay to quantify hydrogen peroxide generated by the redox cycling of compounds in the presence of reducing agents. Assay Drug Dev Technol 6:505–518

22. Soares KM, Blackmon N, Shun TY et al (2010) Profiling the NIH small molecule repository for compounds that generate H2O2 by redox cycling in reducing environments. Assay Dev Technol 8:152–174

23. McGovern SL, Caselli E, Grigorieff N et al (2002) A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. J Med Chem 45:1712–1722

24. Seidler J, McGovern SL, Doman TN et al (2003) Identification and prediction of promiscuous aggregating inhibitors among known drugs. J Med Chem 46:4477–4486

25. Babaoglu K, Simeonov A, Irwin JJ et al (2008) Comprehensive mechanistic analysis of hits from high-throughput and docking screens against beta-lactamase. J Med Chem 51:2502–2511

26. Ferreira RS, Bryant C, Ang KK et al (2009) Divergent modes of enzyme inhibition in a homologous structure-activity series. J Med Chem 52:5005–5008

27. Feng BY, Shoichet BK (2006) A detergent-based assay for the detection of promiscuous inhibitors. Nat Protoc 1:550–553

28. Shoichet BK (2006) Screening in a spirit haunted world. Drug Discov Today 11:607–615

29. Pellecchia M, Bertini I, Cowburn D et al (2008) Perspectives on NMR in drug discovery: a technique comes of age. Nat Rev Drug Discov 7:738–745

30. Mayer M, Meyer B (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. Angew Chem Int Ed Engl 38:1784–1788

31. Becattini B, Culmsee C, Leone M et al (2006) Structure-activity relationships by interligand NOE-based design and synthesis of antiapoptotic compounds targeting Bid. Proc Natl Acad Sci USA 103:12602–12606

32. Becattini B, Pellecchia M (2006) SAR by ILOEs: an NMR-based approach to reverse chemical genetics. Chemistry 12:2658–2662

33. Sledz P, Silvestre HL, Hung AW et al (2010) Optimization of the interligand Overhauser effect for fragment linking: application to inhibitor discovery against Mycobacterium tuberculosis pantothenate synthetase. J Am Chem Soc 132:4544–4545

34. Vanwetswinkel S, Heetebrij RJ, van Duynhoven J et al (2005) TINS, target immobilized NMR screening: an efficient and sensitive method for ligand discovery. Chem Biol 12:207–216

35. Fruh V, Zhou Y, Chen D et al (2010) Application of fragment-based drug discovery to membrane proteins: identification of ligands of the integral membrane enzyme DsbB. Chem Biol 17:881–891

36. Pellecchia M, Becattini B, Crowell KJ et al (2004) NMR-based techniques in the hit identification and optimisation processes. Expert Opin Ther Targets 8:597–611

37. Zartler ER, Shapiro MJ (2006) Protein NMR-based screening in drug discovery. Curr Pharm Des 12:3963–3972

38. Sem DS (2006) NMR-guided fragment assembly. In: Jahnke W, Erlanson DA (eds) Fragment-based approaches in drug discovery. Wiley-VCH, Weinheim, Germany

39. Hajduk PJ, Huth JR, Sun C (2006) SAR by NMR: an analysis of potency gains realized through fragment-linking and fragment-elaboration strategies for lead generation. In: Jahnke W, Erlanson DA (eds) Fragment-based approaches in drug discovery. Wiley-VCH, Weinheim

40. Zartler ER, Mo H (2007) Practical aspects of NMR-based fragment discovery. Curr Top Med Chem 7:1592–1599

41. Hubbard RE, Davis B, Chen I et al (2007) The SeeDs approach: integrating fragments into drug discovery. Curr Top Med Chem 7:1568–1581

42. Dalvit C (2009) NMR methods in fragment screening: theory and a comparison with other biophysical techniques. Drug Discov Today 14:1051–1057

43. Jhoti H, Cleasby A, Verdonk M et al (2007) Fragment-based screening using X-ray crystallography and NMR spectroscopy. Curr Opin Chem Biol 11:485–493

44. Kobayashi M, Retra K, Figaroa F et al (2010) Target immobilization as a strategy for NMR-based fragment screening: comparison of TINS, STD, and SPR for fragment hit identification. J Biomol Screen 15:978–989

45. Gozalbes R, Carbajo RJ, Pineda-Lucena A (2010) Contributions of computational chemistry and biophysical techniques to fragment-based drug discovery. Curr Med Chem 17:1769–1794

46. 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_2011_183

47. Bauman JD, Patel D, Arnold E (2011) Fragment screening and HIV therapeutics. Top Curr Chem. doi:10.1007/128_2011_232

48. Davies TG, Tickle IJ (2011) Fragment screening using X-ray crystallography. Top Curr Chem. doi:10.1007/128_2011_179

49. Hennig M, Ruf A, Huber W (2011) Combining biophysical screening and X-ray crystallography for fragment-based drug discovery. Top Curr Chem. doi:10.1007/128_2011_225

50. Davis AM, St-Gallay SA, Kleywegt GJ (2008) Limitations and lessons in the use of X-ray structural information in drug design. Drug Discov Today 13:831–841
51. Sondergaard CR, Garrett AE, Carstensen T et al (2009) Structural artifacts in protein-ligand X-ray structures: implications for the development of docking scoring functions. J Med Chem 52:5673–5684
52. 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
53. Murray CW, Blundell TL (2010) Structural biology in fragment-based drug design. Curr Opin Struct Biol 20:497–507
54. Rich RL, Myszka DG (2010) Grading the commercial optical biosensor literature-class of 2008: 'The mighty binders'. J Mol Recognit 23:1–64
55. Perspicace S, Banner D, Benz J et al (2009) Fragment-based screening using surface plasmon resonance technology. J Biomol Screen 14:337–349
56. Proll F, Fechner P, Proll G (2009) Direct optical detection in fragment-based screening. Anal Bioanal Chem 393:1557–1562
57. Navratilova I, Hopkins AL (2010) Fragment screening by surface plasmon resonance. ACS Med Chem Lett 1:44–48
58. Rich RL, Myszka DG (2010) Kinetic analysis and fragment screening with Fujifilm AP-3000. Anal Biochem 402:170–178
59. Kreatsoulas C, Narayan K (2010) Algorithms for the automated selection of fragment-like molecules using single-point surface plasmon resonance measurements. Anal Biochem 402:179–184
60. Neumann T, Junker HD, Schmidt K et al (2007) SPR-based fragment screening: advantages and applications. Curr Top Med Chem 7:1630–1642
61. Concepcion J, Witte K, Wartchow C et al (2009) Label-free detection of biomolecular interactions using biolayer interferometry for kinetic characterization. Comb Chem High Throughput Screen 12:791–800
62. Ladbury JE, Klebe G, Freire E (2010) Adding calorimetric data to decision making in lead discovery: a hot tip. Nat Rev Drug Discov 9:23–27
63. Scott AD, Phillips C, Alex A et al (2009) Thermodynamic optimisation in drug discovery: a case study using carbonic anhydrase inhibitors. ChemMedChem 4:1985–1989
64. Erlanson DA, Wells JA, Braisted AC (2004) Tethering: fragment-based drug discovery. Annu Rev Biophys Biomol Struct 33:199–223
65. Hannah V, Atmanene C, Zeyer D et al (2010) Native MS: an 'ESI' way to support structure- and fragment-based drug discovery. Future Med Chem 2:35–49
66. Barker J, Courtney S, Hesterkamp T et al (2006) Fragment screening by biochemical assay. Expert Opin Drug Discovery 1:225–236
67. Slack M, Winkler D, Kramer J et al (2009) A multiplexed approach to hit finding. Curr Opin Drug Discov Devel 12:351–357
68. Godemann R, Madden J, Kramer J et al (2009) Fragment-based discovery of BACE1 inhibitors using functional assays. Biochemistry 48:10743–10751
69. Barker JJ, Barker O, Boggio R et al (2009) Fragment-based identification of Hsp90 inhibitors. ChemMedChem 4:963–966
70. Tsai J, Lee JT, Wang W et al (2008) Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. Proc Natl Acad Sci USA 105:3041–3046
71. Artis DR, Lin JJ, Zhang C et al (2009) Scaffold-based discovery of indeglitazar, a PPAR pan-active anti-diabetic agent. Proc Natl Acad Sci USA 106:262–267
72. Caflisch A, Miranker A, Karplus M (1993) Multiple copy simultaneous search and construction of ligands in binding sites: application to inhibitors of HIV-1 aspartic proteinase. J Med Chem 36:2142–2167

73. Teotico DG, Babaoglu K, Rocklin GJ et al (2009) Docking for fragment inhibitors of AmpC beta-lactamase. Proc Natl Acad Sci USA 106:7455–7460

74. Zoete V, Grosdidier A, Michielin O (2009) Docking, virtual high throughput screening and in silico fragment-based drug design. J Cell Mol Med 13:238–248

75. Law R, Barker O, Barker JJ et al (2009) The multiple roles of computational chemistry in fragment-based drug design. J Comput Aided Mol Des 23:459–473

76. Albert JS, Blomberg N, Breeze AL et al (2007) An integrated approach to fragment-based lead generation: Philosophy, strategy and case studies from AstraZeneca's drug discovery programmes. Curr Top Med Chem 7:1600–1629

77. Congreve M, Carr R, Murray C et al (2003) A 'rule of three' for fragment-based lead discovery? Drug Discov Today 8:876–877

78. Hopkins AL, Groom CR, Alex A (2004) Ligand efficiency: a useful metric for lead selection. Drug Discov Today 9:430–431

79. Hajduk PJ (2006) Fragment-based drug design: how big is too big? J Med Chem 49:6972–6976

80. Kuntz ID, Chen K, Sharp KA et al (1999) The maximal affinity of ligands. Proc Natl Acad Sci USA 96:9997–10002

81. Abad-Zapatero C, Metz G (2005) Ligand efficiency indices as guideposts for drug discovery. Drug Discov Today 10:464–469

82. Bembenek SD, Tounge BA, Reynolds CH (2009) Ligand efficiency and fragment-based drug discovery. Drug Discov Today 14:278–283

83. Orita M, Ohno K, Niimi T (2009) Two 'golden ratio' indices in fragment-based drug discovery. Drug Discov Today 14:321–328

84. Erlanson DA, McDowell RS, O'Brien T (2004) Fragment-based drug discovery. J Med Chem 47:3463–3482

85. Rees DC, Congreve M, Murray CW et al (2004) Fragment-based lead discovery. Nat Rev Drug Discov 3:660–672

86. Jahnke W, Erlanson DA (eds) (2006) Fragment-based approaches in drug discovery. Methods and principles in medicinal chemistry, vol 34. Wiley-VCH, Weinheim, Germany

87. Zartler E, Shapiro M (eds) (2008) Fragment-based drug discovery: a practical approach. Wiley, Hoboken

88. Erlanson DA (2006) Fragment-based lead discovery: A chemical update. Curr Opin Biotechnol 17:643–652

89. Leach AR, Hann MM, Burrows JN et al (2006) Fragment screening: an introduction. Mol Biosyst 2:430–446

90. Ciulli A, Abell C (2007) Fragment-based approaches to enzyme inhibition. Curr Opin Biotechnol 18:489–496

91. Fattori D, Squarcia A, Bartoli S (2008) Fragment-based approach to drug lead discovery: overview and advances in various techniques. Drugs R D 9:217–227

92. Congreve M, Chessari G, Tisi D et al (2008) Recent developments in fragment-based drug discovery. J Med Chem 51:3661–3680

93. Murray CW, Rees DC (2009) The rise of fragment-based drug discovery. Nat Chem 1:187–192

94. Schulz MN, Hubbard RE (2009) Recent progress in fragment-based lead discovery. Curr Opin Pharmacol 9:615–621

95. de Kloe GE, Bailey D, Leurs R et al (2009) Transforming fragments into candidates: small becomes big in medicinal chemistry. Drug Discov Today 14:630–646

96. Coyne AG, Scott DE, Abell C (2010) Drugging challenging targets using fragment-based approaches. Curr Opin Chem Biol 14:299–307

97. Chessari G, Woodhead AJ (2009) From fragment to clinical candidate–a historical perspective. Drug Discov Today 14:668–675

98. Erlanson DA (2010) Fragments in the clinic: 2010 edition. In: Practical fragments. http://practicalfragments.blogspot.com/2010/09/fragments-in-clinic-2010-edition.html. Accessed 23 Dec 2010

99. Li R, Stafford JA (eds) (2009) Kinase inhibitor drugs. Wiley series in drug discovery and development. Wiley, Hoboken

100. Akritopoulou-Zanze I, Hajduk PJ (2009) Kinase-targeted libraries: the design and synthesis of novel, potent, and selective kinase inhibitors. Drug Discov Today 14:291–297

101. Wyatt PG, Woodhead AJ, Berdini V et al (2008) Identification of N-(4-piperidinyl)-4-(2,6-dichlorobenzoylamino)-1H-pyrazole-3-carboxamide (AT7519), a novel cyclin dependent kinase inhibitor using fragment-based X-ray crystallography and structure based drug design. J Med Chem 51:4986–4999

102. Howard S, Berdini V, Boulstridge JA et al (2009) Fragment-based discovery of the pyrazol-4-yl urea (AT9283), a multitargeted kinase inhibitor with potent aurora kinase activity. J Med Chem 52:379–388

103. Erlanson D, Braisted A, Raphael D et al (2000) Site-directed ligand discovery. Proc Natl Acad Sci USA 97:9367–9372

104. Mpamhanga CP, Spinks D, Tulloch LB et al (2009) One scaffold, three binding modes: novel and selective pteridine reductase 1 inhibitors derived from fragment hits discovered by virtual screening. J Med Chem 52:4454–4465

105. Bollag G, Hirth P, Tsai J et al (2010) Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. Nature 467:596–599

106. Flaherty KT, Puzanov I, Kim KB et al (2010) Inhibition of mutated, activated BRAF in metastatic melanoma. N Engl J Med 363:809–819

107. Brough PA, Aherne W, Barril X et al (2008) 4,5-diarylisoxazole Hsp90 chaperone inhibitors: potential therapeutic agents for the treatment of cancer. J Med Chem 51:196–218

108. Brough PA, Barril X, Borgognoni J et al (2009) Combining hit identification strategies: fragment-based and in silico approaches to orally active 2-aminothieno[2,3-d]pyrimidine inhibitors of the Hsp90 molecular chaperone. J Med Chem 52:4794–4809

109. Roughley S, Wright L, Brough P, Massey A, Hubbard RE (2011) Hsp90 inhibitors and drugs from fragment and virtual screening. Top Curr Chem. doi:10.1007/128_181

110. Murray CW, Carr MG, Callaghan O et al (2010) Fragment-based drug discovery applied to Hsp90. Discovery of two lead series with high ligand efficiency. J Med Chem 53:5942–5955

111. Woodhead AJ, Angove H, Carr MG et al (2010) Discovery of (2,4-dihydroxy-5-isopropyl-phenyl)-[5-(4-methylpiperazin-1-ylmethyl)-1,3-di hydroisoindol-2-yl]methanone (AT13387), a novel inhibitor of the molecular chaperone Hsp90 by fragment based drug design. J Med Chem 53:5956–5969

112. Hajduk PJ, Sheppard G, Nettesheim DG et al (1997) Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. J Am Chem Soc 119:5818–5827

113. Hajduk PJ, Shuker SB, Nettesheim DG et al (2002) NMR-based modification of matrix metalloproteinase inhibitors with improved bioavailability. J Med Chem 45:5628–5639

114. Wada CK (2004) The evolution of the matrix metalloproteinase inhibitor drug discovery program at Abbott Laboratories. Curr Top Med Chem 4:1255–1267

115. Oltersdorf T, Elmore SW, Shoemaker AR et al (2005) An inhibitor of Bcl-2 family proteins induces regression of solid tumours. Nature 435:677–681

116. Petros AM, Dinges J, Augeri DJ et al (2006) Discovery of a potent inhibitor of the antiapoptotic protein Bcl-xL from NMR and parallel synthesis. J Med Chem 49:656–663

117. Wendt MD, Shen W, Kunzer A et al (2006) Discovery and structure-activity relationship of antagonists of B-cell lymphoma 2 family proteins with chemopotentiation activity in vitro and in vivo. J Med Chem 49:1165–1181

118. Bruncko M, Oost TK, Belli BA et al (2007) Studies leading to potent, dual inhibitors of Bcl-2 and Bcl-xL. J Med Chem 50:641–662

119. Park CM, Bruncko M, Adickes J et al (2008) Discovery of an orally bioavailable small molecule inhibitor of prosurvival B-cell lymphoma 2 proteins. J Med Chem 51:6902–6915

120. Davies DR, Mamat B, Magnusson OT et al (2009) Discovery of leukotriene A4 hydrolase inhibitors using metabolomics biased fragment crystallography. J Med Chem 52:4694–4715

121. Penning TD, Chandrakumar NS, Chen BB et al (2000) Structure-activity relationship studies on 1-[2-(4-phenylphenoxy)ethyl]pyrrolidine (SC-22716), a potent inhibitor of leukotriene A(4) (LTA(4)) hydrolase. J Med Chem 43:721–735
122. Sandanayaka V, Mamat B, Mishra RK et al (2010) Discovery of 4-[(2S)-2-{[4-(4-chloro-phenoxy)phenoxy]methyl}-1-pyrrolidinyl]butanoic acid (DG-051) as a novel leukotriene A4 hydrolase inhibitor of leukotriene B4 biosynthesis. J Med Chem 53:573–585
123. Nienaber VL, Richardson PL, Klighofer V et al (2000) Discovering novel ligands for macromolecules using X-ray crystallographic screening. Nat Biotechnol 18:1105–1108
124. Sun C (2010) Targeting the intractable. Oral presentation at: Fragment-based lead discovery conference 2010, Philadelphia, 10–13 October 2010
125. Haydon DJ, Stokes NR, Ure R et al (2008) An inhibitor of FtsZ with potent and selective anti-staphylococcal activity. Science 321:1673–1675
126. Czaplewski LG, Collins I, Boyd EA et al (2009) Antibacterial alkoxybenzamide inhibitors of the essential bacterial cell division protein FtsZ. Bioorg Med Chem Lett 19:524–527

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 under-appreciated and impractical due to its relatively low throughput. In this chapter we discuss the various benefits associated with fragment-screening by X-ray crystal-lography, 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

T.G. Davies (✉) and I.J. Tickle
Astex Therapeutics Ltd, 436 Cambridge Science Park, Milton Road, Cambridge CB4 0QA, UK
e-mail: t.davies@astex-therapeutics.com

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 energeti-
cally 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 experi-
ment [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 dissoci-
ation 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
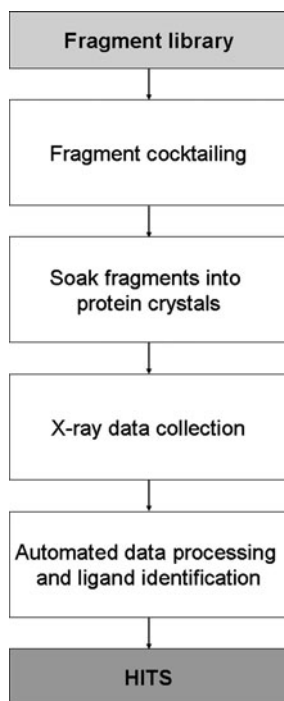
Fig. 1 Work-flow for a typical crystallographic fragment screen

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 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 Å^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

X = C,N

X = C,N,O

X	**Y**
C=O	NH
NH	C=O
C	C
C	O
O	C
NH	C
C	NH

b

Pyrrole   Imidazole   Furan   Benzene   Cyclohexane

Pyridine   Pyrimidine   1,3,5-triazine   Piperidine

Indole   1H-indazole   Benzimidazole   Purine

Napthalene   Quinoline   Isoquinoline   Quinazoline   Pteridine

Fig. 2 (continued)

c

Primary amine   Amidine   Amide

Hydroxyl   Carboxyl

Tetrazole   Sulfonamide   Hydroxamate

d

Fluoro   Chloro   Phenyl   Tetrazole   Methyl   Acetal

Methyl   Methoxy   T-butyl   Trifluoro   Methyl sulfone

Carbon substituents   **Nitrogen substituents**

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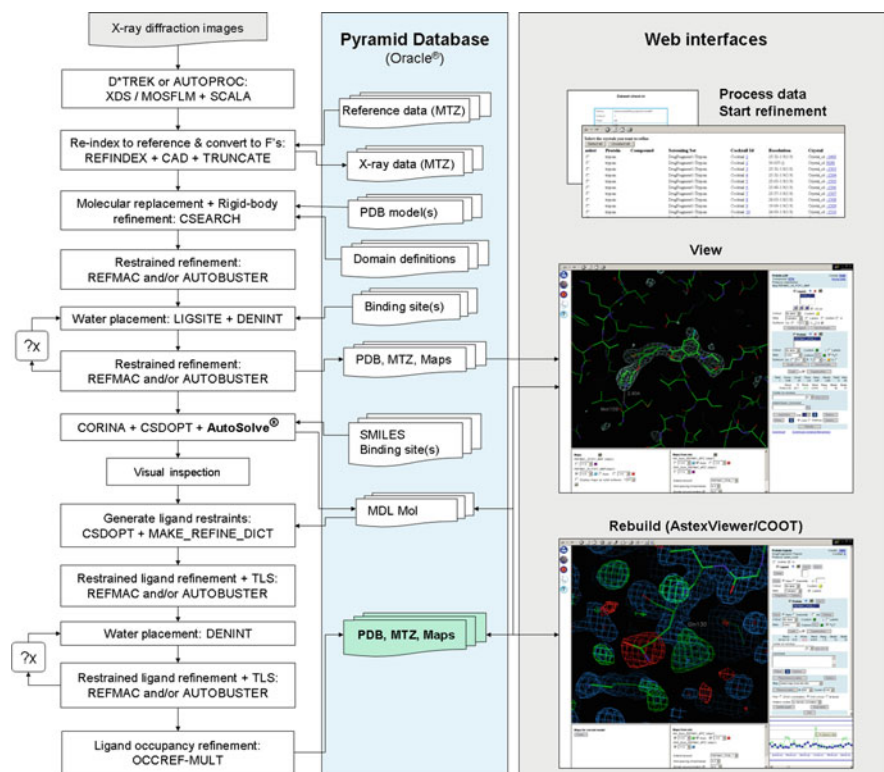
Fig. 3 Flow-diagram summarizing the AutoSolve platform and its automated data processing, refinement and ligand placement procedures. All data handling is carried out within an Oracle database, and the process is driven from a series of web-based interfaces

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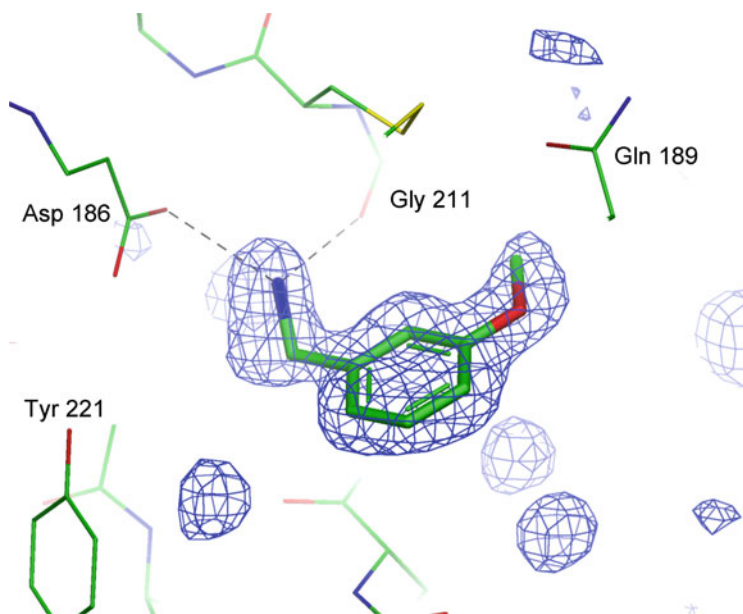
Fig. 4 AutoSolve solution for a fragment hit against trypsin. The initial $mF_o - DF_c$ difference Fourier contoured at 3σ is shown for the active site region. Despite the pseudosymmetric shape of the electron density, AutoSolve correctly orientates the ligand to satisfy the most likely hydrogen bonding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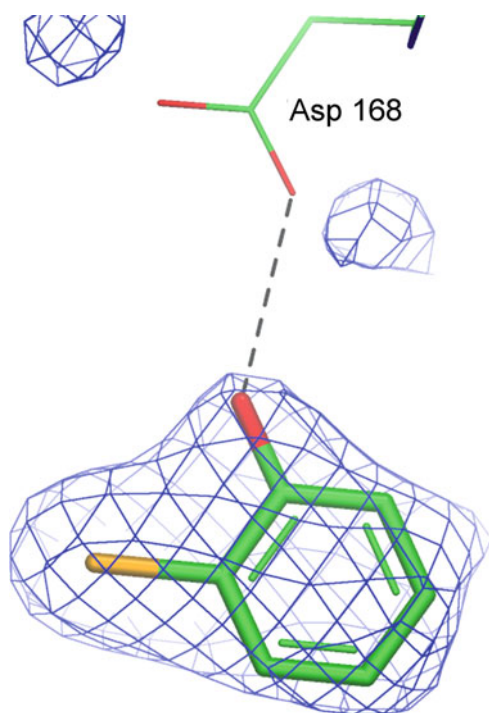

Fig. 5 Top ranked AutoSolve solution for a fragment-screening experiment against the kinase p38. The initial $mF_o - DF_c$ difference Fourier contoured at 3σ is shown for the active site region, and hydrogen bonds between protein and ligand are denoted by *dashed lines*. Figure adapted from Mooij et 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
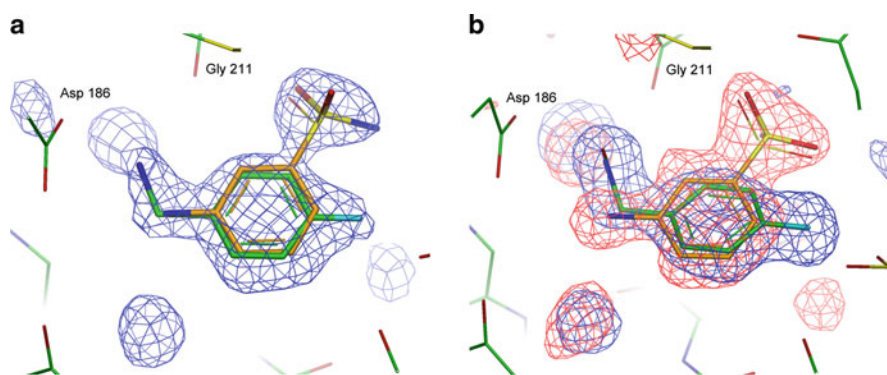
Fig. 6 (**a**) AutoSolve solutions for fragment-screening experiment against trypsin, with simultaneous binding of two compounds from a cocktail of eight. (**b**) Overlay of AutoSolve solutions and electron densities for subsequent deconvolution experiments in which compounds were individually soaked. 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
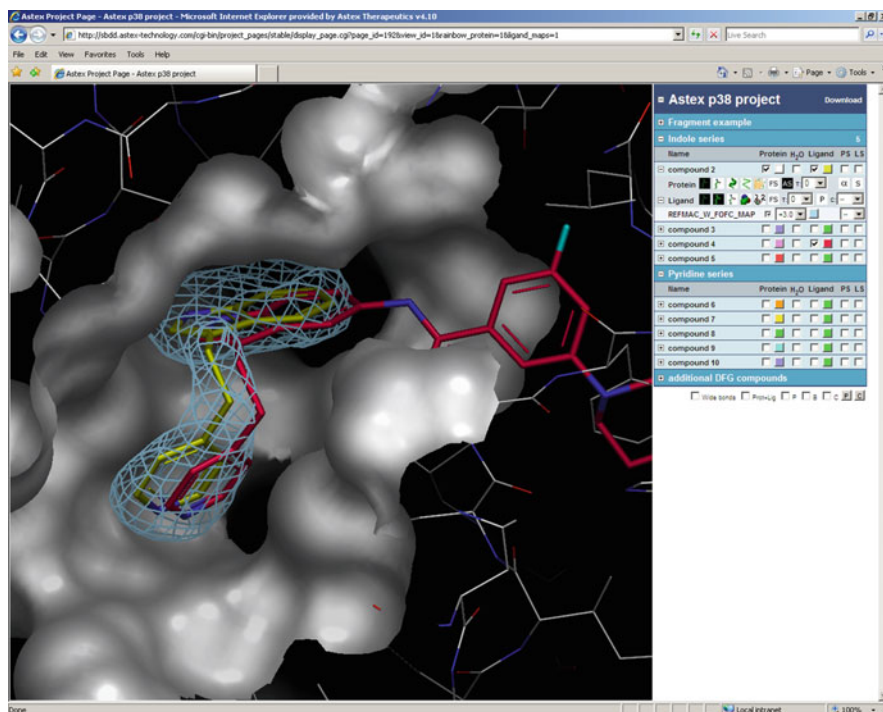
Fig. 7 Overlay page containing protein–ligand structures for the kinase p38. Structures are visualized within AstexViewer (*left-hand pane*), whilst the *right-hand pane* contains folders of display controls for sets of pre-superposed complexes

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⋯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.
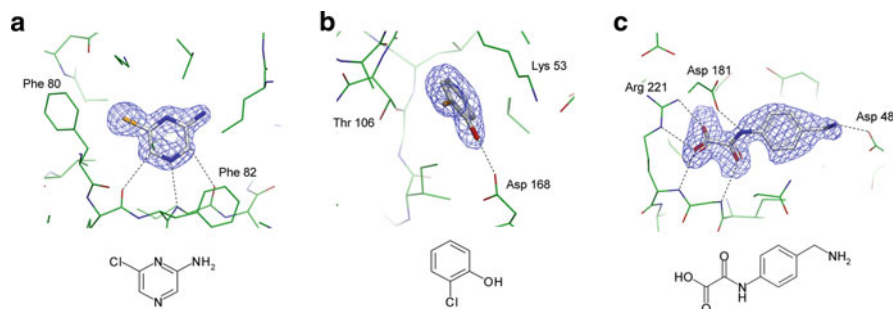
**Fig. 8** Examples of fragment hits against selected targets, illustrating different aspects of molecular recognition. (**a**) CDK2 (neutral hydrogen bonding), (**b**) p38 (lipophilic interactions), (**c**) PTB1B (charge–charge interactions). Hydrogen bonds and electrostatic interactions are denoted by *dashed lines*, and the initial $mF_o - DF_c$ difference Fouriers 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 pyazole, **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
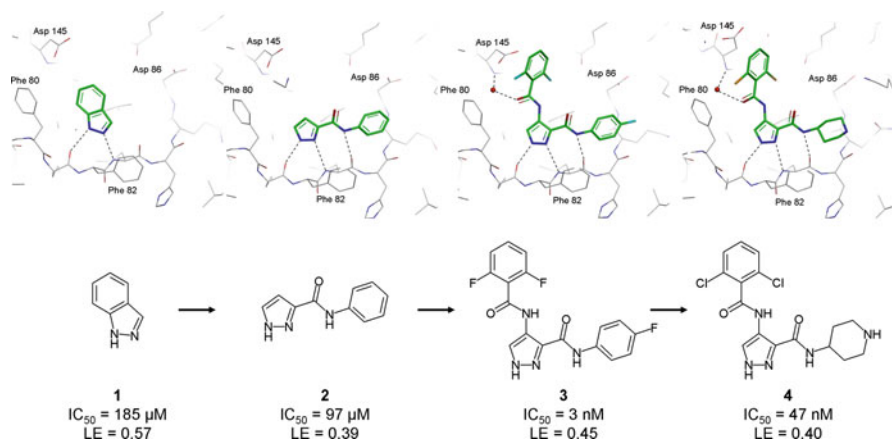
Fig. 9 Fragment evolution for the target CDK2 as described in the text. Key hydrogen bonding interactions with the protein are denoted by *dashed lines*

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 μM 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 μM) revealed that it forms a number of aromatic contacts
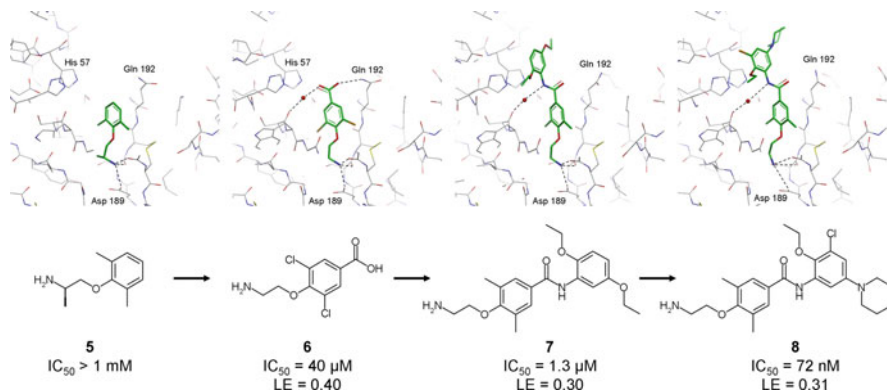
Fig. 10 Fragment evolution for the target urokinase as described in the text. Key hydrogen bonding interactions with the protein are denoted by *dashed lines*

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 ($IC_{50} = 72$ 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 ($F_{rat} = 60\%$). 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.

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 duscovery. 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, Mande SC, Hol WGJ (1997) Antitrypanosomiasis drug development based on structures of glycolitic 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, Jhoti 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, Jhoti 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, Jhoti H (2005) J Med Chem 48:403
44. Blundell TL, Abell C, Cleasby A, Hartshorn MJ, Tickle IJ, Parasini E, Jhoti 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, Jhoti 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 EA, 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 A50:157
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, Zeslawska 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

Hsp90 Inhibitors and Drugs from Fragment and Virtual Screening

**Stephen Roughley, Lisa Wright, Paul Brough, Andrew Massey,
and Roderick E. Hubbard**

Abstract We have previously reported the structure-based optimisation of a number of series of potent compounds progressed as clinical candidates for oncology through inhibition of the ATPase activity of the molecular chaperone, Hsp90. The starting point for these candidates was compounds discovered using a combination of structure-based hit identification methods. This chapter summarises the overall story of how these methods were applied. Virtual screening of commercially available compounds identified a number of classes of compounds. At the same time, an initial fragment screen identified 17 fragments of various classes that bound to the N-terminal domain of Hsp90 with weak (0.5–10 mM) affinity. A subsequent screen identified a total of 60 compounds. This collection of fragments and virtual screening hits were progressed in a number of ways. Although two fragments could be observed binding together in the active site, the synthetic effort required to link these fragments was judged too high. For the resorcinol class of fragments, limited library synthesis generated compounds in the 1–10 μM range. In addition, the resorcinol substructure was used to select commercially available compounds that were filtered using focussed docking in the Hsp90 active site to select further sets of compounds for assay. This identified structural motifs that were exploited during lead optimisation to generate AUY922, currently in Phase II clinical trials. In a separate campaign, features identified in the structures of fragments, evolved fragments and virtual screening hits bound to Hsp90 were combined to generate an oral series of compounds, progressed to preclinical

S. Roughley, L. Wright, P. Brough, and A. Massey
Vernalis (R&D) Ltd, Granta Park, Abington, Cambridge CB21 6GB, UK

R.E. Hubbard (✉)
Vernalis (R&D) Ltd, Granta Park, Abington, Cambridge CB21 6GB, UK
and
YSBL, University of York, Heslington, York YO10 5DD, UK
and
HYMS, University of York, Heslington, York YO10 5DD, UK
e-mail: r.hubbard@vernalis.com

candidates. The crystal structures were determined of many of the fragments bound to Hsp90 and provide examples of both maintenance and change of protein conformation on fragment binding. Finally, we analyse the extent to which our initial set of fragments recapitulates the key structural features of the Hsp90 inhibitors published to date.

Keywords Fragment-based drug discovery · Hsp90 · SAR by catalogue · Structure-based drug discovery · Virtual screening

Contents

1   Introduction

The heat shock protein (Hsp90) family of molecular chaperones is a widely expressed family, comprising cytosolic Hsp90α and Hsp90β, mitochondrial TRAP-1 and endoplasmic reticulum Grp94 [1, 2]. In the cell, Hsp90α and Hsp90β exist as large multiprotein complexes in cohort with a variety of co-chaperones such as Aha1, Cdc37, Hip, Hop, Hsp70 and p23. The major cellular function of Hsp90 is to fold as well as maintain the conformational maturation, stability, and activity of other proteins, often referred to as "clients". The ATPase activity, along with the various co-chaperones, is essential for its biological activity. The chaperone cycle, the mechanism by which Hsp90 functions, is complex and requires the sequential binding and dissociation of various co-chaperones as well as the hydrolysis of ATP.

Molecular targets that modulate the activity of multiple oncogenic processes have attracted considerable interest as cancer therapeutic targets in recent years. Many of the proteins so far identified as Hsp90 clients are key components of the oncogenic phenotype involved in controlling many of the hallmarks of cancer [3]. Inhibiting Hsp90 has the potential to affect all the hallmarks of cancer, making it

an exciting potential therapeutic target [4, 5]. Initial evidence for the role of Hsp90 in cancer came from the natural products geldanamycin and radicicol [2, 6, 7]. Both bind to, and inhibit, the NH_2-terminal ATPase activity of Hsp90, resulting in proteasome-dependent client protein degradation and tumour cell growth arrest and death. The semisynthetic geldanamycin analogue 17-allylamino-17-demethoxygel-danamycin (17-AAG, tanespimycin) has undergone preclinical and clinical studies, further validating Hsp90 as a target. This first-in-class Hsp90 inhibitor has demonstrated evidence of target modulation in melanoma, prostate cancer, renal cancer, multiple myeloma, and trastuzumab-refractory breast cancer [8, 9]. The development of 17-AAG has been hampered by several limitations, which include poor solubility necessitating complex formulations or pro-drug approaches, limited bioavailability and hepatotoxicity. The development of tanespimycin was, however, recently halted.

To overcome the limitations of the ansamycin-derived Hsp90 inhibitors such as 17-AAG, significant effort has been placed in identifying novel, fully synthetic small molecule inhibitors that bind to and inhibit the N-terminal ATPase pocket of Hsp90. Novel agents in clinical trials include NVP-AUY922 (Phase II, Vernalis/Novartis), BIIB021 (Phase II, Biogen Idec), STA9090 (Phase II, Synta), HSP990 (Phase I, Vernalis/Novartis), PF-04929113 (SNX5422, Phase I, Pfizer), AT13387 (Phase I, Astex), IPI-493 (Phase I, Infinity) and XL888 (Phase I, Exelixis) [10–16].

This chapter describes the early hit discovery project to identify inhibitors of Hsp90 at Vernalis in 2002–2003. This generated the initial ideas from which series of preclinical candidate compounds were discovered, some of which continue in clinical trials for treatment of cancer [11, 17]. The very early work (virtual and fragment screening) was carried out within Vernalis before a collaboration was initiated with the group of Paul Workman at the Institute of Cancer Research. Subsequently, the project was partnered with Novartis, who were responsible for preclinical development and clinical trials while work at Vernalis continued on identification and optimisation of backup and follow-up compounds.

The structure of both the ATPase site and full-length Hsp90 has been determined recently and efforts continue to identify and characterise the roles of various co-chaperones [18]. Hsp90 belongs to the family of GHKL (Gyrase B, Hsp90, histidine kinase, MutL) ATPases and consists of three main domains. The N-terminal domain (hereafter Nt-Hsp90) is the most studied and contains an unusually shaped ATP binding cleft characterised by a left handed $\beta-\alpha-\beta$ (Bergerat) fold. ATP binding, hydrolysis and the subsequent release of ADP by the ATPase domain are crucial for Hsp90 function. The middle domain displays a large hydrophobic surface, implicated in stabilising the fold of client proteins but also supplies a core component to the ATPase site necessary for catalytic activity. Protein homodimer-isation occurs through the C-terminal domain. This domain has recently been shown to be important for autophosphorylation of the protein and contains a binding site for novobiocin.

In early 2002, we determined the crystal structure of Nt-Hsp90 from both the α- and β-forms [19]. There was no measurable difference in the inhibition of ATPase activity by the initial tool compound (PU3, compound **2** [20]) between
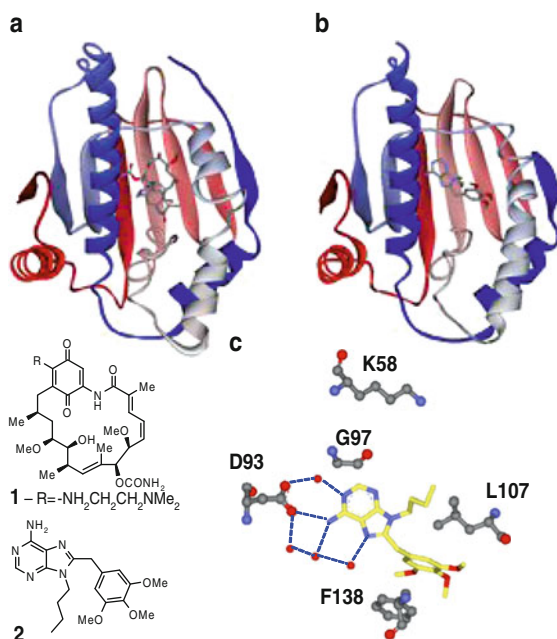
Fig. 1 Crystal structure of the N-terminal domain of human Hsp90 (Nt-Hsp90) in complex with various ligands showing (**a**) protein structure with 17-AAG bound (PDB code: 1OSF); and (**b**) protein structure and (**c**) details of active site and selected solvent molecules and hydrogen bonds (*dashed lines*) for PU3 bound (PDB code: 1UY6). The shaded box highlights residues 110—115 adopt a varied conformation in response to binding of different ligands. In (**c**) and subsequent figures of the Nt-Hsp90 active site, the residues K58, D93, L107, M98, F138 are shown together with selected water molecules

these isoforms, and the few amino acid differences between the proteins are some distance from the ATP binding site. The α-form was more amenable to structural studies and was therefore used for all subsequent studies. The key features of the structure are shown in Fig. 1 for two of the compounds [an ansamycin analogue, 17-AAG (**1a**) and the purine, PU3 (**2**)] that were available as potential tool compounds at the beginning of the project. PU3 was designed [20] based on the adenine core of ADP and the conformation seen in the 17-AAG structure. The subsequent crystal structures [19] showed that although the adenine moiety does adopt the same conformation as seen in the structure of Nt-Hsp90 with ADP [21], the rest of PU3 binds in a very different conformation to that predicted [20]. The crystal structures demonstrate that the key interactions in the binding site are with Asp93 (D93) and an essentially conserved set of water molecules, as shown in Fig. 1c. They also showed that the region containing residues 110–115 at the lip of the binding site can undergo conformational change under the influence of ligand binding. In the case of some ligands (as for PU3) this resulted in the generation of an additional binding site under a helical conformation, which we will call the PU3 conformation.

2 Virtual Screening

There is a considerable literature on the development and evaluation of computational methods for docking ligands into a molecular target [22]. Most docking software can reproduce (60–90% of the time) the correct binding pose (within a 2 Å RMSD margin) for docking a ligand back into the protein crystal structure from which it was taken (e.g. [23]). The success in cross-docking, that is, docking a ligand that is known to bind into a structure obtained either without a ligand bound, or with a different ligand bound, is in general lower (40–70% of the time). Although generating the correct pose is not, in general, the issue, it is the subtleties of the calculation of binding energies that make it difficult to recognise when the correct pose has been identified. Very small changes in the detailed conformation of the protein, or the positioning of a solvent molecule, can have a large impact on the calculated energy. There has been some exploration of whether docking against an ensemble of protein conformations can improve the successful identification of binding pose (see for example a study that included Nt-Hsp90 by Vernalis scientists, [24]). The methods are becoming sufficiently reliable that they can provide useful information on binding of known ligands.

Even more challenging is using such docking calculations to screen a large virtual library of compounds for their potential to bind to a particular target. Despite considerable efforts by many experienced developers over many years, "blind" use of such virtual screening can be barely better than random selection at identifying true hit compounds [23]. However, what most experienced practitioners have learnt is that the methods can be quite successful if additional information and understanding about the active site and selection of compounds is used. The virtual screening campaign against Hsp90 illustrates the importance of appropriate selection of binding site conformation (and solvent) and the careful postfiltering of virtual screening results.

2.1 Virtual Screening Protocols

rCat is our proprietary catalogue of 3.5 million compounds [25] and was assembled before databases such as ZINC became available [26]. A docking library of 0.7 million compounds was selected from rCat based on: (1) minimal drug-likeness (molecular mass 250–550 Da, and six or less rotatable bonds); (2) removal of reactive groups (a list of unstable chemical moieties was compiled based on chemical expertise, and substructural searches were performed to clean the docking library of molecules containing reactive groups); and (3) vendor delivery timelines. The program CORINA (version 2.63) was used to convert the docking library from two dimensions to three dimensions. The same program was used to generate multiple ring conformations; the internal energy threshold was set to 7 kJ/mol.

This allowed the treatment of cyclic groups as rigid bodies. The total number of docked conformers was 1.7 million.

Our analysis of the literature and initial in-house crystal structures (see Fig. 1) highlighted that the conformation of residues 110–115 can be altered significantly upon binding of PU3 [19]. Therefore, two structures of Hsp90 were used for docking: 1YET [27] and 1UY6 – an in-house structure of the PU3–Hsp90 complex [19]. In addition, it was clear that key water molecules (as seen in Fig. 1c) were important and these were included in the active site used for docking. All other water molecules and ligands were removed.

rDock was used for docking in its high-throughput mode [28]. A total of 4,300 compounds were selected using the 1YET cavity and 6,000 using the PU3 cavity, totalling approximately 9,000 non-redundant compounds from the initial screen. As discussed above, the next steps of postfiltering were key in selecting appropriate compounds.

Those compounds binding almost exclusively through polar or apolar interactions were removed to guarantee an adequate balance of these terms, in agreement with the composition of the targeted site. It was clear from the initial structures in the literature (and confirmed in the structures presented in this chapter) that the inhibitors form a hydrogen bond with one carboxylic oxygen of D93 and accept a second hydrogen bond from an interstitial water molecule. The binding poses not satisfying this donor–acceptor motif were discarded.

The remaining compounds were clustered in chemical families to assess diversity and the over-represented families purged. The top 1,000 remaining scorers were selected for purchase. Of those, 719 compounds were actually available and assayed.

Out of 719 compounds assayed, those that inhibited the ATPase activity [29] of Hsp90 by greater than 50% were selected for IC_{50} determination. A total of 13 compounds with $IC_{50} < 100$ μM (1.8% hit rate) and seven with $IC_{50} < 10$ μM (1.0% hit rate) were identified. More than 40% of the purchased compounds that were hits failed on quality control checks (QC) of purity and/or stability. Although many of these QC failures had similar chemotypes to validated hits, they were not considered further.

As mentioned above, subsequent to our initial virtual screen, Vernalis entered a collaboration with a group at the ICR in London. Interestingly, two of the virtual screening hits are closely related to a compound identified by a medium throughput screening (MTS) conducted by the ICR of some 60,000 compounds [30, 31] and show very similar activity.

2.2 Virtual Hits to Nt-Hsp90

Figure 2 shows the crystal structure of four of these 13 virtual screening hit compounds bound to Hsp90. The structures emphasise the importance of the structural water molecules in bridging between protein and ligand and in the conformational
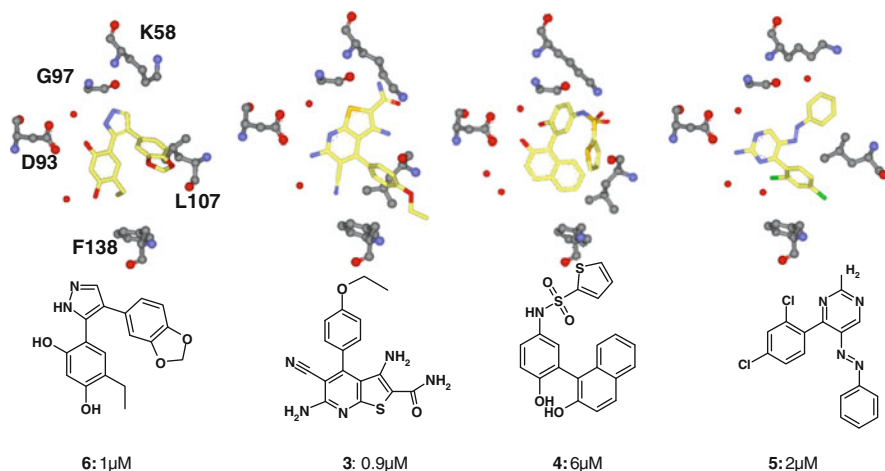
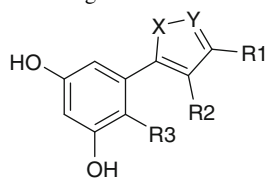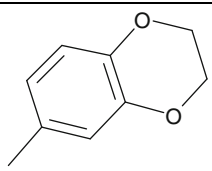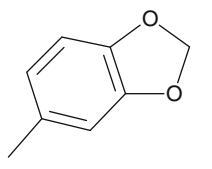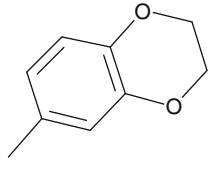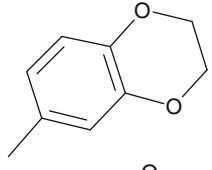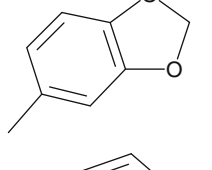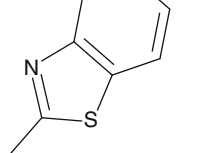**Fig. 2** Details of the crystal structures of the active site of Nt-Hsp90 (see legend to Fig. 1c) and selected solvent molecules for virtual screening hit compounds **3**, **4**, **5** and **6** binding to Nt-Hsp90. The affinities quoted are IC$_{50}$ values for binding in a fluorescence polarisation assay [32]

change seen in the helix. What was striking was the range of different chemotypes identified, with the one common feature being the hydrogen bond directly to D93. Compound **4** was a singleton, and an additional compound was identified of similar central scaffold to compound **3**. However, a range of resorcinol–pyrazole compounds were discovered by this free docking protocol of virtual screening, immediately providing some structure–activity relationship (SAR). The structures and inhibitor activity of some representatives are shown in Table 1. The structure of some of these compounds bound to Nt-Hsp90 were determined and allowed the activity to be rationalised in terms of specific interactions. Group R2 is directed towards the solvent, whereas changing from ethyl to propyl at position R1 made little difference in this early series. The allowed substituents at position R3 are constrained by the structure of the binding site.

This resorcinol–pyrazole series (compounds **6–11**) was rapidly established as the first series suitable to be taken forward for optimisation (see later). However, the other hits identified by virtual screening provided some novel templates. The phenol-naphthol (compound **4**) was not explored further [33]. The phenol makes the key interaction with Asp92; interestingly, this is the key interaction seen in the fragment that inspired the Astex clinical candidate [34]. Although compound **4** was identified by virtual screening, the details of the binding mode produced by docking depends on the protein model used (see [33] for details). This early docking study emphasised the importance of the details of the solvent structure for correct prediction – both presence/absence of particular molecules as well as the detailed position. The other two virtual screening hits (compounds **3** and **5**) highlighted in Fig. 2 provided ideas for the second lead optimisation series discussed in Sect. 6.
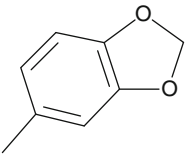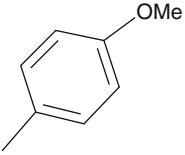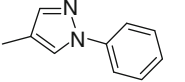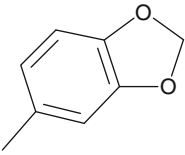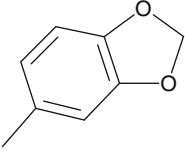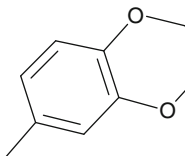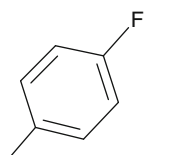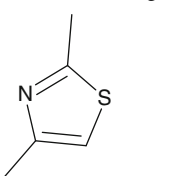
Table 1 Selected resorcinol-containing compounds identified as hits from free docking (virtual) screening or restrained docking calculations

Compound	R1	R2	R3	FP IC$_{50}$[a] (μM)
ICR compound **7** (X = N)	CH$_3$		–CH$_2$–CH$_3$	0.28
Pyrazoles found by free docking (X = N, Y = N)				
6	H		–CH$_2$–CH$_3$	0.8
8	H		–CH$_2$–CH$_3$	0.2
9	CH$_3$		–CH$_2$–CH$_2$–CH$_3$	3.3
10	CH$_3$		–CH$_2$–CH$_3$	0.3
11	CH$_3$		–CH$_2$–CH$_2$–CH$_3$	5

(*continued*)

Table 1 (continued)

Compound	R1	R2	R3	FP IC$_{50}$[a] (μM)
Pyrazoles found by focussed docking (X = N)				
21	H		–CH$_2$–CH$_2$–CH$_3$	0.4
22	H		–CH$_2$–CH$_3$	0.6
23	H		–CH$_2$–CH$_3$	1.8
24	–CO$_2$H		–CH$_2$–CH$_3$	0.3
25	–CF$_3$		–CH$_2$–CH$_3$	0.5
Isoxazoles (X = O, Y = N) found by focussed docking				
26	H		CH$_3$	0.3
27	H		CH$_3$	2.4
Isoxazoles (X = N, Y = O) found by focussed docking				
28	H		CF$_3$	2.8

[a]IC$_{50}$ values measured using a fluorescence polarisation assay

The other feature of the structure of Nt-Hsp90 that was emphasised during these early studies is the flexibility of the helix from residues 110 to 115. The conformational changes first seen on binding of PU3 (the PU3 conformation) could be rationalised through the interactions made by the tri-methoxy benzene ring to a hydrophobic pocket under a helical conformation (Fig. 1c). However, the conformation of these residues varied with other ligands in an unpredictable way, even between ligands soaked into the same batch of apo crystals. This suggests these different conformations of the ligands are of similar energy and that there are subtle changes in the crystal environment induced by ligand binding. These conformational changes are analysed further in Sect. 7.

It is appropriate at this stage in the discussion to comment on the different assays that were used to measure the inhibition of Hsp90 by various compounds. In the absence of co-chaperones, human Hsp90 exhibits weak ATPase activity [35]. The initial assay used to assess the affinity of the virtual screening hits for Hsp90 was a malachite green ATPase assay. This assay did not prove sufficiently robust or sensitive to support the medicinal chemistry program. Therefore, a fluorescence polarisation (FP) assay was developed using an initial fluorescently labelled resorcinol–pyrazole as a probe [32]. As the program progressed, new FP probes with greater affinity for Hsp90 were developed. Additional biophysical methods, such as surface plasmon resonance (SPR) or isothermal titration calorimetry (ITC) were used to assess and characterise inhibitor binding affinity. Although there may be apparent differences in the absolute binding affinity measured by these different types of assay, the relative ranking provides robust indication of changes in affinity.

3   Fragment Screening

During the late 1990s, Vernalis (as the precursor company, RiboTargets) had experimented with various NMR techniques for detecting binding of small ligands to RNA targets such as the ribosome (B. Davis, personal communication). During 2002, Vernalis built on this experience to establish a fragment screening capability. The initial components were a library of suitable compounds of molecular weight 110–250 and a set of NMR experiments able to detect the affinity (as low as 5 mM) for such small compounds. Initial trials highlighted the issue of nonspecific binding at the high concentrations (500 μM–1 mM) used. For these reasons, a competition step was introduced whereby a known ligand was added to the NMR samples and a hit classified as a fragment whose binding was displaced by the competitor.

A paper published in 2004 [36] described the development of the Vernalis fragment library from 2002–2004. The initial screen against Nt-Hsp90 used just the first (so-called SeeDs1) library of 790 compounds. Again, during 2003–2004, the range of NMR experiments used to detect binding was explored and improved (as described in [37]). However, for this first trial, only one-dimensional STD
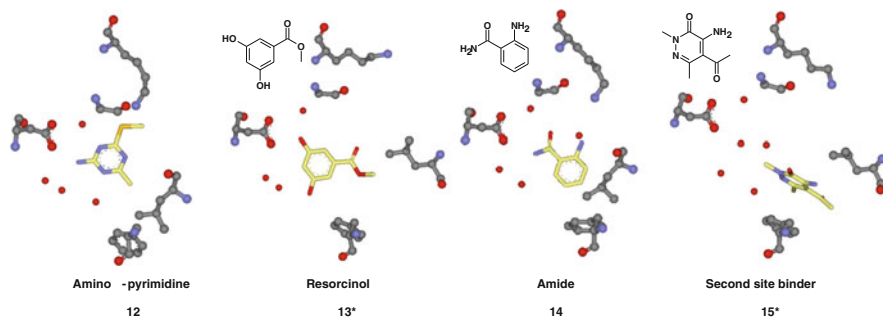
Fig. 3 Details of the crystal structures of the active site (see legend to Fig. 1c) and selected solvent molecules for fragment compounds **12**, **13***, **14** and **15*** binding to Nt-Hsp90. The * by a compound number indicates it came from the second round of fragment screening (see text)

experiments were used, with PU3 used as the competitor ligand. Although this compound only binds with low micromolar affinity, it was able to displace a reasonable number of fragments. The protein was present in the NMR tubes at a concentration of 10 μM, and the STD spectrum measured for fragments added as mixtures of two compounds per tube at a concentration of 1 mM for each compound and again after addition of 100 μM PU3. Any hits were re-screened as singletons at lower protein concentration (which increases the stringency of the binding experiment). Twenty-six hits were identified as competitive with PU3, for which 21 crystal structures were determined (these were fragments that gave crystal structures after no more then two attempts at soaking). The crystal structures of representative fragments bound to Nt-Hsp90 are shown in Fig. 3.

A subsequent re-screen of the target with a larger fragment library (1,350 compounds [36]) at a lower concentration and using a set of three NMR ligand-observed experiments (STD, Water-LOGSY, CPMG [37]) identified some 60 competitive fragment hits for the Nt-Hsp90 binding site, again using PU3 to confirm competition at the ATP binding site.

The discussion in the remainder of this paper will primarily use the fragments from the first screen of the SeeDs1 library. A few fragments will be used from the subsequent screen to make additional points and will be indicated by an * after the compound number.

One such analysis is to ask whether the fragment screen identified all of the resorcinol- or phenol-containing compounds. From the full screen of 1,350 fragments, a total of five resorcinols and four additional phenol-containing compounds were validated as hits. The full fragment library has a further four resorcinol-containing fragments that were not identified as hits and, for these, simple models demonstrate that the compound is elaborated in a way that precludes binding to the Nt-Hsp90 active site. On the other hand, more than 50 additional phenol-containing fragments were present in the library and were not hits. For some of these, the substitution pattern on the fragment precluded binding; for most, however, it was

probably the case that the compounds were not making sufficient additional interactions to augment the binding from the phenol.

4 Fragment Evolution: Linking Fragments

There are three principle approaches for evolving fragments to hit compounds: fragment linking, fragment growth and fragment merging (for summaries and more details of these approaches, see more comprehensive reviews such as those found in [38–40]).

The SAR by NMR approach pioneered by the group at Abbott was the first to demonstrate [41] how to generate high-affinity compounds from fragment hits. In their approach, heteronuclear single quantum coherence (HSQC) NMR spectra from ^{15}N-labelled protein are monitored to identify fragments that bind to the active site. The fragment is then optimised before a second round of screening (usually a library of slightly smaller compounds) is performed in the presence of the optimised first fragment. After some optimisation of this second fragment, a model of the structure is generated by NMR spectroscopic methods. This model is then used to guide chemical strategies to link the two fragments together and in this way generate more potent hit compounds. The Abbott group has reported series of such projects in which linking (sometimes requiring quite considerable chemical effort) has generated advanced lead compounds (e.g. [42–45]). However, there are only a few other reports of successful linking campaigns, with two examples being [46] and [47], probably because of the challenges. The approach requires two distinct binding pockets, tractable synthetic chemistry and the appropriate stereochemistry in the linker to allow the two fragments to maintain their original orientation and position in the final compound.

We noticed a number of so-called, second-site binders in our fragment screens, such as compounds **15*** (Fig. 3) and **17***, (Fig. 4c). The crystal structures showed that these fragments bound to Nt-Hsp90 in the PU3 conformation, interestingly induced in the crystal on soaking of the fragments into apo crystals (as seen by others, [34, 48]). We then determined a number of crystal structures of Nt-Hsp90 with two fragments bound simultaneously. Figure 4 summarises some of the results. The resorcinol **16** shows little change in position and orientation in the active site when soaked into crystals in the presence (Fig. 4b) and absence (Fig. 4a) of second site binder **17***. However, **17*** flips its orientation in the presence (Fig. 4b) and absence (Fig. 4c) of the resorcinol fragment. This same flipping of binding mode is seen in the crystal structure of second site binder **17*** and the amino-pyrimidine **18*** (Fig. 4d).

A team of the chemists working on the project spent some time developing ideas for linking to or exploiting the second site binder. However, the synthetic effort was considerable for most designed compounds and the success of other strategies (see Sects. 5 and 6) meant this linking idea was not pursued further. Interestingly, Evotec have recently disclosed compounds from linking two fragments that were remarkably similar to that shown in Fig. 4d [47].
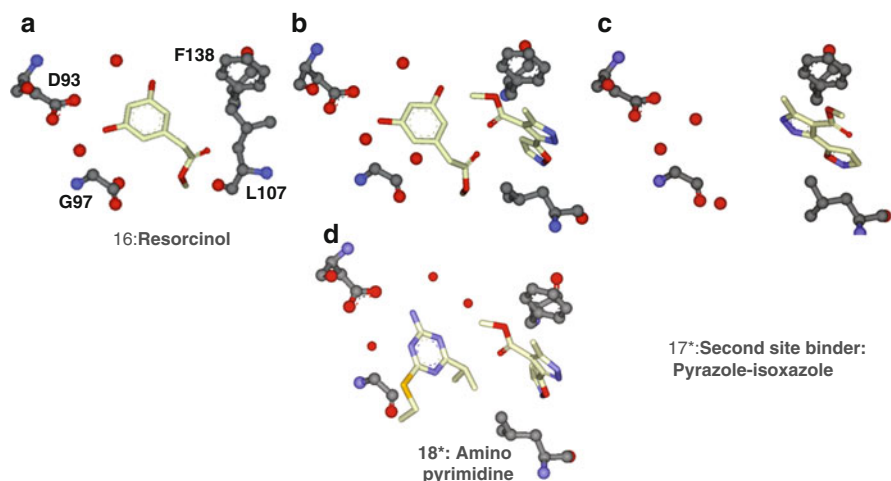
Fig. 4 Attempt at generating hits through a fragment linking strategy: Details of the crystal structures of the active site (see legend to Fig. 1c) and selected solvent molecules for the structures of (**a**) compound **16**, (**b**) compound **17***, (**c**) compounds **16** and **17*** and (**d**) compounds **18*** and **17*** bound to the active site of Nt-Hsp90

5 Fragment Evolution: Fragment Growth and the Discovery of AUY922

5.1 Preliminary Fragment-to-Hit Chemistry by Library Synthesis

An initial attempt was made to generate SAR and evolve a set of resorcinol fragment hits by library synthesis, whereby the selection of library components was guided by virtual construction and docking of potential compounds in the Nt-Hsp90 structure. Some results are shown in Fig. 5. It proved relatively straightforward to increase the affinity from low millimolar to low micromolar for a wide variety of substitutions. This increased the understanding of how the resorcinol template could withstand substitution. These compounds were not progressed further.

5.2 Focussed Docking to Evolve SeeDs

A particularly powerful way to explore the initial evolution of fragment hits is the so-called SAR by catalogue approach. The fragment, or substructure(s) of the fragment chosen through inspection of the crystal structure, is used to search databases of accessible compounds. For Hsp90, we used our virtual library of commercially available compounds (rCat, [25]). This was searched for compounds
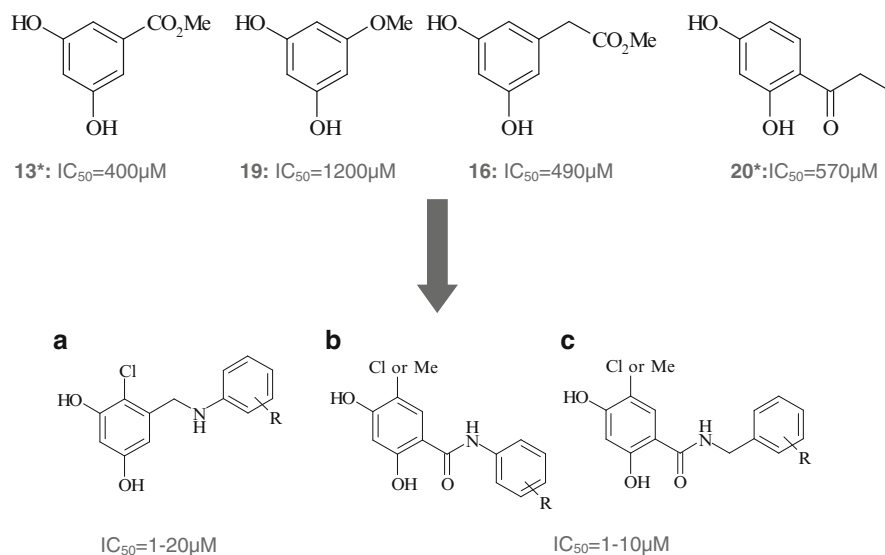
Fig. 5 Fragment growth by limited library synthesis. See text for details

containing the resorcinol substructure; these compounds were then docked into the active site of Nt-Hsp90 in the PU3 conformation, taken from the 1UY6 [19] crystal structure. Such an SAR by catalogue search was performed for a number of the fragments. The results for the resorcinol fragment search are reported here as they illustrate a number of important features of such an approach.

The search of the in-house virtual library identified more than 1,000 compounds containing the resorcinol substructure that were assessed for fit to the Nt-Hsp90 active site. Some 225 compounds were selected, of which 170 were delivered and assayed. Some of these compounds are shown in Table 1. Not only did this approach retrieve the same resorcinol **7** found by virtual and experimental screening [31] and encouraged us to generate additional SAR around the resorcinol–pyrazole template (compounds **21–25**), it also identified isoxazole (**26–28**) replacements for the pyrazole, which were explored in lead optimisation.

This example and Fig. 6a, b emphasise how fragments can be used to mine databases of available compounds as a very rapid way of going from the millimolar binding affinity seen for a fragment to the micromolar affinity expected for a hit or, in this case, lead compound that can be taken forward for optimisation.

5.3 Structure-Guided Design of AUY922

The lead compound **7** was identified by all the various hit identification techniques: virtual screening, experimental screening of a relatively small (<60,000) compound collection and SAR by catalogue based on the resorcinol motif seen in **20***.

a

13*:FP IC$_{50}$= ~1mM
Starting fragment

b

7: FP IC$_{50}$= 0.28μM
GI$_{50}$= 6μM

Hit from SAR by Catalog
(and docking and MTS)

c

29:FP IC$_{50}$= 0.009μM
GI$_{50}$= 0.014μM

Phase II Candidate
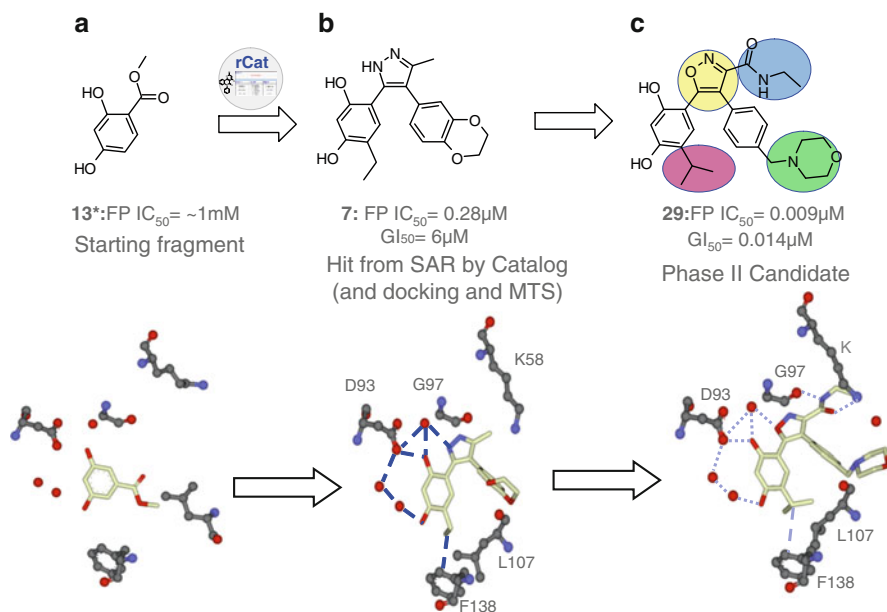
Fig. 6 Fragment growth to identify the clinical candidate AUY922. IC$_{50}$ was measured in a fluorescence polarisation assay; GI$_{50}$ values are for growth inhibition of HCT116 cells. See text for details

Compound **7** was chosen for optimisation. The issues with compound **7** were clear. The affinity needed to be improved from 300 nM to low nanomolar; the solubility needed to be improved from its initial < 12 μM; and it had very poor pharmacokinetics due to rapid glucoronidation. The details of the medicinal chemistry strategy in optimisation are described in various papers [11, 17, 30, 49]. The key features are illustrated in Fig. 6c. The addition of the amide group to the pyrazole (Fig. 6c, top right) gave a big leap in potency, whereby the amide completed a hydrogen bonding network between the main chain carbonyl of Gly97 (G97) and the side-chain terminal amine of Lys58 (K58). The crystal structure of **7** shows that the benzodioxane substituent is pointing out into solvent; the replacement by the methylmorpholino (Fig. 6c, bottom right) gave a big improvement in compound solubility. The substituent at the 5-position (Fig. 6c, bottom left) on the resorcinol ring was explored in the final series of compounds for the appropriate balance of biological properties. The isopropyl was eventually chosen as giving the best in vivo profile [10]. This is probably due to two effects: first, the isopropyl exploits a small hydrophobic region between Phe138 (F138) and Leu107 (L107); and second, the additional bulk will have some influence on the speed at which the compound is glucoronidated. Finally, the replacement of the pyrazole by isoxazole (Fig. 6c, top left, and suggested by the hits from SAR by catalogue) had a dramatic impact on the cell potency in certain cell lines. Subsequent measurement (by SPR) of the binding kinetics of paired compounds with a pyrazole or an isoxazole showed that this

increase in efficacy is due to a dramatic increase in the rate of dissociation of the compound from the target (J. Murray, unpublished results). This increase in so-called residence time has been noted by others as an important attribute for increased efficacy for some targets [50]. In the case of Hsp90, the combination of very slow off-rate binding properties of the compound, the high concentration of Hsp90 in tumour cells and the likelihood of a particular co-chaperone environment within the cancer cells [51] have an impact on drug binding. This leads to rapid absorption and long-term retention of the compound in tumour cells in vivo and to rapid clearance of the compound from the plasma and non-tumour tissues. Compound **29** (AUY922) is currently undergoing Phase II clinical trials in multiple myeloma, breast, lung and gastric cancers by Novartis.

6 Fragment Evolution: Merging Fragments and the Discovery of BEP800

The third strategy for exploiting fragment hits is to combine the information about scaffolds and interactions seen in crystal structures with the information available from existing compounds and virtual screening hits. Examples of this so-called "merging" approach exist for kinases [52] and the Hsp90 preclinical candidate BEP800. A paper from 2009 describes the medicinal chemistry program that generated this orally bioavailable compound [17]. The essential features of the structure-guided design are summarised in Fig. 7. The crystal structures of
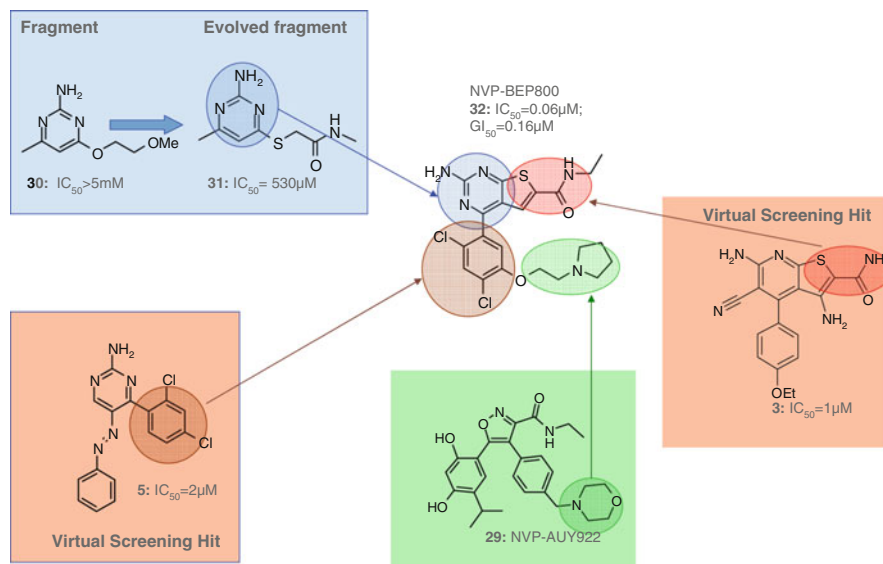
Fig. 7 Fragment merging to identify the preclinical candidate BEP800. See text for details

fragments such as **30** bound to Nt-Hsp90 were used to identify SAR by catalogue hits such as **31**, which explored the SAR around the fragments. In addition, the crystal structures of virtual screening hits such as **3** and **5** bound to Nt-Hsp90 were considered. An overlay of these various compounds correctly suggested that a thienopyrimidine scaffold (merging of the top left and top right features in Fig. 7) would bind to the active site of Nt-Hsp90 and provide accessible chemistry and vectors for further optimisation. Further comparison of the virtual screening compounds suggested the addition of a di-chloro benzene (Fig. 7, bottom left feature) would enhance activity. Finally, comparison with the crystal structure of the clinical candidate AUY922 (**29**) bound to NtHsp90, showed where to place a suitable solubiliser. The resulting compound **32** showed efficacy when dosed orally and was taken forward for preclinical development [53].

7 Analysis of Water Position Movement in the Fragments

The N-terminus of Hsp90 crystallises readily in a crystal form that is suitable both for co-crystallisation and soaking experiments. To increase throughput, most of the nearly 300 crystal structures determined at Vernalis in the Hsp90 project were by soaking of ligands into preformed apo crystals, with occasional checks by co-crystallisation. It is arguable that most of these crystal structures did not give additional information and that the essential features of many of the structures could have been predicted by molecular modelling. However, as it was possible to determine the structures, it was useful to check the subtleties of side chain flexibility and, in particular, solvent position whenever possible. Some interesting general features were observed.

Figure 8 shows details of the active site of nine crystal structures of Nt-Hsp90. A number of interesting phenomena can be observed:

- There are three water molecules at the protein–ligand interface (shown as spheres), which are preserved across all crystal structures. There are only minor movements of these water molecules, mostly within experimental error and to accommodate the binding of ligands.
- There are two water molecules at the protein–ligand interface (shown as darker coloured spheres in Fig. 8b, h and i), which are present in the apo structure and some structures with fragments, but which are displaced in other structures.
- L107 (highlighted with circles) is in one of two positions, depending on the conformation of the helix at residues 110–115 (as discussed in Sect. 2). The formation of the folded helix is stabilised by ligands such as PU3 (Fig. 6i). This conformation for L107 (Fig. 8a, d, and i) can also be induced by fragments, as seen for the second site binder, compound **17*** (Fig. 6d) but it can also be formed when there is no compound binding into the pocket, as for compound **13*** (Fig. 6a). However, the alternative conformation is seen in the apo structure
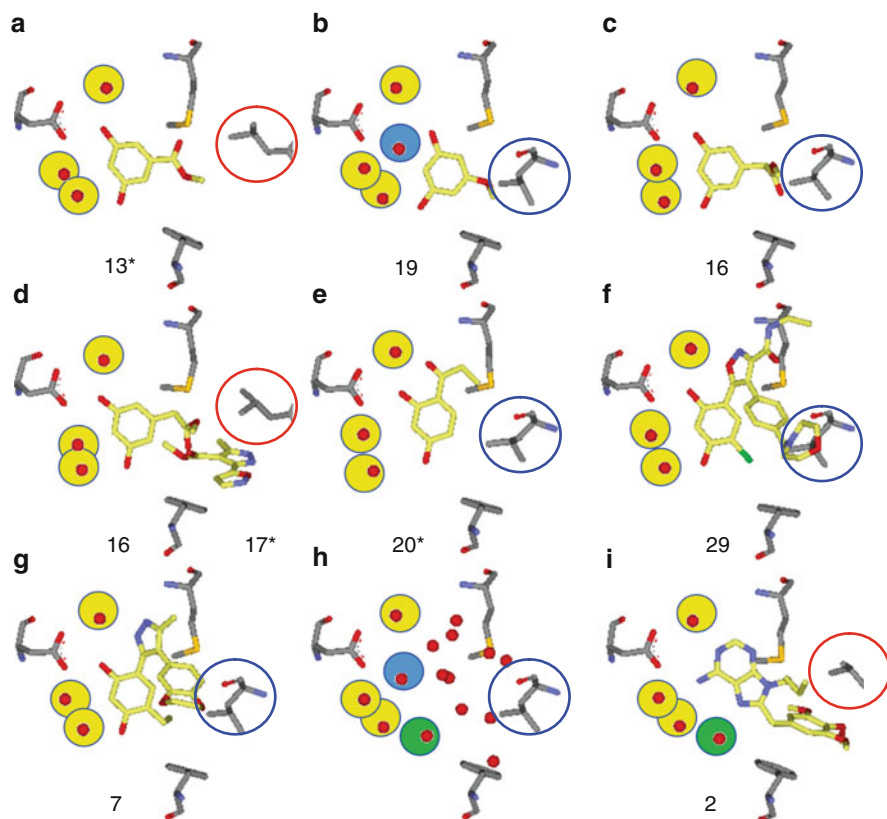
Fig. 8 Detail of the crystal structures of the Nt-Hsp90 active site (see legend to Fig. 1c) and selected solvent molecules and ligands in crystal structures of various individual fragments (**a–c**, **e**); a double fragment soak (**d**); the initial lead compound (**g**) that led to AUY922 (**f**); apo Nt-Hsp90 (**h**); and PU3 (**i**). *Circles* indicate the two positions of L107; *spheres* indicate water molecules at the protein–ligand interface. See text for details

   (Fig. 6h) when compound **19** binds (Fig. 6b) and is preserved when the lead compound **7** binds (Fig. 6g).

- Although the binding mode of resorcinol-containing compounds is strongly preserved (as in Fig. 6a), differences can be seen, albeit rarely, as for compound **19** (Fig. 6b).

8 Published Hsp90 Inhibitors

The ATPase site of Hsp90 has proved an exciting target for the development of novel, small molecule inhibitors for the treatment of cancer. A large number of inhibitors have now been described, many of which are currently in Phase I and II
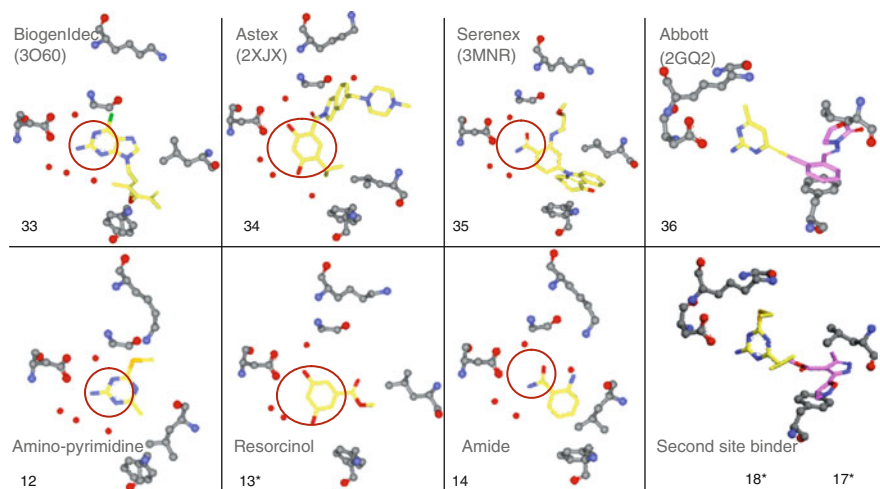
Fig. 9 Fragments recapitulate the central features (*circled*) of published Hsp90 clinical candidates. Detail of the crystal structures of Nt-Hsp90 active site (see legend to Fig. 1c) and selected solvent molecules and ligands in crystal structures of fragments (*bottom panel* of four structures) and clinical candidates (*top panel* of four structures)

trials with others in preclinical development and poised to enter the clinic. For many of these inhibitors, the crystal structure of the inhibitor bound to Hsp90 has been described. It is interesting to review to what extent the chemical entities identified in our fragment screens conducted in 2002–2003 are recapitulated in the compounds subsequently developed, using both structure/fragment based methods and conventional medicinal chemistry. A more thorough review of this topic has been published recently [57] but Fig. 9 summarises some of the key results.

The amino-pyrimidine motif (essentially a substructure within purine seen in ADP) is echoed in the Vernalis preclinical candidate, BEP800 and in the compound BIIB021 now being developed by Biogen Idec [12] (**33**). The resorcinol motif is found in both the Vernalis/Novartis Phase II candidate, AUY922 and the Astex compound AT13387 [15] (**34**). The benzamide motif is seen in a number of the Pfizer/Serenex compounds [54] (**35**). Finally, there have been two published reports of successful fragment linking approaches which have combined a core, purine replacement scaffold with a compound that binds in the methoxy-benzene second site pocket. Both examples [from Evotec [47] and from Abbott [42] (**36**)] have generated extended compounds where the affinity is greater than an individual fragment, but not the full additivity of energy expected from optimal combination. To date, there has been no report of either of these compounds progressing further in optimisation or clinical trials.

9   Concluding Remarks

Many different companies and organisations have successfully generated Hsp90 inhibitors over the past 10 years. The first clinical candidates evolved from the natural product ansamycins (17-AAG and IPI-504), whereas the Vernalis project was the first to identify and publish novel inhibitors discovered using structure- and fragment-based discovery. Subsequent to this, many similar structurally inspired Hsp90 inhibitors have been published. As well as providing a promising range of clinical candidates, Hsp90 as a target has proved particularly tractable for structure- and fragment-based discovery and it has been a useful testbed for the development and characterisation of new techniques and technologies [55]. Perhaps more importantly, the clinical trials of several Hsp90 inhibitors continue and there is a growing understanding that compounds that affect multiple signalling pathways may prove useful in cancer therapy.

References

1. Dutta D et al (2009) The molecular chaperone heat shock protein-90 positively regulates rotavirus infectionx. Virology 391(2):325–333
2. Dollins DE et al (2007) Structures of GRP94-nucleotide complexes reveal mechanistic differences between the hsp90 chaperones. Mol Cell 28(1):41–56
3. Hanahan D, Weinberg RA (2000) The hallmarks of cancer. Cell 100(1):57–70
4. Whitesell L, Lindquist SL (2005) HSP90 and the chaperoning of cancer. Nat Rev Cancer 5(10):761–772
5. Trepel J et al (2010) Targeting the dynamic HSP90 complex in cancer. Nat Rev Cancer 10(8):537–549
6. Supko JG et al (1995) Preclinical pharmacologic evaluation of geldanamycin as an antitumor agent. Cancer Chemother Pharmacol 36:305–315
7. Soga S, Shiotsu Y, Akinaga S, Sharma SV (2003) Development of radicicol analogues. Curr Cancer Drug Targets 3(5):359–369
8. Gao Z, Garcia-Echeverria C, Jensen MR (2010) Hsp90 inhibitors: clinical development and future opportunities in oncology therapy. Curr Opin Drug Discov Dev 13(2):193–202
9. Taldone T et al (2008) Targeting Hsp90: small-molecule inhibitors and their clinical development. Curr Opin Pharmacol 8(4):370–374

10. Eccles SA et al (2008) NVP-AUY922: a novel heat shock protein 90 inhibitor active against xenograft tumor growth, angiogenesis, and metastasis. Cancer Res 68(8):2850–2860

11. Brough PA et al (2008) 4,5-Diarylisoxazole Hsp90 chaperone inhibitors: potential therapeutic agents for the treatment of cancer. J Med Chem 51(2):196–218

12. Lundgren K et al (2009) BIIB021, an orally available, fully synthetic small-molecule inhibitor of the heat shock protein Hsp90. Mol Cancer Ther 8(4):921–929

13. Wang Y et al (2010) STA-9090, a small-molecule Hsp90 inhibitor for the potential treatment of cancer. Curr Opin Investig Drugs 11(12):1466–1476

14. Huang KH et al (2009) Discovery of novel 2-aminobenzamide inhibitors of heat shock protein 90 as potent, selective and orally active antitumor agents. J Med Chem 52(14):4288–4305

15. Woodhead AJ et al (2010) Discovery of (2,4-dihydroxy-5-isopropylphenyl)-[5-(4-methylpiperazin-1-ylmethyl)-1,3-di hydroisoindol-2-yl]methanone (AT13387), a novel inhibitor of the molecular chaperone Hsp90 by fragment based drug design. J Med Chem 53(16): 5956–5969

16. Biamonte MA et al (2010) Heat shock protein 90: inhibitors in clinical trials. J Med Chem 53(1):3–17

17. Brough PA et al (2009) Combining hit identification strategies: fragment-based and in silico approaches to orally active 2-aminothieno[2,3-d]pyrimidine inhibitors of the Hsp90 molecular chaperone. J Med Chem 52(15):4794–4809

18. Pearl LH, Prodromou C (2006) Structure and mechanism of the Hsp90 molecular chaperone machinery. Annu Rev Biochem 75:271–294

19. Wright L et al (2004) Structure-activity relationships in purine-based inhibitor binding to HSP90 isoforms. Chem Biol 11(6):775–785

20. Chiosis G et al (2003) Development of purine-scaffold small molecule inhibitors of Hsp90. Curr Cancer Drug Targets 3(5):371–376

21. Obermann WM et al (1998) In vivo function of Hsp90 is dependent on ATP binding and ATP hydrolysis. J Cell Biol 143(4):901–910

22. Leach AR, Shoichet BK, Peishoff CE (2006) Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. J Med Chem 49(20):5851–5855

23. Warren GL et al (2006) A critical assessment of docking programs and scoring functions. J Med Chem 49(20):5912–5931

24. Barril X, Morley SD (2005) Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. J Med Chem 48(13):4432–4443

25. Baurin N et al (2004) Drug-like annotation and duplicate analysis of a 23-supplier chemical database totalling 2.7 million compounds. J Chem Inf Comput Sci 44(2):643–651

26. Irwin JJ, Shoichet BK (2005) ZINC–a free database of commercially available compounds for virtual screening. J Chem Inf Model 45(1):177–182

27. Stebbins CE et al (1997) Crystal structure of an Hsp90-geldanamycin complex: targeting of a protein chaperone by an antitumor agent. Cell 89(2):239–250

28. Morley SD, Afshar M (2004) Validation of an empirical RNA-ligand scoring function for fast flexible docking using Ribodock. J Comput Aided Mol Des 18(3):189–208

29. Aherne W et al (2003) Assays for HSP90 and inhibitors. Methods Mol Med 85:149–161

30. Brough PA et al (2005) 3-(5-Chloro-2,4-dihydroxyphenyl)-pyrazole-4-carboxamides as inhibitors of the Hsp90 molecular chaperone. Bioorg Med Chem Lett 15(23):5197–5201

31. Cheung KM et al (2005) The identification, synthesis, protein crystal structure and in vitro biochemical evaluation of a new 3,4-diarylpyrazole class of Hsp90 inhibitors. Bioorg Med Chem Lett 15(14):3338–3343

32. Howes R et al (2006) A fluorescence polarization assay for inhibitors of Hsp90. Anal Biochem 350(2):202–213

33. Barril X et al (2005) Structure-based discovery of a new class of Hsp90 inhibitors. Bioorg Med Chem Lett 15(23):5187–5191

34. Murray CW et al (2010) Fragment-based drug discovery applied to Hsp90. Discovery of two lead series with high ligand efficiency. J Med Chem 53(16):5942–5955

35. Panaretou B et al (2002) Activation of the ATPase activity of hsp90 by the stress-regulated cochaperone aha1. Mol Cell 10(6):1307–1318
36. Baurin N et al (2004) Design and characterization of libraries of molecular fragments for use in NMR screening against protein targets. J Chem Inf Comput Sci 44(6):2157–2166
37. Hubbard RE et al (2007) The SeeDs approach: integrating fragments into drug discovery. Curr Top Med Chem 7(16):1568–1581
38. Congreve M et al (2008) Recent developments in fragment-based drug discovery. J Med Chem 51(13):3661–3680
39. Fischer M, Hubbard RE (2009) Fragment-based ligand discovery. Mol Interv 9(1):22–30
40. Schulz MN, Hubbard RE (2009) Recent progress in fragment-based lead discovery. Curr Opin Pharmacol 9(5):615–621
41. Shuker SB et al (1996) Discovering high-affinity ligands for proteins: SAR by NMR. Science 274(5292):1531–1534
42. Huth JR et al (2007) Discovery and design of novel HSP90 inhibitors using multiple fragment-based design strategies. Chem Biol Drug Des 70(1):1–12
43. Oltersdorf T et al (2005) An inhibitor of Bcl-2 family proteins induces regression of solid tumours. Nature 435(7042):677–681
44. Hajduk PJ (2006) SAR by NMR: putting the pieces together. Mol Interv 6(5):266–272
45. Hajduk PJ, Greer J (2007) A decade of fragment-based drug design: strategic advances and lessons learned. Nat Rev Drug Discov 6(3):211–219
46. Howard N et al (2006) Application of fragment screening and fragment linking to the discovery of novel thrombin inhibitors. J Med Chem 49(4):1346–1355
47. Barker JJ et al (2010) Discovery of a novel Hsp90 inhibitor by fragment linking. ChemMedChem 5(10):1697–1700
48. Barker JJ et al (2009) Fragment-based identification of Hsp90 inhibitors. ChemMedChem 4 (6):963–966
49. Dymock BW et al (2005) Novel, potent small-molecule inhibitors of the molecular chaperone Hsp90 discovered through structure-based design. J Med Chem 48(13):4212–4215
50. Copeland RA, Pompliano DL, Meek TD (2006) Drug-target residence time and its implications for lead optimization. Nat Rev Drug Discov 5(9):730–739
51. Chiosis G, Neckers L (2006) Tumor selectivity of Hsp90 inhibitors: the explanation remains elusive. ACS Chem Biol 1(5):279–284
52. Hubbard RE (2008) Fragment approaches in structure-based drug discovery. J Synchrotron Radiat 15(Pt 3):227–230
53. Massey AJ et al (2010) Preclinical antitumor activity of the orally available heat shock protein 90 inhibitor NVP-BEP800. Mol Cancer Ther 9(4):906–919
54. Fadden P et al (2010) Application of chemoproteomics to drug discovery: identification of a clinical candidate targeting hsp90. Chem Biol 17(7):686–694
55. Hubbard RE, Murray JB (2011) Experiences in fragment-based lead discovery. Methods Enzymol 493:509–531
56. Dymock B et al (2004) Adenine derived inhibitors of the molecular chaperone HSP90-SAR explained through multiple X-ray structures. Bioorg Med Chem Lett 14(2):325–328
57. Roughley SD, Hubbard RE (2011) How well can fragments explore accessed chemical space? A case study from Heat Shock Protein 90. J Med Chem. DOI: 10.1021/jm200350g

Combining NMR and X-ray Crystallography in Fragment-Based Drug Discovery: Discovery of Highly Potent and Selective BACE-1 Inhibitors

Daniel F. Wyss, Yu -Sen Wang, Hugh L. Eaton, Corey Strickland, Johannes H. Voigt, Zhaoning Zhu, and Andrew W. Stamford

Abstract Fragment-based drug discovery (FBDD) has become increasingly popular over the last decade. We review here how we have used highly structure-driven fragment-based approaches to complement more traditional lead discovery to tackle high priority targets and those struggling for leads. Combining biomolecular nuclear magnetic resonance (NMR), X-ray crystallography, and molecular modeling with structure-assisted chemistry and innovative biology as an integrated approach for FBDD can solve very difficult problems, as illustrated in this chapter. Here, a successful FBDD campaign is described that has allowed the development of a clinical candidate for BACE-1, a challenging CNS drug target. Crucial to this achievement were the initial identification of a ligand-efficient isothiourea fragment through target-based NMR screening and the determination of its X-ray crystal structure in complex with BACE-1, which revealed an extensive H-bond network with the two active site aspartate residues. This detailed 3D structural information then enabled the design and validation of novel, chemically stable and accessible heterocyclic acylguanidines as aspartic acid protease inhibitor cores. Structure-assisted fragment hit-to-lead optimization yielded iminoheterocyclic BACE-1 inhibitors that possess desirable molecular properties as potential therapeutic agents to test the amyloid hypothesis of Alzheimer's disease in a clinical setting.

Keywords Alzheimer's disease· Aspartic acid protease· BACE-1· Fragment-based drug discovery· Structure-based design

D.F. Wyss (✉), Y.-S. Wang, H.L. Eaton, C. Strickland, J.H. Voigt, Z. Zhu, and A.W. Stamford
Global Structural Chemistry and Global Chemistry, Merck Research Laboratories, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA
e-mail: daniel.wyss@merck.com

Contents

Abbreviations

AD   Alzheimer's disease
ADMET   Absorption distribution, metabolism, excretion, and toxicity
APP   Amyloid precursor protein
Aβ   Amyloid beta peptides ranging from 37 to 42 amino acids in length
BACE-1   β-site APP cleaving enzyme
cLogP   Computed partition coefficient of a compound
CNS   Central nervous system
c-STD   Competition saturation transfer difference
FBDD   Fragment-based drug discovery
FBS   Fragment-based screening
FQ   Fit quality
HCS   High concentration functional screening
HSQC   Heteronuclear single quantum coherence
HTS   High-throughput screening
IC_{50} Half maximal inhibitory concentration
ITC   Isothermal calorimetry
K_D Equilibrium dissociation constant
K_I Equilibrium inhibition constant
LE   Ligand efficiency
LLE   Ligand lipophilicity efficiency
MW   Molecular weight
NMR   Nuclear magnetic resonance
PK   Pharmacokinetics
SAR   Structure–activity relationship
SPR   Surface plasmon resonance
STD   Saturation transfer difference

1   Introduction

Fragment-based drug discovery (FBDD) is an emerging field in which much lower molecular weight (MW) compounds are screened relative to those in high-through-put screening (HTS) campaigns [1–15]. In theory, fragment-based methods offer the possibility of identifying novel leads with improved pharmaceutical properties and the prospect of tackling less tractable drug targets, and the rationale behind these fragment-based strategies makes intuitive sense. However, optimization of weak-binding fragments into potent leads can be challenging, and fragment-based lead discovery can be difficult in practice. Nevertheless, FBDD has become increasingly popular over the last decade in both the pharmaceutical industry and academia [6]. Both the discovery and advancement of fragment hits are areas of intense research. Although there is still much work to be done to fully exploit the potential of this approach, the increasing number of successful applications that have appeared in the literature [1–15], including the first examples of clinical drug candidates [6, 9, 11] originating from this approach, strongly suggest its viability.

Advantages of fragment-based screening (FBS) over HTS are, first, more efficient sampling due to the smaller chemical space of fragment-sized compounds [16, 17] and, second, a higher probability of fragments possessing good complementarity with the target [18]. Since fragment-based hits are typically weak inhibitors and/or binders (half maximal inhibitory concentration (IC_{50}) and/or the equilibrium dissociation constant (K_D) is in the micromolar to millimolar range) due to their low MW, they need to be screened at higher concentrations using suitable detection techniques that can reliably detect weakly interacting compounds, e.g., nuclear magnetic resonance (NMR), surface plasmon resonance (SPR), high concentration functional screening (HCS), or X-ray crystallography. All in all, FBS leads to higher hit rates, and only relatively low numbers of compounds (thousands) need to be screened to identify interesting hits [7], even against challenging targets [12, 19]. However, fragment hits have lower affinities towards the target. As a consequence, more effort has to be spent on optimization to obtain lead compounds with an acceptable affinity and, arguably, structural biology may play a crucial role in accomplishing this goal efficiently [12].

Although fragment hits are simpler, less functionalized compounds [20] than HTS hits, with correspondingly lower potencies, they typically possess good ligand efficiency (LE) [21–28] and ligand lipophilicity efficiency (LLE) [29, 30], especially after some initial analoging or exploratory elaboration. Fragments are therefore highly suitable for optimization into clinical candidates with good drug-like properties. This means that the number of atoms involved in the desired interaction with the drug target is usually high for such fragment hits. Typical HTS hits, on the other hand, tend to be larger and, although having higher potency, contain portions in the molecule that are not directly involved in the desired interaction with the drug target. Therefore, the hit-to-lead optimization process is fundamentally different between fragment hits and typical hits from HTS. Fragment hits need to be extended into nearby binding pockets by increasing their MW to gain potency,

whereas the potency of HTS hits often need to be increased without a significant increase of the MW of the initial hit [3]. Strategies have been proposed to guide and evaluate the fragment hit-to-lead optimization process [31–34]. These strategies aim at the efficient optimization of fragment hits while maintaining their generally good physicochemical properties. A recent review suggests, however, that typical medicinal chemistry approaches for lead optimization may fail at accomplishing this task [35], and a larger focus on enthalpy-driven lead optimization may be required [35, 36]. Nevertheless, less complex, polar, low MW hits should serve as better starting points for optimization [37–39] if unfavorable property shifts can be avoided during fragment hit-to-lead optimization.

In our laboratory, we have used a highly structure-driven, iterative FBDD approach composed of fragment-based NMR screening, X-ray crystallography, target-based NMR, computational chemistry, and structure-assisted chemistry. We have thus focused on targets that would be amenable to such a structure-based drug discovery (SBDD) approach, initiating protein production for both NMR and X-ray crystallography early in a project. To focus resources and to maximize impact we have applied this FBDD approach strategically to early targets, high priority targets, and those struggling for leads. Since exploratory chemistry is required for fragment hit-to-lead progression, we also paid special attention to prioritize those internal projects for which chemistry resources would be available to follow up attractive fragment hits. Further emphasis was then given to those fragment hits for which 3D structural data was available to support efficient fragment hit-to-lead progression. As a result, for 73% of the FBDD targets we have followed up fragment-based NMR screening hits through exploratory chemistry and generated 3D structural data of fragment hits when bound to the drug target. This approach has yielded valid lead series in the submicromolar potency range in about one third of those projects.

In this chapter, we first discuss fragment-based NMR screening, then suggest how to progress fragment hits into valid lead series, and finally describe a successful FBDD campaign that yielded a clinical candidate for BACE-1.

2 Fragment-Based NMR Hit Identification

Different FBS techniques have been developed and applied successfully for FBDD, as well documented in the literature (e.g., [15, 40–42]). NMR methods are among the most widely used FBS techniques [40] because they can provide useful information throughout a FBDD campaign. Versatile NMR methods are available to study the interaction of a ligand with its drug target. Such methods can be used for fragment-based NMR screening, the subsequent progression of fragment hits into leads based on structure–activity relationship (SAR) and structural information, and to support different stages in the lead generation process, ranging from hit characterization early in the process to late-stage lead optimization. Techniques can be broadly categorized into target- versus ligand-based NMR methods, depending on

whether signals from the drug target or the ligand are detected to characterize the intermolecular interaction. Each of these methods has advantages and limitations and can provide information about the ligand–target interaction at various levels of detail, including the determination of ligand affinities and potencies or their binding site and binding mode when bound to the drug target. NMR experiments can be selected to fit the target size and type, the program status, and the resources that are available. Therefore, different NMR screening and follow-up strategies may be selected for different FBDD campaigns.

2.1   NMR Screening Methods

Target-detected NMR methods (Fig. 1a) have the distinct advantage that they reveal structural information about the ligand binding site and its binding mode with the drug target, can detect site-specific ligand binding over a virtually unlimited affinity range, are very robust and reliable, and can be used to derive ligand affinities for weak fragment hits that are in fast exchange on the NMR time scale ($K_D > {\sim}10\ \mu M$) or for submicromolar hits when combined in a competition format (Table 1). However, they require large amounts of isotope-labeled drug target, necessitating expression of the protein in a host (typically *Escherichia coli*) that allows high expression yields ($> {\sim}1$ mg/L) and cost-effective isotope-labeling, and also require knowledge of the 3D structure of the drug target and NMR assignments (or at least a map) of the active site residues to reveal active site binders. Therefore, target-detected NMR approaches are limited to a subset of drug targets (MW $<$ 40–60 kDa) that give quality NMR spectra and do not aggregate at relatively high concentrations (${\sim}25$–$80\ \mu M$) in an aqueous NMR buffer.

Target-detected NMR screens monitor chemical shift perturbations in the heteronuclear single quantum coherence (HSQC) spectrum of an isotope-labeled protein as a small molecule (or mixture of small molecules) is added [43]. The most commonly used labeling scheme is to uniformly isotope-label the protein with ^{15}N. The HSQC spectrum of a uniformly ^{15}N-labeled protein contains a resonance for almost every amide N–H pair in the protein, and if these resonances have previously been assigned to the primary sequence of the protein, the binding site of the small molecule can be localized to several residues in the protein. If resonance assignments are not available, but there are reference compounds that are known to bind to the target, these reference compounds can be used to "map" residues in the binding site. If these same residues are perturbed during a fragment screen, it is likely that the screened molecule binds at the same site as the reference molecule. Even if no reference compound is available, the pattern of perturbed residues can be used to "bin" small molecules into potentially overlapping binding regions. Finally, for small molecules that are in fast exchange on the chemical shift time-scale, an NMR-K_D can be determined by titrating the protein with the small molecule and monitoring the magnitude of the chemical shift perturbations as the concentration of small molecule increases. NMR-K_D determination is particularly
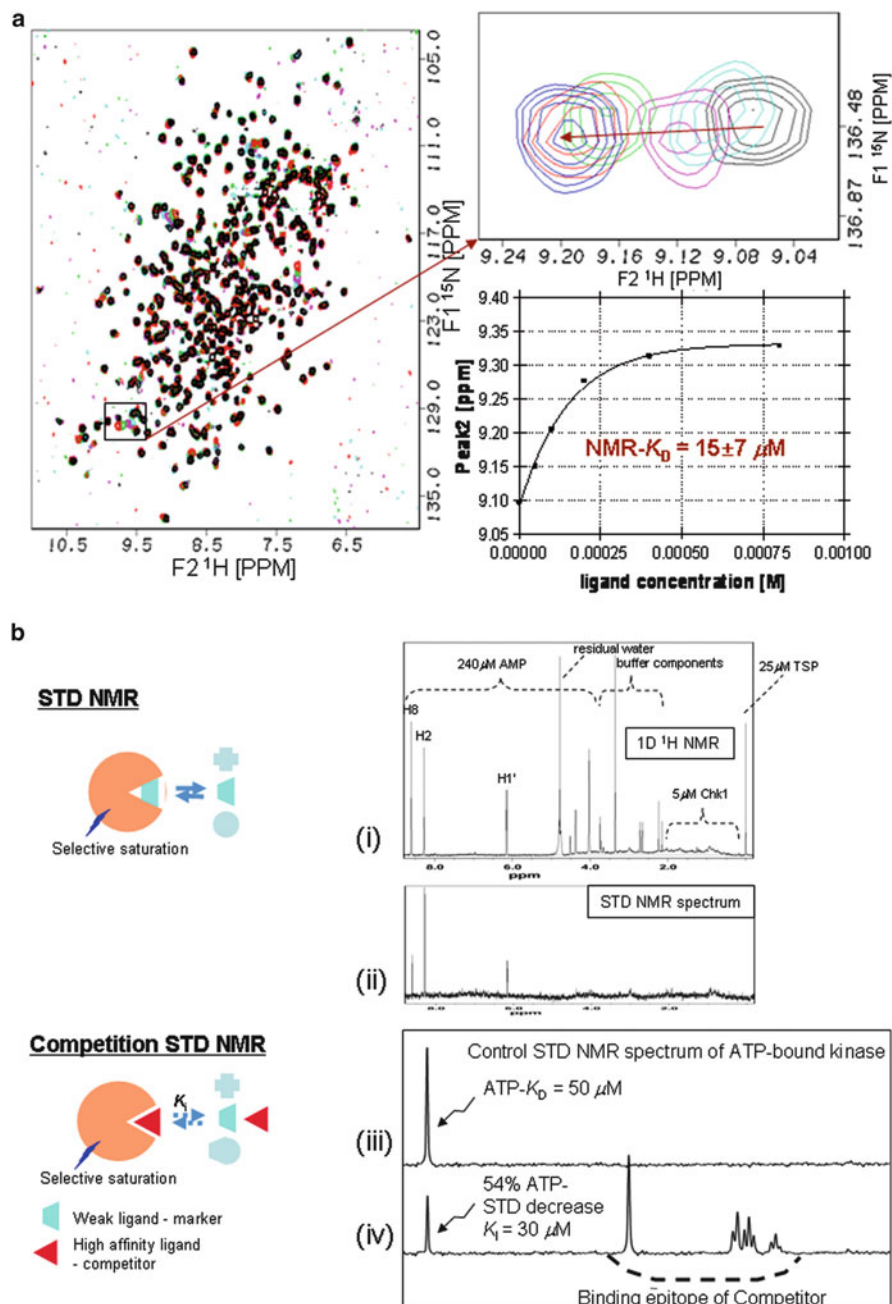
Fig. 1 NMR tools to support fragment hit identification and progression. Lead identification and optimization can broadly be categorized into target- versus ligand-detected methods depending on whether signals from the target or the ligand are detected to monitor binding. (**a**) Target-detected

useful when functional assays for a target have not been developed or are problematic for detecting weak fragment hits.

HSQC of uniformly ^{15}N-labeled protein can work well up to about 40–60 kDa. For protein targets larger than this, spectral overlap becomes a major problem, and methods that simplify the spectrum and improve the signal-to-noise ratio are needed. HSQC of proteins in which methyl groups are labeled with ^{13}C has been used to simplify spectra while still providing good coverage of the target [44]. This isotope-labeling scheme also has the advantage of yielding a favorable threefold sensitivity increase. In order to further simplify the HSQC spectrum of a large protein, amino-acid-type-selective (AATS) labeling can be used with either ^{15}N-labeling or ^{13}C-labeling of methyl carbons. In AATS labeling, the labeling is confined to either a single amino acid type (i.e., Phe) or a small group of amino acids types (i.e., Ile, Leu, Val). Choice of which amino acid types to label is based on the presence of an amino acid type in the binding site (if the binding site is known) and/or the distribution of the amino acid in the primary sequence of the protein. Not every protein is amenable to the labeling schemes required for target-based fragment screens or may not produce quality NMR spectral data. In these cases, ligand-based fragment screens may be employed.

Ligand-detected NMR methods (Fig. 1b, Table 1) can be applied much more broadly than target-detected fragment screens because they require about 1–10% the amount of drug target, do not require isotope-labeling, and have no upper MW size limitation (in fact they work better on large proteins). Although some details about the ligand binding epitope can be obtained, ligand-detected NMR methods do not reveal the ligand binding site on the drug target. Ligand-based screens rely on monitoring the change in some NMR parameter of the ligand upon its binding to the protein. One of the most useful of these NMR methods is saturation transfer difference (STD) spectroscopy [45], and its variant, competition-STD (c-STD) spectroscopy [46, 47]. If spins anywhere in the protein are saturated, the saturation will quickly spread throughout the protein by spin diffusion, and will be transferred to a ligand if it has a long-enough residence time in the binding site. If the ligand has a fast-enough off-rate, the bound-state saturation will be observed on the free state of the ligand, with its narrow resonances. In practice, the STD experiment works well for the range $0.1 \ \mu M < K_D < 1 \ mM$, with protein concentration of 0.5–$5.0 \ \mu M$ and ligand present in 50- to 500-fold molar excess.

The presence of signal in the STD spectrum of a ligand–protein complex must be interpreted in the broadest possible sense: there might be relatively tight binding at one binding site, weak binding at multiple sites, or some combination of the two.

NMR method: in this case ^{15}N-HSQC, depends on following the movement of cross-peaks as a small molecule is added. If a titration is performed, an NMR-K_D can be extracted, as shown in the graph. (**b**) Ligand-detected STD NMR method: (*i*) 1D control spectrum of AMP/kinase; (*ii*) STD spectrum of AMP/kinase; only resonances of atoms that contact the protein are present in the STD spectrum; (*iii*) STD spectrum of ATP/kinase complex; (*iv*) STD of ATP/kinase/competitor; the STD signal due to ATP is decreased because ATP is partially displaced from the binding site by the competitor, and new STD signals for the competitor appear, compared to spectrum (*iii*)

Table 1 Advantages and disadvantages of target- and ligand-based NMR screening methods

Screening method	Advantages	Disadvanatges
Target-detected		
	Detects high- and low-affinity ligands	Requires large amounts of isotope-labeled protein (~250 mg)[a]
	Structure-based	Limited to smaller protein targets (< ~40–60 kDa)
	Yields ligand binding site information	High protein concentration required (~25–80 μM)
	Detects site-specific binding only; nonspecific binding not detected	Knowledge of 3D structure of target protein and NMR assignments (or map) of at least active site residues required to identify active site hits
	Very robust and reliable	
	SAR for weak ligands ($K_D > $ ~10 μM)	
	SAR for higher affinity ligands by competition	
Ligand-detected		
STD	No isotope labeling required (~20 mg)[a]	Does not reveal ligand binding site
	Lower protein concentration required (~1–5 μM)	Signal may be due to binding at multiple sites
	No protein size limitation	
	No quality protein NMR spectra required	
Competition STD	Infer ligand binding site and K_I relative to "marker" with known binding site and K_D	Requires "marker" with known K_D and/or binding site
	Detects low- and high-affinity ligands	

[a]For a typical screen of about 1,000–2,000 fragments against a 50-kDa protein, including fragment hit validation and initial SAR development

If there is a reference ligand with known binding site, c-STD may be used to localize the binding site of a screened molecule. Competition-STD is a two-part experiment. First, the STD spectrum of the reference molecule is obtained. Next, the competitor is added, and the STD spectrum of the ternary mixture (reference molecule, competitor molecule, protein) is obtained. If both molecules are competing for the same binding site, the STD signal of the reference molecule will decrease. The magnitude of the decrease can be used to estimate the affinity of the competitor if the affinity of the reference is known and the two molecules are strictly competitive with each other for the same binding site [48]. Since c-STD can help rule out weak, nonspecific binding, it is a highly valuable addition if well-characterized reference molecules are known for the target.

Finally, substrate-based functional NMR assays can be used to derive the percentage inhibition or IC_{50} values [49]. In our experience, functional NMR assays can also reveal valuable details about the mode of action of modulators, since the

Table 2 NMR methods for determining ligand affinities/potencies

NMR method	Affinity/potency range
Target-based (2D HSQC)	$K_D > \sim10\ \mu M$
Target-based (competition 2D HSQC)	$K_I < \mu M$
Ligand-based (1D c-STD)	$K_I < nM–mM$
Substrate-based functional assays (1D NMR)	% Inhibition; IC_{50}

substrate, the product, and the ligand can be monitored in simple one-dimensional (1D) ^1H NMR spectra.

From the previous discussion it becomes clear that depending on the knowledge and characteristics of a drug target, an appropriate NMR screening method needs to be selected for any given FBDD campaign. Moreover, suitable NMR methods can be selected to derive ligand affinity or potency to assist SAR development (Table 2).

2.2 Library Considerations

Fragment-based approaches probe chemical space more efficiently than HTS approaches, are less dependent on legacy compound collections, and can provide hits for challenging targets. One of the great advantages of NMR-based methods is the ability to reliably identify weak binders with K_Ds in the low millimolar range, while still obtaining useful structural information about their potential binding site(s). With this affinity cut-off, screening a library of 1,000–2,000 fragments will result in multiple hits for most targets. The selection of these 1,000–2,000 compounds for an NMR-based screening library can be crucial to the success of the endeavor, and details of this important topic have been described in a number of publications (e.g., [50–54]). Candidate molecules are filtered to ensure favorable physicochemical properties and a lack of reactive functional groups. Issues of "chemical diversity" versus "drug likeness" must be balanced. Library members might be synthetic cores for which chemically elaborated back-up libraries are readily available for fast SAR. The chosen fragment screening method may to some degree also influence the design of such a fragment library [53]. If 3D structural information is available for the drug target, virtual screening may be employed to select focused FBS libraries to increase hit rates [15]. Several companies nowadays sell FBS libraries as part of their business since FBDD has become increasingly popular over the last decade.

Once candidate library members have been chosen, they must be validated by experiment. For each library member, the chemical structure is verified, the purity of the sample determined, and aqueous solubility measured. In addition, the fragments should be tested for their potential to aggregate at the high screening concentrations used for fragment-based NMR screening [51]. DMSO-d_6 stock solutions of the library must be plated and stored in a way that minimizes freeze/

thaw cycles and exposure to atmospheric water. In order to facilitate the identification of hits in ligand-detected NMR methods, library members are plated so that each screening cluster is "chemical shift encoded"; that is, within each cluster there are no degenerate chemical shifts between cluster members. Target-detected NMR screening methods do not have such requirements, but fragments in an "active" cluster must be deconvoluted to identify hit(s).

3 Fragment Hit-to-Lead Progression

3.1 Fragment Hit Validation and Initial SAR Development

Cross-validation of NMR results with information yielded by other biophysical, biochemical, and cell-based assays can be crucial to the progression of a fragment hit. Access to other assay methods is especially important when STD is used as the NMR screening method because STD reveals no information on the ligand binding site and is more susceptible to unrecognized nonspecific binding. Results from biophysical methods such as SPR, thermal denaturation, and isothermal calorimetry (ITC), in addition to X-ray crystallography and structure-based NMR studies, can be used to validate NMR hits. If available, biochemical and cell-biological functional assays are valuable tools for probing the interaction of a fragment hit with its target.

Even before project chemists become actively involved, SAR can be quickly progressed by testing obvious analogs of the initial fragment hit from readily available commercial or internal sources, which may include "expansion" libraries that have been prepared on the basis of members of the screening library. The value of a chemotype or structural motif becomes clearer if a series of molecules has been studied, and some initial SAR is seen. On the basis of results from the first round of analoging, project chemists will usually have ideas for further SAR development. It is important that the "iteration time" between submission of new compounds for testing and the reporting of test results back to the project team be as short as possible to maintain project momentum.

3.2 Evaluation of Binding Site and Binding Mode

Target-based NMR methods can often provide this crucial information, especially if site-specific assignments are available from the literature or can be obtained internally, and the 3D structure of the drug target is known. The detailed binding mode of a fragment hit by NMR can, however, only be obtained for smaller targets with MWs up to about 20–30 kDa, and requires significant resources. Thus, X-ray crystallography becomes the method of choice for determining the detailed binding mode of a fragment hit.

The preferred binding mode within a chemical series can change even within the same binding site as substituents are changed, thus confusing SAR development. In these cases, knowledge of the detailed binding modes of key members within a lead series is crucial for efficient fragment hit progression.

3.3 Ligand Efficiency Indices to Guide Fragment Hit Selection and Progression

Traditionally, affinity/potency has been the primary factor for hit selection and optimization. However, there is a strong correlation between increased MW and improved affinity/potency. Moreover, lead optimization typically yields bigger and more lipophilic compounds [30]. However, almost all absorption, distribution, metabolism, excretion, and toxicity (ADMET) parameters deteriorate with either increasing MW and/or partition coefficient (logP) [55] and good physicochemical properties help to reduce the attrition rate in late stage clinical trials [56]. Therefore, selecting appropriate hits with a good balance of MW and lipophilicity, and monitoring this balance in addition to affinity/potency during hit optimization, have been recognized as important factors for successful drug discovery.

Many fragment hits found through NMR screening will have weak affinity and will require substantial modification to become viable leads. As fragment hits are different from traditional HTS hits, a process tailored for fragment hit progression is required. Several LE indices have been proposed for guiding this process [34]. Thus, weak binders identified by fragment-based NMR screening might be good starting points for lead generation if they exhibit good LE and good LLE. LE and related indices estimate the efficiency of a binding interaction with respect to the number of non-hydrogen atoms and is a way of normalizing the binding energy by the size of the molecule [21–23]. Because LE cannot be evaluated independently of the molecular size [24], scaled LE scores have been proposed to enable a size-independent comparison of ligands [25–28]. LLE is a measure of the minimally acceptable lipophilicity per unit of in vitro potency: LLE(Leeson) = $pIC_{50}-cLogP$ (computed partition coefficient) [29] or a normalized LLE(Astex) = $0.111-(-1.36 \times$ LLE(Leeson)/number of heavy atoms) can be used for practical reasons for fragment hits [57]. Chemists will have the freedom to elaborate low MW, high LE hits before reaching unacceptable limits of MW and complexity, which often lead to compounds that exhibit unacceptable solubility, absorption, and permeability properties. Similarly, fragments with good LLE provide the opportunity to increase lipophilicity during lead optimization without reaching an unfavorable physical profile for the drug candidate. However, LLE does not include LE and vice versa. Since there is a significant predisposition towards improving potency simply by adding lipophilicity, LELP = logP/LE was proposed as a useful function to depict the price of LE paid in logP [30].

Affinity, or binding energy, comprises two components: enthalpy and entropy. It has recently been proposed that there are advantages to starting with enthalpically

driven leads [35, 58, 59], in which binding arises from specific molecular interactions such as hydrogen bonds, salt bridges, and van der Waals interactions. In contrast, entropically driven binding generally arises from nonspecific hydrophobic interactions. ITC is the tool of choice for determining the relative contributions of entropy and enthalpy to binding affinity [60]. The information from ITC is best interpreted in conjunction with a detailed structural model of the binding interaction (usually from X-ray crystallography) and provides a strong starting point for optimization of a lead series. The relative balance of entropy and enthalpy will, of course, change as optimization progresses, but thermodynamic analysis and detailed structure models can go a long way towards explaining unexpected SAR and in providing guidance on where to focus synthetic efforts. Thus, an enthalpic efficiency (EE) and a specific EE were proposed as additional tools for selecting compounds in lead discovery and for aiding lead optimization [36].

The tractability of a fragment hit for chemical elaboration is judged by project chemists, who have the expert knowledge needed to assess the possibilities for elaboration of a fragment hit with substituents, or recasting of a chemotype into an isostere. Chemists also assess the fragment hit for potential chemical novelty, especially important if the target has been extensively studied by other groups. Close interaction with project chemists is crucial to the success of a project. In the early stage of a project, a core structure that can easily be derivatized is advantageous for fragment hit progression.

Structural information about the binding mode of a fragment hit can be crucial for efficient hit-to-lead optimization, as discussed above. Therefore, we prefer to apply this FBDD approach to high-priority targets and drug targets for which X-ray or NMR structures can be obtained. Whenever possible, with this approach we like to provide the chemist with low MW, high LE, and high LLE compounds for which we know their binding mode to the drug target, thereby providing chemists with more room for optimizing pharmokinetic (PK)/ADMET properties during the lead optimization process. Thus, a structure-focused FBDD approach can produce leads for very challenging targets where other methods may fail (see BACE example below).

Follow-up strategies for fragment hits strongly depend on the nature and characteristics of the drug target and the fragment hits. For more challenging targets, structural data is crucial for efficient fragment hit-to-lead optimization, whereas for other targets with deep, well-defined active sites this may not necessarily be the case. In the latter case, high-concentration biochemical screening of libraries that contain "lead-like" compounds [39] may be more efficient than a structure-based NMR fragment screen, especially if a robust functional assay can be developed. High-concentration biochemical screens have the distinct advantage that they already provide a functional readout for the fragment hit, and the hit-to-lead process follows a traditional progression path. However, HCS of fragment libraries could be prone to larger numbers of "false positives," and orthogonal biophysical methods might become important for pruning fragment hit lists.

Although tethering/linking fragments that bind to proximal binding sites can in principle yield high-affinity linked molecules, this approach is often not very

practical due to difficulties in finding proximal binders, knowing their detailed binding mode, and due to limitations in linker chemistry and optimization [61]. Thus, expanding or growing initial fragment hits into more potent leads has become much more common than tethering for FBDD. FBDD approaches may also become very useful in better understanding the contributions of individual components of an existing lead [62], or for improving an existing lead by "fragment hopping" [63].

4 Structure-Based FBDD Approach Applied to BACE-1

4.1 BACE-1 as a Drug Target for Alzheimer's Disease

Alzheimer's disease (AD) is a progressive, ultimately fatal neurodegenerative disease that gives patients an average life expectancy of 7–10 years after diagnosis [64]. It is the leading cause of dementia in the elderly population, causing gradual loss of mental and physical function. In addition to the devastating physical and emotional impact of AD on patients and their families, all patients at an advanced stage of the disease will inevitably need long-term care, which places a huge social and economic burden on their families and society [65]. In the USA alone, there are currently four million AD patients, with an additional eight million subjects diagnosed with mild cognitive impairment (MCI), of whom many will progress to AD [66]. This number is expected to quadruple in the next three decades unless therapies that impact the underlying pathophysiology of AD can be identified.

Current therapies for AD, comprising acetylcholine esterase inhibitors and an NMDA receptor antagonist, offer only symptomatic relief by compensating for the neuronal and synaptic losses in AD patients through prolonging activation of the remaining neuronal network [67]. These therapies offer patients transient improvements in cognition and daily living functions, but do not halt disease progression. Thus, there are enormous unmet medical needs for the AD population.

The pathological hallmarks observed in the brains of AD patients are the extracellular amyloid plaques, mainly composed of an amyloid-β peptide with 42 amino acids in length (Aβ42), and the intracellular neurofibrillary tangles of hyperphosphorylated tau protein. According to the amyloid hypothesis [68–73], the prevailing theory in the field, the underlying cause of AD is the aggregation and deposition of Aβ42 in the brain due to its overproduction and/or diminished clearance. This hypothesis is supported by strong genetic, histopathological, and clinical evidence. All early-onset familial AD is identified by genetic mutations in amyloid precursor protein (APP) or presenilins (PS1 and PS2) that result in increased Aβ peptide production. Down's syndrome patients, who have an extra copy of chromosome 21 containing the APP gene, or individuals who have a duplication of only a portion of chromosome 21 that contains the APP gene, produce more Aβ peptides and develop early-onset AD [74, 75]. One Down's

individual [76], whose extra copy of the portion of chromosome 21 lacked the APP gene, did not develop AD. Among all β-amyloid species, Aβ42 is most prone to aggregation and most cytotoxic in vitro [77–80]. Lastly, active immunization against Aβ peptides reduced amyloid load in animal models [81, 82] and was associated with cognitive improvement for AD patients who developed robust anti-Aβ titers in human clinical trials [83–87].

Aβ peptides, ranging from 37 to 42 amino acids in length, differ from each other at the C-terminus. They are produced as minor products (5–10%) of the metabolism of the membrane-bound APP via two consecutive cleavages: first by β-site APP cleaving enzyme (BACE-1, also known as β-secretase or memapsin-2) [88–91], followed by γ-secretase, in competition with the major pathway (90–95%) of non-amyloidogenic processing of APP by α-secretase. There are two BACE isoforms, with BACE-1 mainly expressed in the central nervous system (CNS) and responsible for Aβ peptide production. BACE-2 cleaves APP at a different site to the BACE-1 cleavage and is mainly expressed in the periphery [92, 93]. BACE-1 knockout (KO) mice are normal, do not produce Aβ peptides, and have few overt phenotypes [94–97]. Crossing BACE-1 KO with transgenic mice that overproduce human APP eliminates Aβ production and amyloid plaque formation and rescues memory dysfunction [98]. These data suggest that Aβ peptide inhibition through small molecule BACE-1 inhibitors is highly promising as a disease-modifying therapy that may halt or even reverse the progression of AD. Therefore, BACE-1 has been a high priority therapeutic target for the treatment of AD throughout the pharmaceutical industry over the last decade.

BACE-1 is a membrane-anchored aspartic acid protease that is localized to the acidic compartments of endosomes and lysosomes in the CNS and has an optimal enzymatic activity at around pH 5. As a consequence, a BACE-1 inhibitor needs to be able to cross the blood–brain barrier and to have a significant non-protein bound fraction in order to reach the active site of the enzyme. This makes traditional aspartic protease inhibitors, which typically are large and peptidic, unsuitable as BACE-1 inhibitors. Moreover, the BACE-1 active site is extended, shallow and hydrophilic (Fig. 2) [99]. Therefore, the development of potent, selective, orally active, and brain penetrant low MW compounds has been a big challenge for the pharmaceutical industry [101, 102].

Many of the early drug discovery efforts focused on the development of transition state peptidomimetics that were known from the aspartic acid protease field [99, 103]. Although this approach yielded highly potent and selective BACE-1 inhibitors, the resulting compounds lacked in vivo efficacy probably due to their large MW and suboptimal PK properties. We review here how we have used a highly structure-driven approach, consisting of the integrated application of target-detected fragment-based NMR screening, X-ray crystallography, structure-based design and structure-assisted chemistry together with innovative biology, to develop a first-in-class clinical candidate as a potential proof-of-concept for the inhibition of BACE-1 in AD [104, 105]. Recently, several other fragment-derived BACE-1 inhibitors have also been described [106–113].
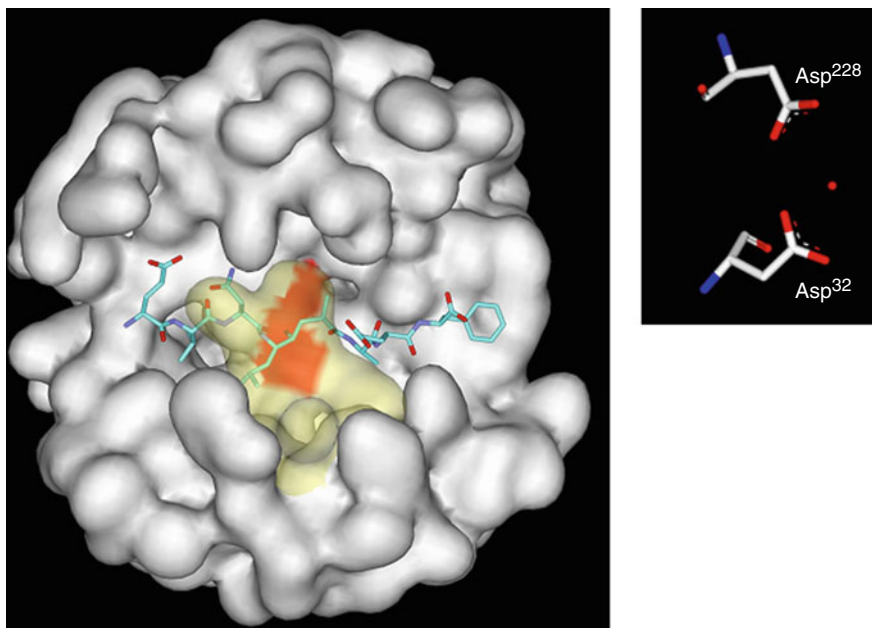
Fig. 2 BACE-1 characteristics. The overall fold of BACE-1 is typical for an aspartic acid protease, consisting of an N- and C-terminal lobe with the substrate binding site located in a crevice between the two lobes [99, 100]. A flexible hairpin, called the flap (*Yellow see-through surface*), partially covers the active site of BACE-1 and can adopt many different conformations as a result of inhibitor binding. In the center of the active site are the two aspartic acid residues (*orange* and *inset*) that are involved in the enzymatic reaction

4.2 Fragment Hit Identification

We developed an efficient protocol for the large scale production of a fully processed soluble version of ^{15}N-labeled BACE-1 for fragment-based NMR screening and X-ray crystallography in which the pre- and pro-sequences are autocatalytically removed within about 3 days at room temperature or 18 days at 4 °C at protein concentrations of ~5–10 mg/mL [114]. This refolding protocol from inclusion bodies yielded around 40 mg BACE-1/L cell paste. We used NMR to monitor structural details of the autocatalytic conversion, which revealed a major structural rearrangement in the N-terminal lobe from a partially disordered to a well-folded conformation suggesting that the pro-sequence may assist the proper folding of the protein. Once the protein was completely folded, we could recycle it multiple times for fragment-based NMR screening.

We screened over 10,000 fragments of a custom-built fragment library [7] at high concentrations (100 µM–1 mM each) in cocktails of 12 to identify active-site-directed hits by ^{15}N-HSQC NMR [104]. About half of these fragments were strictly rule-of-three compliant [20], whereas a large majority followed "reduced complexity" rules (MW < 350, cLogP ≤ 2.2, H-bond donor ≤ 3, H-bond acceptor

≤ 8, rotatable bonds ≤ 6, heavy atom count ≤ 22) [39]. At first, we did not have protein NMR resonance assignments for BACE-1. In order to not delay fragment-based NMR screening, we initially identified peaks of active site residues of BACE-1 by binding peptide inhibitors known from the literature and then screening for fragments that showed chemical shift perturbations of some of those peaks. Eventually, we obtained sequence-specific NMR resonance assignments for BACE-1, which then allowed us to study ligand binding in more structural detail [115]. Overall we identified nine distinct chemical classes of active site binders to BACE-1 in the 30 μM to 3 mM K_D range, as determined by NMR titration experiments (Fig. 3a).

Fig. 3 (continued)

c

(1) directed SbN

n=1 aminoimidazole
n=2 aminopyridine

Compound 2
K_D(NMR) = 15 μM
BACE-1 IC_{50} ~ 210 μM
MW = 273
LE = 0.39
FQ = 0.54
cLogP = 2.23
LLE = 2.59

(2) SBDD

n=1 iminohydantoin
n=2 iminopyrimidinone

d

Compound 2
K_D(NMR) = 15 μM
IC_{50}(HTRF) ~ 210 μM
pKa ~ 9

directed NMR screens

2-aminopyridines
Compound 3
K_D(NMR) = 32 μM
pKa ~ 7.2
MW = 233
LE = 0.39
FQ = 0.51
cLogP = 3.43
LLE = 1.06

+2-amino pyridines
(exploratory chemistry)
+2-amino imidazoles
+2-amino benzimidazoles
+2-amino triazines
+benzoamidines

Asp^{228}
2.8Å
2.9Å
2.9Å
2.7Å
2.7Å
Asp^{32}

crystal structures of **compound 2** (white) and **compound 3** (yellow)

Fig. 3 (continued)

Fig. 3 BACE-1 fragment hit identification and fragment hit-to-lead progression. (**a**) Fragment-based NMR screening hits for BACE-1. Nine classes of BACE-1 active-site-directed NMR hits were identified by screening 10,000 compounds from a customized NMR fragment library by ^{15}N-HSQC NMR. (**b**) Isothiourea fragment hit identification and optimization by NMR and X-ray crystallography. (**c**) Search for heterocyclic isothiourea isosteres. (**d**) 2-Aminopyridines and related heterocyclic isothiourea isosteres were identified through directed fragment-based NMR screening. (**e**) Structure-based design of prototype iminohydantoins yielded attractive starting points for the development of novel low MW BACE-1 inhibitors. See text for details

Among our initial fragment hits were several amidine-containing chemotypes, including the isothiourea **1** (Fig. 3b). We then tested over 200 analogs by NMR to derive initial SAR and discovered isothiourea **2**, which showed an NMR-K_D of 15 μM (LE = 0.39) and weak activity in an enzymatic assay. The NMR chemical shift perturbation data suggested that compound **2** binds to the two active site aspartates and extends into the S3 pocket while leaving the flap untouched in its "open" apo-conformation. Subsequently, the co-crystal structure of compound **2** with BACE-1 revealed details about how the isothiourea moiety forms an extensive H-bond donor acceptor array with the two active site aspartates and places the chloro-phenyl ring into the S1 pocket and extends deep into the shallow S3 pocket through the butyl-ether group. From that point on, this fragment was used in an X-ray soaking system to solve the X-ray structures of over 1,000 BACE-1 inhibitors that followed in this project.

When we discovered this NMR fragment hit several years ago, this type of hydrogen-bond network to the two active site aspartates was unprecedented in the

aspartic acid protease field. Unfortunately, potential hydrolytic instability of the isothiourea moiety of compound **2** renders it unsuitable for drug development. Therefore, we started an extensive search for heterocyclic isothiourea isosteres that would be pharmaceutically attractive with an appropriate basicity (pK_a range 6–10) to maintain the crucial H-bonding network with the two active site aspartates while limiting the number of H-bond donating groups and have molecular properties compatible with brain penetration. We pursued two approaches (Fig. 3c). In the first approach, we carried out focused NMR screens to identify heterocyclic structures including 2-aminoimidazoles and 2-aminopyridines to bind into the active site of BACE-1 [104]. In the second approach, we designed cyclic acylguanidines, including iminohydantoins and iminopyrimidinones [105].

4.3 Focused Search for Pharmaceutically Attractive Isothiourea Isosteres

While our general NMR fragment screening was still in progress, we initiated focused directed NMR screens of heterocyclic isothiourea isosteres that were available from our corporate library. During this process, we identified several heterocyclic cores as active site BACE-1 binders, which included 2-aminopyridines, 2-aminoimidazoles, 2-aminobenzimidazoles, 2-aminotriazines, and benzoamidines, whereas other related cores were not identified as hits (Fig. 3d). In the 2-aminopyridine series, we discovered compound **3**, which bound to the two active site aspartates with an NMR-K_D of 32 μM (LE = 0.39) as judged by the NMR chemical shift perturbation data. Compound **3** thus had LE [21] and fit quality (FQ) [25, 26] values similar to those of compound **2**. Its LLE [29] was, however, significantly reduced due to its increased hydrophobicity. Interestingly, the X-ray crystal structure of this fragment in complex with BACE-1 revealed the same H-bonding network as previously seen for compound **2**. Only a few months into the FBDD campaign, compound **3** provided the first attractive starting point for chemical elaboration. Exploratory chemistry on the 2-aminopyridine series was initiated. Small chemical libraries based on the 2-aminopyridine-phenethyl core were built to explore this chemotype. Several analogs with activities in the micromolar range were identified, and crystal structures for some of these suggested the synthesis of 3,6-disubstituted 2-aminopyridine, which yielded the first submicromolar inhibitors in this series. However, the planar nature of the 2-aminopyridine core and difficulties in synthesizing 3,6-disubstituted analogs prevented the easy development of more potent BACE-1 inhibitors with lead-like properties in this series.

In an alternate approach, novel cyclic acylguanidine active-site-binding cores such as iminohydantoin and iminopyrimidinone were conceptualized (Fig. 3c) in which the crucial aspartate-binding amidino motif, common to fragment-based NMR screening hits **2** and **3** and of similar weak basicity, is conserved. It was suggested that disubstitution at C5 (iminohydantoin) or C5 and C6

(iminopyrimidinone) would simultaneously provide direct access to both prime and non-prime binding sites adjacent to the catalytic aspartate residues, with substitution on the second ring nitrogen providing a further handle for accessing binding pockets adjacent to the active site. To test this hypothesis, the prototype iminohydantoin (compound **4**) and its N1-analog were designed and synthesized. The 3-chlorobenzyl substituent was predicted to bind in the S1 pocket, in analogy to **2** and **3**. We were delighted to find that iminohydantoin **4** bound to BACE-1 with an NMR-K_D of 200 μM, whereas no binding was observed for its N1-analog. Despite its weak binding activity, compound **4** showed promising LE and LLE values for fragment hit progression. An X-ray structure of **4** in complex with BACE-1 confirmed that **4** bound as predicted (Fig. 3e). We then tested several related N3 and N1 analogs. We consistently found by NMR that the N3-, but not the N1-prototype iminohydantoins bound into the active site of BACE-1. About a year into the FBDD approach, we had now discovered a very attractive novel core structure that was chemically stable, had a pK_a compatible with CNS penetration, and provided ample opportunities to extend the molecule into nearby substrate binding pockets using well-known hydantoin chemistry.

4.4   Fragment Hit-to-Lead Progression

During fragment hit-to-lead progression we quickly identified a second binding mode of the iminohydantoin core in the active site of BACE-1 using X-ray crystallography. This is represented by compound **5**, in which an extensive ligand–BACE-1 H-bonding network is maintained, but the iminohydantoin core is flipped in the active site (Fig. 4a). This observation turned out to be highly significant because this mode proved to be the preferred binding mode as lead optimization evolved. NMR chemical shift perturbation data could be used to quickly categorize ligands with respect to these two binding modes (Fig. 4b). Simple changes in the substituents could not only cause the iminohydantoin core to flip, but also to tilt or slightly shift in the binding pocket while maintaining an extensive H-bond network with the two active site aspartates. Therefore, X-ray structural data was crucial for medicinal chemists to understand otherwise confusing SAR (Fig. 4b).

It was important to demonstrate quickly that we can produce potent iminohydantoin BACE-1 inhibitors that had submicromolar IC$_{50}$s in the enzymatic assay. The binding mode of iminohydantoin **7** (Fig. 4b) suggested that cyclohexylmethyl and cyclohexylethyl extensions into the respective hydrophobic S1 and S2′ pockets should achieve this goal (Fig. 4c). The resulting iminohydantoin **8** was in fact the first submicromolar inhibitor in this series. Its crystal structure confirmed the underlying structure-based design and suggested that a further increase in potency should be possible by introducing a cyclic urea with the propyl extension in the proper (**S**)-configuration. Again, the ensuing iminohydantoin **9** bound to BACE-1 as expected and showed an increased potency in the enzymatic assay. Isolation of the single stereoisomer with 4(**S**)/4(**R**) configuration yielded compound **10** with
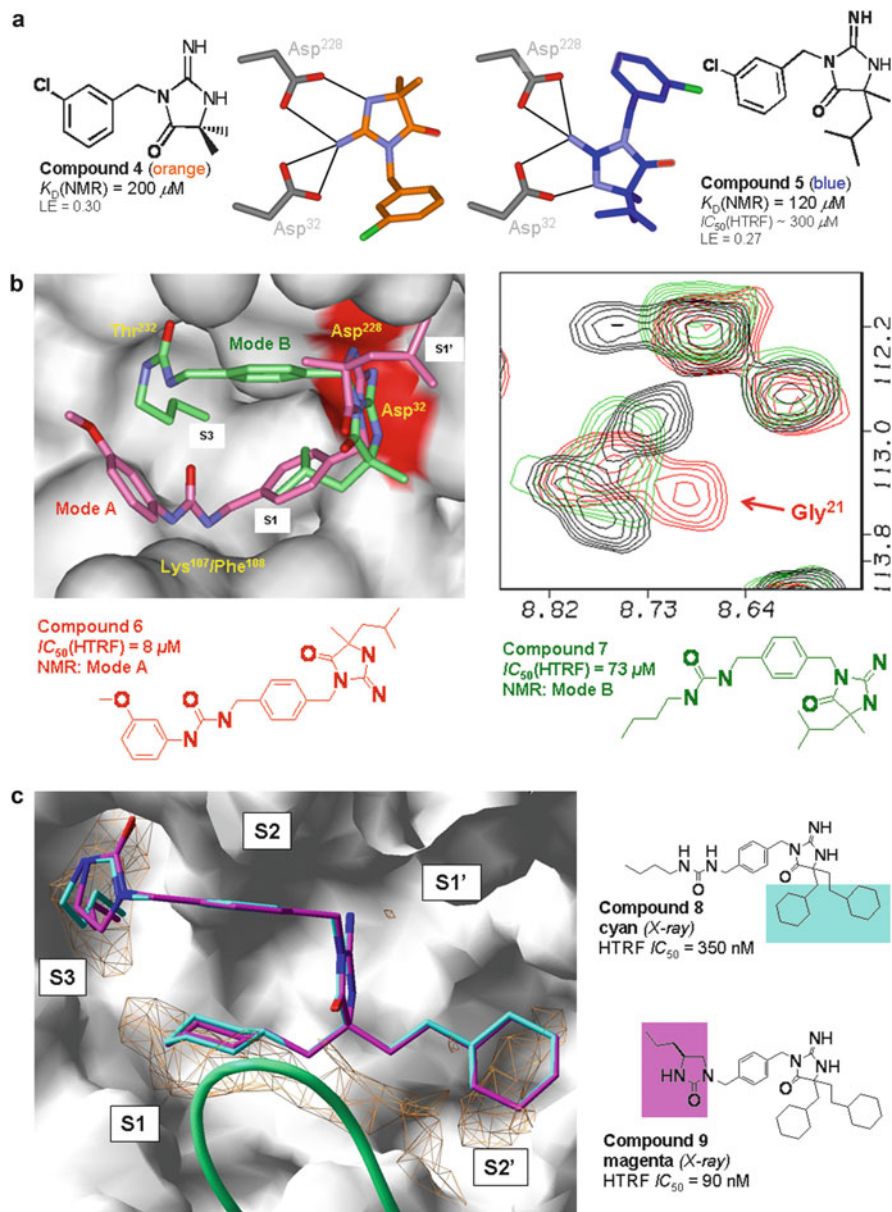
Fig. 4 (continued)

a cellular IC_{50} in the submicromolar range (Fig. 4d). However, the resultant compounds became non-leadlike with significantly reduced LE (despite an improved FQ), increased cLogP (yielding a very poor LLE), poor rat PK, and no

d Compound 10
BACE-1 K_i = 67 nM
cell IC_{50} = 695 nM
CatD/BACE-1 = 0.5
rapid rat AUC = 0.58 μM.h
MW = 536
LE = 0.25
FQ = 0.67
cLogP = 7.5
LLE = -0.3

Compound 11
BACE-1 K_i = 650 nM
cell IC_{50} = 2600 nM
CatD/BACE-1 = 1
rapid rat AUC = 1.03 μM.h
MW = 319
LE = 0.37
FQ = 0.64
cLogP = 5.4
LLE = 0.79

S1- S3 Approach

e Compound 11
BACE-1 K_i = 650 nM
cell IC_{50} = 2600 nM
CatD/BACE-1 = 1
rapid rat AUC = 1.03 μM.h

MW = 319
LE = 0.37
FQ = 0.64
cLogP = 5.4
LLE = 0.79

rotatable bonds

Compound 12
BACE-1 K_i = 3.7 μM
cell IC_{50} = 13 μM
CatD/BACE-1 > 20
rapid rat AUC = 5 μM.h
brain / plasma = 1.8
MW = 265
LE = 0.37
FQ = 0.58
cLogP = 1.85
LLE = 3.58

Fig. 4 (continued)

selectivity over the aspartic acid protease Cathepsin D. Molecular modeling suggested that we should be able to truncate compound **10** to a *N*-methyl and extend the iminohydantoin core deep into the S3 subpocket (S3$_{sp}$) more directly through a contiguous hydrophobic patch without adding as much MW to the iminohydantoin core. It was good to see that the truncated *N*-methyl iminohydantoin analog (compound **11**) showed much higher LE than compound **10** (while maintaining a good FQ) and only a three- to fourfold loss in cellular 70–75% decerase in potency (Fig. 4d).
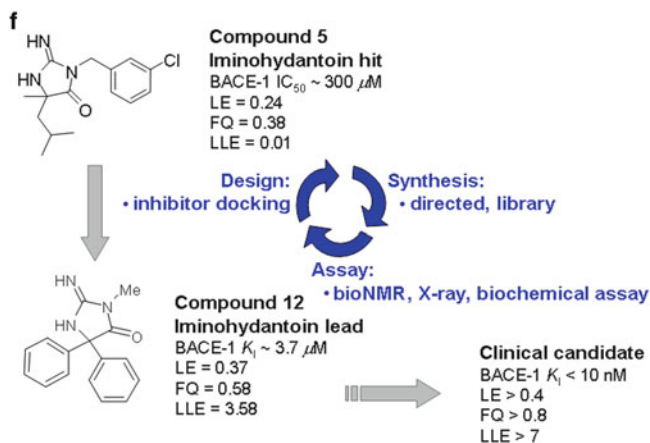
Fig. 4 Iminohydantoin fragment hit progression. (**a**) A second binding mode of the iminohydantoin core in the active site of BACE-1 was revealed by X-ray crystallography. (**b**) Simple changes in the substituents could cause the iminohydantoin core to flip, tilt, or shift in the active site while maintaining an extensive H-bond network with the two aspartates, thus structural data simplified SAR development. (**c**) Structure-based design of the first series of submicromolar iminohydantoin BACE-1 inhibitors. (**d**) Truncated *N*-methyl iminohydantoins provided a more direct way to build toward S3 through a contiguous hydrophobic patch from S1 through S3 into S3$_{sp}$. (**e**) Truncated *N*-methyl iminohydantoins showed improved LE, with compound **12** showing excellent lead-like properties. The X-ray crystal structures of BACE-1 in complex with compound **11** and **12** showed relatively open flap conformations, with flap residue Tyr71 (shown in *cyan* in the "closed" flap, peptidomimetic inhibitor conformation [99]) displaced by one of the phenyls at the 5-position of compound **12** [105]. (**f**) Iterative structure-assisted chemistry was able to improve ligand efficiency indices during fragment hit-to-lead optimization and lead optimization in the iminohydantoin series. See text for details

Knowing that we needed to develop a CNS drug, we then reduced the number of rotatable bonds of compound **11** and designed the rigid, compact 5,5′-diphenyl iminohydantoin core structure (compound **12**) which, despite being only weakly active, now showed excellent lead-like properties with good LE, much better LLE, and an overall favorable profile with respect to cellular potency, selectivity, rat PK, and brain penetration (Fig. 4e). Thus, compound **12** was the superior choice for lead optimization.

The X-ray structure of BACE-1 in complex with compound **11** (Fig. 4e) revealed a relatively open flap conformation, with the aliphatic chain of the inhibitor projecting towards S2′ in close proximity to a pocket that we termed F′. In X-ray structures of peptidomimetic inhibitors bound to BACE-1 [99], the F′ pocket is occupied by flap Tyr71 but is vacated in X-ray crystal structures of iminohydantoin **11** and related compounds. Incorporation of 5-phenyl substitution to exploit occupancy of F′ was a key design concept that resulted in identification of 5,5′-diphenyl-iminohydantoin **12**, in which one of the C5 phenyl substituents now

occupied the unique F′ pocket, again yielding a more "open" flap conformation (Fig. 4e).

During iminohydantoin fragment hit-to-lead optimization, which involved an iterative process of molecular modeling, structure-assisted synthesis, and functional and structural evaluation, the LE between the initial fragment hit and the optimized fragment lead was increased significantly from 0.24 to 0.37 kcal/mol/heavy atom (Fig. 4f), yielding a corresponding increase in FQ. Due to a significant reduction in cLogP, compound **12** also showed a much improved LLE as compared to the initial iminohydantoin fragment hit. Thus, the primary goal during lead optimization was to increase potency and selectivity of the iminohydantoin lead series while maintaining good LE and molecular properties that would be compatible with brain penetration.

4.5   Iminohydantoins: S1–S3 Occupancy

The truncated *N*-methyl iminohydantoins (compounds **11** and **12**) showed much higher LE than compound **10**, and provided opportunities to build into the S3 pocket more directly without increasing the MW of the iminohydantoins as much as in the earlier series, which was extended at the N1-position towards the S2 pocket. Compound **12** possesses a diphenyl substitution at C5, with one of the C5 phenyl substituents occupying a unique binding pocket designated F′ that is normally filled by the enzyme flap Tyr residue in the closed-flap enzyme conformation (Fig. 5). Thus, compound **12** offered several opportunities to extend the iminohydantoin core from the C5 position into the surrounding S1–S3 and S2′ substrate binding pockets. Based on the X-ray structure of compound **12** in complex with BACE-1, molecular modeling suggested that we could extend the phenyl in the S1 pocket at the *meta*-position toward the S3 pocket. We then tested this hypothesis by synthesizing analogs that probed different extensions at this *meta*-position (Fig. 5). SAR revealed that a phenyl extension is tolerated and that small hydrophobic substituents at the 3-position of this distal phenyl improved the enzymatic K_{I} by about an order of magnitude, yielding several submicromolar BACE-1 inhibitors. The X-ray structure of the diphenyl-iminohydantoin with a 3-pyridine extension (compound **13**) in complex with BACE-1 exhibited an H-bond to a bound water molecule in the S3 subpocket and could explain SAR that showed a preference of the 3-pyridine over the 4-pyridine analog. It could explain additional SAR that revealed a preference of substitutions at the 3- over the 4-position at the distal phenyl in the S3 pocket. Substitutions at the 3-position presumably could reach deep into the S3 subpocket by replacing this nonstructural bound water molecule in the S3 subpocket.

Despite this structural knowledge it still turned out to be challenging to significantly improve the potency of the iminohydantoin series with respect to cellular potency and PK properties. However, by use of structure-assisted SAR development the team was ultimately able to develop BACE-1 inhibitors with high affinity,
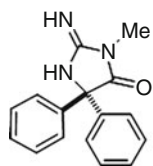
Compound 12
BACE-1 K_I = 3.7 µM
cell Aβ40 IC$_{50}$ = 13µM
CatD/BACE1 >20
LE: 0.37
FQ = 0.58
cLogP = 1.85
LLE = 3.58
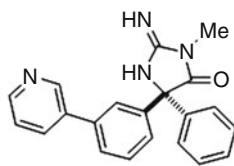
Compound 13
BACE-1 K_I = 109 nM
cell Aβ40 IC$_{50}$ = 633 nM
CatD/BACE1 = 120
LE: 0.37
FQ = 0.71
cLogP = 2.84
LLE = 4.12

+/-
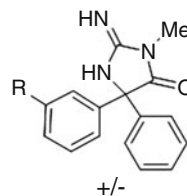
R	K_i µM
Ph	3.25
3-MePh	0.55
3-CBPh	0.37
3-ClPh	0.30
3-MeOPh	0.19
4-MeOPh	3.80
3-Py	0.53
4-Py	3.80
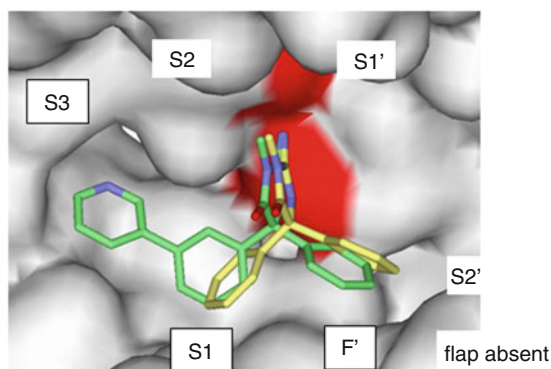
Fig. 5 S1–S3 occupancy in iminohydantoins. The X-ray crystal structures of compound **12** (*yellow*) and compound **13** (*green*) are shown superimposed when bound to BACE-1. The bound structure of compound **13** could explain the SAR shown in the inserted table. See text for details

selectivity, and excellent PK properties to achieve brain penetration and CNS efficacy in vivo [116].

5 Conclusion and Perspectives

We have used a highly structure-driven approach composed of fragment-based NMR screening, X-ray crystallography, and structure-assisted chemistry to develop a first-in-class clinical candidate as a potential proof-of-concept for the inhibition of BACE-1 in AD. Crucial to this achievement was the initial identification of a ligand-efficient isothiourea fragment and its X-ray crystal structure, which revealed an extensive H-bond network with the two active site aspartates. This interaction was unprecedented in the aspartic acid protease field when we discovered it several

years ago. This detailed 3D structural information then enabled the design and validation of novel, chemically stable and accessible heterocyclic acylguanidines as aspartic acid protease inhibitor cores. Lead optimization guided by structure-based design afforded unique, low MW, high affinity, selective iminopyrimidinones as BACE-1 inhibitors in which the hydrophobic interactions in the S1, S3, and $S3_{sp}$ pockets were optimized to achieve excellent cellular potency. The resulting leads were conformationally restricted with few rotatable bonds, which contribute to their high LE indices. These iminoheterocyclic BACE-1 inhibitors possess desirable molecular properties as potential therapeutic agents to test the amyloid hypothesis in a clinical setting. Optimized iminopyrimidinones have shown high oral bioavailability, good CNS penetration, and robust reductions of cerebrospinal fluid and brain $A\beta$ in animal models.

Combining biomolecular NMR, X-ray crystallography, and molecular modeling with structure-assisted chemistry and innovative biology as an integrated approach for FBDD can solve very difficult problems, as illustrated in this chapter. BACE-1 has been a challenging CNS target for small molecule drug discovery, where more conventional lead generation approaches had failed despite extensive efforts for over a decade. However, none of the components mentioned above would have been successful if applied in isolation. Therefore, the future for FBDD looks bright as long as an appropriate infrastructure can be provided for this technology to tackle appropriate problems in drug discovery.

References

1. Jahnke W, Erlanson DA (eds) (2006) Fragment-based approaches in drug discovery. In: Mannhold R, Kubinyi H, Folkers G (series eds) Methods and principles in medicinal chemistry, Vol 34. Wiley-VCH, Weinheim, Germany
2. Zartler ER, Shapiro MJ (eds) (2008) Fragment-based drug discovery: a practical approach. Wiley, Chichester, UK
3. Rees DC, Congreve M, Murray CW, Carr R (2004) Fragment-based lead discovery. Nat Rev Drug Discov 3:660–672
4. Erlanson DA, McDowell RS, O'Brien T (2004) Fragment-based drug discovery. J Med Chem 47:3463–3482
5. Erlanson DA (2006) Fragment-based lead discovery: a chemical update. Curr Opin Biotechnol 17:643–652
6. Hajduk PJ, Greer J (2007) A decade of fragment-based drug design: strategic advances and lessons learned. Nat Rev Drug Discov 6:211–219

7. Wyss DF, Eaton HL (2007) Fragment-based approaches to lead discovery. Front Drug Des Discov 3:171–202

8. Alex AA, Flocco MM (2007) Fragment-based drug discovery: what has it achieved so far? Curr Top Med Chem 7:1544–1567

9. Congreve M, Chessari G, Tisi D, Woodhead AJ (2008) Recent developments in fragment-based drug discovery. J Med Chem 51:3661–3680

10. de Kloe GE, Bailey D, Leurs R, de Esch IJP (2009) Transforming fragments into candidates: small becomes big in medicinal chemistry. Drug Discov Today 14:630–646

11. Chessari G, Woodhead AJ (2009) From fragment to clinical candidate-a historical perspective. Drug Discov Today 14:668–675

12. Murray CW, Rees DC (2009) The rise of fragment-based drug discovery. Nat Chem 1:187–192

13. Schulz MN, Hubbard RE (2009) Recent progress in fragment-based lead discovery. Curr Opin Pharm 9:615–621

14. Murray CW, Blundell TL (2010) Structural biology in fragment-based drug design. Curr Opin Struct Biol 20:497–507

15. Gozalbes R, Carbajo RJ, Pineda-Lucena A (2010) Contributions of computational chemistry and biophysical techniques to fragment-based drug discovery. Curr Med Chem 17:1769–1794

16. Fink T, Bruggesser H, Reymond JL (2005) Virtual exploration of the small-molecule chemical universe below 160 Daltons. Angew Chem Int Ed Engl 44:1504–1508

17. Fink T, Reymond JL (2007) Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F: assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery. J Chem Inf Model 47:342–353

18. Hann MM, Leach RL, Harper G (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. J Chem Inf Comput Sci 41:856–864

19. Coyne AG, Scott DE, Abell C (2010) Drugging challenging targets using fragment-based approaches. Curr Opin Chem Biol 14:299–307

20. Congreve M, Carr R, Murray CW, Jhoti H (2003) A 'rule of three' for fragment-based lead discovery? Drug Discov Today 8:876–877

21. Hopkins AL, Groom CR, Alex A (2004) Ligand efficiency: a useful metric for lead selection. Drug Discov Today 9:430–431

22. Abad-Zapatero C, Metz JT (2005) Ligand efficiency indices as guideposts for drug discovery. Drug Discov Today 10:464–469

23. Verdonk ML, Rees DC (2008) Group efficiency: a guideline for hits-to-leads chemistry. ChemMedChem 3:1179–1180

24. Reynolds CH, Bembenek SD, Tounge BA (2007) The role of molecular size in ligand efficiency. Bioorg Med Chem Lett 42:4258–4261

25. Reynolds CH, Tounge BA, Bembenek SD (2008) Ligand binding efficiency: trends, physical basis, and implications. J Med Chem 51:2432–2438

26. Bembenek SD, Tounge BA, Reynolds CH (2009) Ligand efficiency and fragment-based drug discovery. Drug Discov Today 14:278–283

27. Nissink JWM (2009) Simple size-independent measure of ligand efficiency. J Chem Inf Model 49:1617–1622

28. Orita M, Ohno K, Niimi T (2009) Two 'golden ratio' indices in fragment-based drug discovery. Drug Discov Today 14:321–328

29. Leeson PD, Springthorpe B (2007) The influence of drug-like concepts on decision-making in medicinal chemistry. Nat Rev Drug Discov 6:881–890

30. Keseru GM, Makara GM (2009) The influence of lead discovery strategies on the properties of drug candidates. Nat Rev Drug Discov 8:203–212

31. Hajduk PJ (2006) Fragment-based drug design: how big is too big? J Med Chem 49:6972–6976

32. Perola E (2010) An analysis of the binding efficiencies of drugs and their leads in successful drug discovery programs. J Med Chem 53:2986–2997
33. Abad-Zapatero C, Perisic O, Wass J, Bento AP, Overington J, Al-Lazikani B, Johnson ME (2010) Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an atlas-like representation. Drug Discov Today 15:804–811
34. Schultes S, de Graaf C, Haaksma EEJ, de Esch IJP, Leurs R, Kramer O (2010) Ligand efficiency as a guide in fragment hit selection and optimization. Drug Discov Today Technol 7:e157–e162
35. Ferenczy GG, Keseru GM (2010) Thermodynamics guided lead discovery and optimization. Drug Discov Today 15:919–932
36. Ladbury JE, Klebe G, Freire E (2010) Adding calorimetric data to decision making in lead discovery: a hot tip. Nat Rev Drug Discov 9:23–27
37. Teague SJ et al (1999) The design of lead like combinatorial libraries. Angew Chem Int Ed Engl 38:3743–3748
38. Oprea TI et al (2001) Is there a difference between leads and drugs? A historical perspective. J Chem Inf Comput Sci 41:1308–1315
39. Hann MM, Oprea TI (2004) Pursuing the leadlikeness concept in pharmaceutical research. Curr Opin Chem Biol 8:255–263
40. Dalvit C (2009) NMR methods in fragment screening: theory and a comparison with other biophysical techniques. Drug Discov Today 14:1051–1057
41. Retra K, Irth H, Muijlwijk-Koezen JE (2010) Surface plasmon resonance biosensor analysis as a useful tool in FBDD. Drug Discov Today Technol 7:e181–e187
42. Jhoti H, Cleasby A, Verdonk M, Williams G (2007) Fragment-based screening using X-ray crystallography and NMR spectroscopy. Curr Opin Chem Biol 11:485–493
43. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Discovering high affinity ligands for proteins: SAR by NMR. Science 274:1531–1534
44. Hajduk PJ, Augeri DJ, Mack J, Mendoza R, Yang J, Betz SF, Fesik SW (2000) NMR-based screening of proteins containing 13 C-labeled methyl groups. J Am Chem Soc 122:7898–7904
45. Mayer M, Meyer B (2001) Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. J Am Chem Soc 123:6108–6117
46. Wang YS, Liu D, Wyss DF (2004) Competition STD NMR for the detection of high-affinity ligands and NMR-based screening. Magn Reson Chem 42:485–499
47. McCoy MA, Senior MM, Wyss DF (2005) Screening of protein kinases by ATP-STD NMR spectroscopy. J Am Chem Soc 127:7978–7979
48. Dalvit C (2008) Theoretical analysis of the competition ligand-based NMR experiments and selected applications to fragment screening and binding constant measurements. Concepts Magn Reson A 32A:341–372
49. Dalvit C, Ardini E, Flocco M, Fogliatto GP, Mongelli N, Veronesi M (2003) A general NMR method for rapid, efficient, and reliable biochemical screening. J Am Chem Soc 125:14620–14625
50. Jacoby E, Davies J, Blommers MJ (2003) Design of small molecule libraries for NMR screening and other applications in drug discovery. Curr Top Med Chem 3:11–23
51. Baurin N, Aboul-Ela F, Barril X, Davis B, Drysdale M, Dymock B, Finch H, Fromont C, Richardson C, Simmonite H, Hubbard RE (2004) Design and characterization of libraries of molecular fragments for use in NMR screening against protein targets. J Chem Inf Comput Sci 44:2157–2166
52. Blomberg N, Cosgrove DA, Kenny PW, Kolmodin K (2009) Design of compound libraries for fragment screening. J Comput Aided Mol Des 23:513–525
53. Chen IJ, Hubbard RE (2009) Lessons for fragment library design: analysis of output from multiple screening campaigns. J Comput Aided Mol Des 23:603–620
54. Boyd SM, de Kloe GE (2010) Fragment library design: efficiently hunting drugs in chemical space. Drug Discov Today Technol 7:e173–e180

55. Gleeson MP (2008) Generation of a set of simple, interpretable ADMET rules of thumb. J Med Chem 51:817–834
56. Wenlock MC, Austin RP, Barton P, Davis AM, Leeson PD (2003) A comparison of physicochemical property profiles of development and marketed oral drugs. J Med Chem 46:1250–1256
57. Mortenson PN, Murray CW (2009) Ligand lipophilicity efficiency – assessing lipophilicity of fragments and early hits. Presented at RSC Fragments 2009, Astra Zeneca Alderley Park, UK, 4–5 March 2009, Poster 9
58. Freire E (2008) Do enthalpy and entropy distinguish first in class from best in class? Drug Discov Today 13:869–874
59. Scott AD, Phillips C, Alex A, Flocco M, Bent A, Randall A, O'Brien R, Damian L, Jones LH (2009) Thermodynamic optimisation in drug discovery: a case study using carbonic anhydrase inhibitors. ChemMedChem 4:1985–1989
60. Ward HJ, Holdgate GA (2001) Isothermal titration calorimetry in drug discovery. Prog Med Chem 38:309–376
61. Chung S, Parker JB, Bianchet M, Amzel LM, Stivers JT (2009) Impact of linker strain and flexibility in the design of a fragment-based inhibitor. Nat Chem Biol 5:407–413
62. Barelier S, Pons J, Marcillat O, Lancelin J-M, Krimm I (2010) Fragment-based deconstruction of Bcl-xL inhibitors. J Med Chem 53:2577–2588
63. Ji H, Li H, Martasek P, Roman LJ, Poulos TL, Silverman RB (2009) Discovery of highly potent and selective inhibitors of neuronal nitric oxide synthase by fragment hopping. J Med Chem 52:779–797
64. Levy-Lehad E, Wijsman E, Nemens E, Anderson A, Goddard KA, Weber JL (1995) A familial Alzheimer's disease locus on chromosome 1. Science 269:970–973
65. Melnikova I (2007) Therapies for Alzheimer's disease. Nat Rev Drug Discov 6:341–342
66. Mount C, Downton C (2006) Alzheimer disease: progress or profit? Nat Med 12:780–784
67. Moreira PI, Zhu X, Nunomura A, Smith MA, Perry G (2006) Therapeutic options in Alzheimer's disease. Expert Rev Neurother 6:897–910
68. Hardy J, Selkoe DJ (2002) The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. Science 297:353–356
69. Haass C, Selkoe DJ (2007) Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β-peptide. Nat Rev Mol Cell Biol 8:101–112
70. Korczyn AD (2008) The amyloid cascade hypothesis. Alzheimers Dement 4:176–178
71. Hardy J (2006) Has the amyloid cascade hypothesis for Alzheimer's been proved? Curr Alzheimer Res 3:71–73
72. Selkoe D (2001) Alzheimer's disease: genes, proteins, and therapy. Physiol Rev 81:741–766
73. Archer HA, Edison P, Brooks DJ, Barnes J, Frost C, Yeatman T (2006) Amyloid load and cerebral atrophy in Alzheimer's disease: a 11 C-BIP positron emission tomography study. Ann Neurol 60:145–147
74. Olson MI, Shaw CM (1969) Presenile dementia and Alzheimer's disease in mongolism. Brain 92:147–156
75. Mann DMA (1998) Alzheimer's disease and Down's syndrome. Histopathology 13:125–137
76. Prasher VP, Farrer MJ, Kessling AM, Fisher EMC, West RJ, Barber SPC, Butler AC (1998) Molecular mapping of IIv Alzheimer-type dementia in Down's syndrome. Ann Neurol 43:380–383
77. Pike CJ, Walencewicz AJ, Glabe CG, Cotman CW (1991) In vitro aging of amyloid-beta protein causes peptide aggregation and neurotoxicity. Brain Res 573:311–314
78. Lorenzo A, Yankner B (1994) Beta-amyloid neurotoxicity requires fibril formation and is inhibited by congo red. Proc Natl Acad Sci USA 91:12243–12247
79. Iversen LL, Mortishire-Sith RJ, Pollack SJ, Shearman MS (1995) The toxicity in vitro of beta-amyloid protein. Biochem J 311:1–16
80. Tsai J, Grutzendler J, Duff K, Gan W (2004) Fibrillar amyloid deposition leads to local synaptic abnormalities and breakage of neuronal branches. Nat Neurosci 7:1181–1183

81. Mor F, Monsonego A (2006) Immunization therapy in Alzheimer's disease. Expert Rev Neurother 6:653–659
82. Qu B, Boyer PJ, Johnston SA, Hynan LS, Rosenberg RN (2006) Aβ42 gene vaccination reduces brain amyloid plaque burden in transgenic mice. J Neurol Sci 244:151–158
83. Bayer AJ, Bullock R, Jones RW, Wilkinson D, Paterson KR, Jenkins L (2005) Evaluation of the safety and immunogenicity of syntheticAβ42 (AN1792) in patients with AD. Neurology 64:94–101
84. Gilman S, Koller M, Black RS, Jenkins L, Griffith SG, Fox NC (2005) Clinical effects of Aβ immunization (AN1792) in patients with AD in an interrupted trial. Neurology 64:1553–1562
85. Hock C, Konietzko U, Streffer JR, Tracy J, Signorell A, Muller-Tilmanns B (2003) Antibodies against b-amyloid slow cognitive decline in Alzheimer's disease. Neuron 38:547–554
86. Fox NC, Black RS, Gilman S, Rossor MN, Griffith SG, Jenkins L (2005) Effects of Ab immunization (AN1792) on MRI measures of cerebral volume in Alzheimer disease. Neurology 64:1563–1572
87. DaSilva KA, Aubert I, McLaurin J (2006) Vaccine development for Alzheimer's disease. Curr Pharm Des 12:4283–4293
88. Hussain I, Powell D, Howlett DR, Tew DG, Meek TD, Chapman C, Gloger IS, Murphy KE, Southan CD, Ryan DM, Smith TS, Simmons DL, Walsh FS, Dingwall C, Christie G (1999) Identification of a novel aspartic protease (Asp 2) as beta-secretase. Mol Cell Neurosci 14:419–427
89. Sinha S, Anderson JP, Barbour R, Basi GS, Caccavello R, Davis D, Doan M, Dovey HF, Frigon N, Hong J, Jacobson-Croak K, Jewett N, Keim P, Knops J, Lieberburg I, Power M, Tan H, Tatsuno G, Tung J, Schenk D (1999) Purification and cloning of amyloid precursor protein beta secretase from human brain. Nature 402:537–540
90. Vassar R, Bennett BD, Babu-Khan S, Kahn S, Mendiaz EA, Denis P, Teplow DB, Ross S, Amarante P, Loeloff R, Luo Y, Fisher S, Fuller J, Edenson S, Lile J, Jarosinski MA, Biere AL, Curran E, Burgess T, Louis JC (1999) Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. Science 286:735–741
91. Yan R, Bienkowski MJ, Shuck ME, Miao H, Tory MC, Pauley AM, Brashier JR, Stratman NC, Mathews WR, Buhl AE, Carter DB, Tomasselli AG, Parodi LA, Heinrikson RL, Gurney ME (1999) Membrane-anchored aspartyl protease with Alzheimer's disease beta-secretase activity. Nature 402:533–537
92. Farzan M, Schnitzler CE, Vasilieva N, Leung D, Choe H (2000) BACE2, a beta-secretase homolog, cleaves at the beta site and within the amyloid-beta region of the amyloid-beta precursor protein. Proc Natl Acad Sci USA 97:9712–9717
93. Saunders AJ, Kim T-M, Tanzi RE (1999) BACE maps to chromosome 11 and a BACE homolog, BACE2, reside in the obligate Down syndrome region of chromosome 21. Science 286:1255a
94. Luo Y, Bolon B, Kahn S, Bennett BD, Babu-Khan S, Denis P, Fan W, Kha H, Zhang J, Gong Y, Martin L, Louis JC, Yan Q, Richards WG, Citron M, Vassar R (2001) Mice deficient in BACE1, the Alzheimer's beta secretase, have normal phenotype and abolished beta-amyloid generation. Nat Neurosci 4:231–232
95. Luo Y, Bolon B, Damore MA, Fitzpatrick D, Liu H, Zhang J, Yan Q, Vassar R, Citron M (2003) BACE1 (beta secretase) knockout mice do not acquire compensatory gene expression changes or develop neural lesions over time. Neurobiol Dis 14:81–88
96. Willem M, Garratt AN, Novak B, Citron M, Kaufmann S, Rittger A, DeStrooper B, Saftig P, Birchmeier C, Haass C (2006) Control of peripheral nerve myelination by the β-secretase BACE1. Science 314:664–666
97. Wang H, Song L, Laird F, Wong PC, Lee H-K (2008) BACE1 knock-outs display deficits in activity-dependent potentiation of synaptic transmission at mossy fiber to CA3 synapses in the hippocampus. J Neurosci 28:8677–8681

98. Ohno M, Sametsky EA, Younkin LH, Oakley H, Younkin SG, Citron M, Vassar R, Disterhoft JF (2004) BACE1 deficiency rescues memory deficits and cholinergic dysfunction in a mouse model of Alzheimer's disease. Neuron 41:27–33

99. Ghosh AK, Shin D, Downs D, Koelsch G, Lin X, Ermolieff J, Tang J (2000) Design of potent inhibitors for human brain memapsin 2 (β-secretase). J Am Chem Soc 122:3522–3523

100. Hong L, Koelsch G, Lin X, Wu S, Terzyan S, Ghosh AK, Zhang XC, Tang J (2000) Structure of the protease domain of memapsin 2 (β-secretase) complexed with inhibitor. Science 290:150–153

101. Stachel SJ (2009) Progress toward the development of a viable BACE-1 inhibitor. Drug Dev Res 70:101–110

102. Durham TB, Shepherd TA (2006) Progress toward the discovery and development of efficacious BACE inhibitors. Curr Opin Drug Discov Dev 9:776–791

103. Maillard MC, Hom RK, Benson TE, Moon JB, Mamo S, Bienkowski M, Tomaselli AG, Woods DD, Prince DB, Paddock DJ, Emmons TL, Tucker JA, Dappen MS, Brogley L, Thorsett ED, Jewett N, Sinha S, Varghese J (2007) Design, synthesis, and crystal structure of hydroxyethyl secondary amine-based peptidomimetic inhibitors of human beta -secretase. J Med Chem 50:776–781

104. Wang YS, Strickland C, Voigt JH, Kennedy ME, Beyer BM, Senior MM, Smith EM, Nechuta TL, Madison VS, Czarniecki M, McKittrick BA, Stamford AW, Parker EM, Hunter JC, Greenlee WJ, Wyss DF (2010) Application of fragment-based NMR screening, X-ray crystallography, structure-based design, and focused chemical library design to identify novel μM leads for the development of nM BACE-1 (β-site APP cleaving enzyme 1) inhibitors. J Med Chem 53:942–950

105. Zhu Z, Sun ZY, Ye Y, Voigt J, Strickland C, Smith EM, Cumming J, Wang L, Wong J, Wang YS, Wyss DF, Chen X, Kuvelkar R, Kennedy ME, Favreau L, Parker E, McKittrick BA, Stamford A, Czarniecki M, Greenlee W, Hunter JC (2010) Discovery of cyclic acyl-guanidines as highly potent and selective β-site amyloid cleaving enzyme (BACE) inhibitors: part I – inhibitor design and validation. J Med Chem 53:951–965

106. Geschwindner S, Olsson LL, Albert JS, Deinum J, Edwards PD, de Beer T, Folmer RH (2007) Discovery of a novel warhead against β-secretase through fragment-based lead generation. J Med Chem 50:5903–5911

107. Kuglstatter A, Stahl M, Peters JU, Huber W, Stihle M, Schlatter D, Benz J, Ruf A, Roth D, Enderle T, Hennig M (2008) Tyramine fragment binding to BACE-1. Bioorg Med Chem Lett 18:1304–1307

108. Murray CW, Callaghan O, Chessari G, Cleasby A, Congreve M, Frederickson M, Hartshorn MJ, McMenamin R, Patel S, Wallis N (2007) Application of fragment screening by X-ray crystallography to β-secretase. J Med Chem 50:1116–1123

109. Congreve M, Aharony D, Albert J, Callaghan O, Campbell J, Carr RA, Chessari G, Cowan S, Edwards PD, Frederickson M, McMenamin R, Murray CW, Patel S, Wallis N (2007) Application of fragment screening by X-ray crystallography to the discovery of aminopyr-idines as inhibitors of β-secretase. J Med Chem 50:1124–1132

110. Edwards PD, Albert JS, Sylvester M, Aharony D, Andisik D, Callaghan O, Campbell JB, Carr RA, Chessari G, Congreve M, Frederickson M, Folmer RH, Geschwindner S, Koether G, Kolmodin K, Krumrine J, Mauger RC, Murray CW, Olsson LL, Patel S, Spear N, Tian G (2007) Application of fragment-based lead generation to the discovery of novel, cyclic amidine β-secretase inhibitors with nanomolar potency, cellular activity, and high ligand efficiency. J Med Chem 50:5912–5925

111. Yang W, Fucini RV, Fahr BT, Randal M, Lind KE, Lam MB, Lu W, Lu Y, Cary DR, Romanowski MJ, Colussi D, Pietrak B, Allison TJ, Munshi SK, Penny DM, Pham P, Sun J, Thomas AE, Wilkinson JM, Jacobs JW, McDowell RS, Ballinger MD (2009) Fragment-based discovery of nonpeptidic BACE-1 inhibitors using tethering. Biochemistry 48:4488–4496

112. Godemann R, Madden J, Krämer J, Smith M, Fritz U, Hesterkamp T, Barker J, Höppner S, Hallett D, Cesura A, Ebneth A, Kemp J (2009) Fragment-based discovery of BACE1 inhibitors using functional assays. Biochemistry 48:10743–10751
113. Madden J, Dod JR, Godemann R, Kraemer J, Smith M, Biniszkiewicz M, Hallett DJ, Barker J, Dyekjaer JD, Hesterkamp T (2010) Fragment-based discovery and optimization of BACE1 inhibitors. Bioorg Med Chem Lett 20:5329–5333
114. Wang Y-S, Beyer BM, Senior MM, Wyss DF (2005) Characterization of autocatalytic conversion of precursor BACE1 by heteronuclear NMR spectroscopy. Biochemistry 44:16594–16601
115. Liu D, Wang Y-S, Gesell JJ, Wilson E, Beyer BM, Wyss DF (2004) Backbone resonance assignments of the 45.3 kDa catalytic domain of human BACE1. J Biomol NMR 29:425–426
116. Stamford A (2010) Discovery of small molecule, orally active and brain penetrant BACE1 inhibitors. Paper presented at 239th ACS National Meeting, San Francisco, CA, 21–25 March 2010

Combining Biophysical Screening and X-Ray Crystallography for Fragment-Based Drug Discovery

Michael Hennig, Armin Ruf, and Walter Huber

Abstract Over the past decade, fragment-based drug discovery (FBDD) has gained importance for the generation of novel ideas to inspire synthetic chemistry. In order to identify small molecules that bind to a target protein, multiple approaches have been utilized by various groups in the pharmaceutical industry and by academic groups. The combination of fragment screening by biophysical methods and in particular with surface plasmon resonance technologies (SPR) together with the visualization of the binding properties by X-ray crystallography offers a number of benefits. Screening by SPR identifies ligands for a target protein as well as provides an assessment of the binding properties with respect to affinity, stoichiometry, and specificity of the interaction. Despite the huge technology advances of the past years, X-ray crystallography is still a resource-intensive technology, and SPR binding data provides excellent measures to prioritize X-ray experiments and consequently enable a better success rate in obtaining structural information. Information on the chemical structures of fragments binding to a protein can be used to perform similarity searches in compound libraries in order to establish structure–activity relationships as well as to explore particular scaffolds. At Roche we have applied this workflow for a number of targets and the experiences will be outlined in this review.

Keywords BACE · Chymase · Fragment-Based lead discovery · Library Design · Screening · Surface Plasmon Resonance · X-ray Crystallography

M. Hennig (✉), A. Ruf and W. Huber
F. Hoffmann-La Roche Ltd., Pharma Research and Early Development, Discovery Technologies, Basel 4070, Switzerland
e-mail: Michael.Hennig@Roche.com

Contents

Abbreviations

BACE   β-Secretase
HTS   High-throughput screening
ITC   Isothermal calorimetry
LIMS   Lab information management system
NMR   Nuclear magnetic resonance
SLS   Swiss light source
SPR   Surface plasmon resonance
Stdv   Standard deviation
wt   Wild type

1   Introduction

Fragment screening methods have evolved in the last decade from a serendipitous observation of solvent molecules in crystal structures to a new technology for generating ligand binding information in drug discovery [1–5]. In contrast to screening methods of random compound libraries like high-throughput screening (HTS), fragment screening uses a compound library that contains substances that are selected to follow mainly three basic rules: a molecular weight of less than about 300 Da, no more than three hydrogen bonding donors or acceptors, and a computed partition coefficient (clogP) of less than three [4]. Additional selection criteria might be added, such as no more than three rotatable bonds and a polar surface of less than 60 $\mathring{A}^2$. The small size and limited potential for formation of diverse interaction of the fragments leads to a higher degree of promiscuity of binding. These properties lead to a number of advantages. Compared to HTS, the screening effort can be limited to hundreds or a few thousand compounds to explore

the chemical space of a binding site. Optimization of a hit or lead towards a drug molecule benefits from favorable physicochemical properties and low chemical complexity. Also, the ligand efficiency, as defined by binding energy per nonhydrogen atom [6, 7], is typically higher for fragments than for HTS hits.

The disadvantage of the fragment screening approach is the psychological and technological hurdle to synthetic chemistry efforts with ligand binding affinities in the micromolar (μM) to millimolar (mM) range, including the higher error in their determination. Another consequence of the low binding affinity can be the lack of functional activity of such compounds in cellular and in vivo assays of the initial hits, and the requirement of sometimes significant chemistry effort to synthesize compounds that show such properties. The potential lack of selectivity of small compounds (promiscuity) is in our experience not a problem and selectivity is quickly achieved during lead optimization.

In this review, we highlight the importance of biophysical methods like surface plasmon resonance (SPR), NMR, isothermal titration calorimetry (ITC) and others for the fragment screening approach because these methods are used as a primary filter to select the compounds with higher likelihood of being visualized with X-ray crystallographic methods compared to the use of X-ray methods for primary screening. Advances regarding sensitivity and throughput, especially in SPR methods, have enabled evaluation of several thousand compounds in a few days or weeks and many examples of successful identification of new binding motifs have been reported. Some challenges like deviations in the buffer conditions between the methods remain and are potential ways for further improvement of the procedures.

2   Biophysical Methods for Fragment Screening

It has been well recognized that the application of several screening technologies in parallel, followed by diligent analysis of the data on the basis of the strengths and limits of the respective methods, is crucial for the identification of novel chemical scaffolds with high potential for generating new therapeutic agents [8, 9].

An enzymatic or ligand displacement assay as generally used in HTS campaigns seems to be the most straightforward approach in the identification of biochemically active fragments and there is a wide variety of such assays used by different companies [10]. The lack of sensitivity of such assays for the characterization of fragment binding in routine HTS settings demands alternative methods, despite the success in investigating particular protein targets where more sensitive biochemical assays could be established [11].

In this review, an approach is described that overcomes these difficulties using biophysical methods. The advantages and limitations of the various technologies are discussed in some detail in order to give guidance for the selection of the most appropriate methods for fragment screening [12]. The main focus, however, is on a detailed description of how to use SPR-based methods.

2.1 Biophysical Methods for Detection of Ligand Binding: An Overview

Among biophysical methods, high-throughput crystallography is the most elegant way to detect ligand binding since it provides direct structural information [13]. Several technological innovations such as improvements in expression systems and methodology of cloning and expression; advances in robotics, liquid handling and miniaturization; improvements in working with large cocktails of test compounds; and increased efficiency in data collection, processing, and analysis have made this a realistic and practical proposition. Application, however, is limited to targets for which a robust crystallographic system is available that allows the production of large numbers of diffraction quality crystals for soaking or cocrystallization experiments. The essential prerequisites of this technology are discussed by Davies and Tickle [14].

All other methods for label-free binding studies can be assigned to basically two different classes. In one class, the binding event is measured in homogeneous solutions (homogeneous assays) and in the other at a solid–liquid interface with one of the binding partners immobilized on the solid phase (heterogeneous assays). The class of homogeneous assays are based on detection technologies such as NMR [8, 15–18], mass spectrometry [19, 20], ITC [21, 22], thermal shift assays (also called ThermoFluor) [23–25], and backscattering interferometry [26, 27]. Among these, NMR is the most widely used technique in fragment screening [13, 28]. For all the other technologies, their application in the assessment of fragment binding has been demonstrated, but there is only limited data available on applications in screening of large fragment libraries.

The strength of NMR-based technologies is the ability to use changes in one or more NMR parameters, including chemical shift (^{1}H, ^{15}N, ^{13}C), anisotropy measurements, transverse and longitudinal relaxation of the ligand or protein, cross relaxation in the protein–fragment complex, or cross relaxation between the fragment and the protein-bound water-molecules. Zartler and Huaping [28] emphasize that the type of NMR method selected for a screening effort depends on different factors (size of target protein and quality of spectra, protein consumption, number of measurements planned etc.) but that the first and foremost factor should be the type of information that is expected from the experiment. They identified five different data types:

1. Does the ligand bind (Yes/No)?
2. Which ligand is binding (from a mixture)?
3. How is the ligand binding?
4. Where is the ligand binding?
5. What is the structural and dynamic implication of binding?

The large number of experimental NMR methods can be subdivided into two main classes: ligand-observed and protein-observed methods. The ligand-observed experiments are differentiated by the type of the magnetization and how the pulse-sequence delays are set. The two main unlabeled experiments are STD [28, 29] and WaterLOGSY [8, 15, 30]. The ligand-observed experiments deliver data about

whether a ligand binds or not, about the identity of the bound ligand and also (with certain limitations) how, i.e., in what orientation, the ligand binds to the target [28]. Protein-observed methods are limited to targets with a molecular weight less than 30–40 kDa due to line width and relaxation considerations. Protein-observed methods frequently require spectrum simplification through ^{15}N and ^{13}C labeling. Assignment of resonance lines to amino acids is advantageous because it allows determination of structural constraints. The investment in time and effort required for full assignment of resonances can be reduced by selective isotope labeling of one or more types of amino acid. Typically, these protein-observed experiments are heteronuclear single quantum coherence (HSQC) experiments using ^{15}N or ^{13}C as the heteronucleus [31]. The first example of protein-based screening was structure–activity relationship (SAR) by NMR [17]. In the meantime, automated data evaluation methods were developed to handle large sets of heteronuclear correlation spectra [32, 33].

The class of heterogeneous assays includes detection technologies that are based on optical transducers such as SPR [34], guided mode reflectance filter [35], and white light interferometer devices. All the optical devices are able to detect either a small change in the refractive index [36] or a change of the thickness of an adlayer occurring upon binding of molecules to their surface. Although examples of the use of all these technologies to monitor small ligand binding have been presented at scientific meetings, the only application to fragment screening reported in scientific journals has been for the SPR-based systems from Biacore [37–40], FujiFilm [41], and SensiQ [42]. The limited feasibility of the methods to work with fragment-sized compounds results from special limitations and challenges. Since the refractive index change, and hence the response, scales with the molecular mass of the ligand, the technology has to be pushed to its detection limit. Consequently, immobilization strategies must be developed that lead to high densities of active biomolecule on the surface. Due to the low affinity of the fragments, screening has to be performed at high concentrations, which makes the method susceptible to unspecific binding and false positive hits. The use of suitable control proteins is highly recommended to circumvent this problem. Hämäläinen et al. [37] suggested for thrombin fragment screening the use of a blocked thrombin as control for unspecific binding as well as proteins like serum albumin and carbonic anhydrase as further control proteins. Nordström et al. [38] worked with an active site mutated matrix metalloproteinase, MMP-12, as a control protein to identify fragments that interact specifically with the active site of the protein. Perspicace et al. [39] used the zymogen form of chymase as a control protein in which an N-terminally attached small proregion is bound to the active site and blocks the protein. An additional method to validate active site binding of a ligand is a competition assay with known active site binders [39].

2.2 Choice of Assay Methods (Criteria for Selection)

Combining complementary technologies is beneficial for identifying and reconfirming new chemical scaffolds that can be exploited in a fragment-based approach. For cost and efficiency reasons it is advantageous to select one leading

method as a workhorse for the primary screen and to use the other method solely for hit confirmation. Some of the main aspects to be considered in such a method selection are discussed below in more detail. The most prominent methods used in label-free interactions analyses are listed in Table 1 with their respective properties. Potential weaknesses highlighted here need not disqualify a certain method, but should indicate that the impact of any potential issues must be carefully considered in the application of the technique to fragment screening.

2.2.1 Statistical Assay Control

Screening is about making a decision on the interaction of a particular compound with a biological target. Independent from the read-out technology used for screening, the data on which such decision is based are subject to variability and hence uncertainty. However, the degree of uncertainty can be evaluated and estimated by application of statistical tools. These statistical criteria are useful to monitor during all states of a screening workflow as they help to assess reliability, reproducibility and sensitivity of a given assay and hence deliver experimental facts to investigate the quality of the assay. Finally the statistical tools can be used in data analysis to distinguish, based on statistical arguments, between positive and negative signals in screenings. Along this line, reproducibility and robustness of SPR like assays can be tested with the same tools as biochemical HTS assays [37, 39]. There are technologies, however, such as NMR, ITC and thermofluor where such statistical assay controls could not be defined because the response evolution is not independent from the molecule under investigation.

2.2.2 Material Consumption

Although fragment screening involves testing of a relatively low number of compounds compared to HTS, material consumption and costs are an important argument in technology selection. Cost considerations should include all disposables (plates, tips, sensor chips etc.) as well as the biological material consumed. For example protein production in a quantity and quality required for methods with low sensitivity like NMR, ITC or Thermofluor can limit the number of compounds tested in a fragment screening effort or the application of the method overall. Methods with higher sensitivity due to the high density of the immobilized biological target and, in addition, opportunities for regeneration (such as the SPR based) have a clear advantage in this respect since the once immobilized protein can be regenerated and reused for many experiments. This is not the case for all methods working in homogeneous solutions and also not for the SPR based technology from Corning. The intrinsic sensitivity for the Corning technology is the same as for other SPR based methods, however, the higher demand for protein results from the set-up in disposable micro titer well plates requiring freshly immobilized protein in each and every experiment.

Table 1 Comparison of various biophysical methods

| | NMR | | | ITC | Thermofluor | SPR | | | Fortebio |
	STD	Waterlogsy	HSQC			Biacore	Zeiss	Corning	
Target modification	No	No	Yes	No	No	Yes	Yes	Yes	Yes
Target consumption	High	High	High	High	High	Low	Low	High	Low
Dynamic range	High	High	High	Limited	Limited	Limited	Limited	Limited	Limited
Statistical control	No	No	No	No	No	Yes	Yes	Yes	Yes
False positive susceptibility	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Throughput	Low	Low	Low	Low	High	Medium	Medium	High	Medium

2.2.3   Throughput

Fragment screening involves testing for binding of many candidate ligands requiring a robust and reliable approach for data acquisition as well as data evaluation. Thermofluor and the SPR method from Corning are the methods with the highest potential throughput. Both technologies are based on 384 well plates, and the high degree of parallelization allows carrying out several ten-thousands of binding experiments per day. At the other end of the scale is ITC with much lower degree of automation and throughput of experiments. All the other techniques have a throughput of several hundreds to a few thousand binding experiments a day, which is sufficient to deal with several thousands of fragment molecules that are typical for such libraries.

2.2.4   False Positive Susceptibility

All assay technologies are susceptible to false positives that are caused by the imperfection of the in-vitro model system. Compared to biochemical or functional assays, direct binding assay technologies add unspecific binding as a possible cause for false positives. If properly designed (exclusion of pH, salt effects) protein observed NMR is probably the technology with the lowest susceptibility to unspecific binding. Ligand promiscuity, i.e., aggregation of ligands in solution will give responses in STD and WaterLOGSY experiments similar to that in the presence of protein binding [8] and can be eliminated by performing control experiments in the absence of target protein. Label free methods with the binding event occurring at a solid/liquid interface are highly susceptible to false positive hits since any deposition of material at this interface will lead to a positive response if no special measures are taken. Unspecific binding can be accounted for in such methods by parallel immobilization of reference proteins [36, 37, 39, 43]. Reference proteins could either be an unrelated protein (for example carbonic anhydrase), or better the identical protein target with a blocked binding site. Blockage of the active site of an enzyme or the anticipated drug binding site can be achieved by several methods like introduction of binding site destructive mutants or the binding of irreversible inhibitors.

2.2.5   Modification of Target

The reliability of the biological system under investigation is extremely important. Consequently, assay methods are preferred where none of the interacting partners has to be modified by labeling or immobilization. Immobilization can induce a severe modification of the protein with respect to structure, flexibility and consequently activity. It must be part of the assay development to select an adequate immobilization procedure that does not modulate protein activity and to thoroughly checking intactness of target protein with control measurements using positive controls.

2.2.6   Dynamic Range

Dalvit [8] described dynamic range limitations for certain technologies that could lead to higher numbers of false negatives. He argues that with many of the technologies protein/fragment interactions can be detected only at fragment concentrations close to the equilibrium binding constant K_D. For these technologies, for instance SPR, the observed response is directly proportional to the ratio L/K_D with L being the fragment concentration in solution. In SPR experiments with fragments the lower limit of LT/K_D leading to a detectable response is probably in the order of 0.2. This limitation is of concern for very low affinities when K_D is significantly higher than the solubility limits of the fragments. In this respect the thermofluor methods is probably the method with the highest limits because the low affinity of the fragments might not lead to a strong stabilization of the protein and hence to a non detectable shift in the protein melting temperature [12]. By contrast, it has been shown that NMR methods can have a higher dynamic range. In WaterLOGSY experiments it was demonstrated that binding was still observed even if the ratio of LT/K_D is as low as 0.07 [44].

2.3   The SPR Based Binding Assay for Screening

2.3.1   The Hardware

Commercially available instruments that can be used for SPR based experiments are available from several vendors (see Table 1). The set-up of an SPR based binding assay given in this review is related to the use of a Biacore A100 instrument that achieves higher throughput of measurements by parallelization. It enables parallel testing of four ligands independently in four flow through channels. Each channel provides the possibility to immobilize four proteins in parallel. For example one channel allows the measurement of the wt-protein and 1–3 reference proteins to eliminate false positives due to unspecific binding or to characterize specificity of binding. The sensor chips most often used are the so called CM5 sensor for covalent immobilization of the target via amide coupling chemistry. Recently a C7 sensor was launched with a much higher binding capacity for protein immobilization. The CM5 and the CM7 sensors are both equipped with a carboxymethyldextran adlayer [45]. Alternatively, sensors with immobilized Ni-chelator or streptavidin have been used if the protein that has to be immobilized contains the appropriate affinity tags, a polyhistidine sequence [46–48] or biotin [49, 50].

2.3.2   The Immobilization Strategy

The most frequently applied method to immobilize soluble proteins on the sensor chip surface is amine coupling. Covalent binding is achieved by activation of

carboxylic acid groups on the surface of a CM7 (CM5) sensor and subsequent linkage of these activated carboxylic acids via the amino acid side chains of lysine of the protein. No upfront biomolecular engineering or chemical modification of the protein is necessary. Amine coupling is probably the method by which the highest density of immobilized protein can be achieved. The irreversible covalent coupling makes the set-up extremely robust with respect to leakage of protein, however, the random immobilization is often seen as a disadvantage because of the potential for loss of active protein [51]. Another disadvantage is the lack of feasibility of regenerating a sensor chip after adhesion of undesired compounds (promiscuous binders) as it has been demonstrated that such adhered substances can significantly influence the outcome of follow-up binding experiments in a screening effort.

In this respect reversible capturing of histidine tagged proteins has clear advantages because it enables full regeneration of the surface (removal of protein and ligand) and reconstruction with fresh histidine-tagged protein after each binding experiment [48]. However, such strategies require larger amounts of protein.

2.3.3 Assay Quality

The quality of an SPR based direct binding assay can be described in a similar manner to an HTS assay by measures that characterize the robustness and the reproducibility of the assay. In order to determine the reproducibility of a screen, a set of compounds is tested in replicate. It is important that all experimental steps of a given screen such as sample preparation, injection mode, washing procedures, data evaluation are included. The statistical data of the correlation (for instance slope and standard error) are indicative for the reproducibility of the data [39].

The Z' factor introduced by Zang is a well accepted measure for the robustness of HTS screens. It is calculated according to (1):

$$Z' = 1 - \frac{3\sigma_{\mathrm{s}} + 3\sigma_{\mathrm{b}}}{R_{\mathrm{s}} - R_{\mathrm{b}}}. \tag{1}$$

In this equation the indices s and b denote the variation (σ) or the average response (R) of the positive (s) and a negative (b) control. R_{s} is determined at saturation concentration of the positive response. With certain limitations, the Z' factor can be used for expressing the robustness of an SPR based fragment screen. One such limitation is the molecular weight dependency of SPR responses. Since SPR measurements are dependent on the molecular weight of the compound, Z' factors are only relevant measures for robustness if they are determined for control compounds that have a molecular weight comparable to the average molecular weight of the compounds to be tested in a screen. Low Z'-factors result either from high standard deviations for the negative control and/or the positive controls. This can often be optimized by optimizing the running buffer, regeneration and washing conditions but also the sample preparation steps. Another source of small values of Z' is a low density of active protein on the sensor surface that is reflected in small

saturation responses for the positive controls. In this case the optimization of the assay might be achieved by a better purity and activity of the target protein as well as an alternative immobilization procedure. It has been discussed, that molecular weight dependent Z'-factors can be used to determine the minimum molecular weight and the percentage of compounds of a given library for which statistically relevant data could be expected for that screen [39].

2.3.4 The Screening Cascade

The screening cascade in SPR based fragment screening contains a series of assays that enable the application of different filter criteria for the selection of true positive binders. An overview on the most commonly used filters is given in Table 2.

2.3.5 Single Concentration Affinity Filter

The measured responses at the given concentration should be located in a window that is defined by the average responses and the respective standard deviation of negative and positive controls. The lower limit of a positive response is usually taken as three times the standard deviation of a negative control. The upper limit of such a window is less well defined as many of the compounds show over-stoichiometric binding at high concentration. Nonoptimal behavior with respect to stoichiometry does not per se disqualify compounds as interesting binders that could appear as positives by X-ray crystallography. It has been suggested to differentiate between nonstoichiometric binders and "superstoichiometric" binders (>5 times the saturation response of positive control) which are disqualified for follow up work [52].

Table 2 Overview of selection filters and the respective assay types

Filter	Filter criteria	Type of assay performed
Affinity filter	Response at screening concentration $>3x$ standard deviations of negative control	Single concentration binding assay with wild-type protein
Promiscuity filter	Curve shape during association and/or dissociation, superstoichiometry etc.	Single concentration binding assay with wild-type protein
Specificity filter	Response ratio of responses measured on target and on suitable reference protein (active site mutation, blockage)	Single concentration assay with parallel immobilization of wild type and reference protein
	Displacement of test compound by reference compound	Competition assay with control analyte molecule
Dose response filter	Ratio of responses at different concentrations	Screening at different concentrations
	Shape of dose response (saturation, slope etc.)	Dose response assay with concentration series

2.3.6 Promiscuity Filter

The term promiscuous binders has recently been applied to a class of compounds that often show up in high through put screens as false positive hits due to their ability to inhibit a broad spectrum of different protein classes. The investigation of promiscuous binding in solution indicates that such compounds form soluble or colloidal aggregates that envelop the protein. It was recently demonstrated that such promiscuous binding can easily be identified in time resolved SPR experiments and a number of mechanisms for the inhibition of protein function is suggested. The classification scheme presented in this work can be used during the evaluation of single concentration data to rapidly characterize and eliminate such compounds [52].

2.3.7 Specificity Filters

In SPR technology, any adsorption of compounds at the sensing surface will lead to a signal response and the observed signal is a superposition of specific binding to the desired binding sites on the target biomolecule and nonspecific binding to any place on the surface of the biomolecule or anywhere on the surface of the sensor. Special care is required to design an experimental setup that can distinguish between specific and nonspecific binding in order to deselect compounds that lack specific binding properties to the site of interest. Most of the approaches are based on preparing reference channels by immobilizing proteins that are structurally related to the target, but have a blocked active site or binding site of interest disabling specific binding of the analyte.

Blockage of the binding site can be achieved by site directed mutagenesis, i.e., by impairing or modifying the targeted site of a given protein via the exchange of one or several essential amino acids [43, 51] possibility is the use of a covalent irreversible inhibitor [36, 37] or an inactive form of the active protein (a zymogen) as reference protein [39].

Another possibility to filter for specific binding is a competition experiment with compounds known to bind to the binding site of the target [39, 43]. In this case the binding experiments have to be performed with (1) the pure test analyte (2) the reference compound and (3) with a mixtures of both substances. Generally, the compound concentrations in mixtures are the same as those in the solutions that contain analyte and reference alone. In the case of noncompetitive binding (different binding sites of analyte and reference compound) the sensor signal of the mixture is the sum of the sensor signals that were measured for the two compounds alone. In the case of competitive binding (binding of the analyte and reference compound to the same binding site) the resulting signal of the mixture is of intermediate strength between the two signals measured for the compounds alone. The expected signal for the mixture can be calculated taking the fractional occupancies at the binding site by the competitor and test analyte into consideration [39]. They can be derived by applying the law off mass action under the assumption that the concentration of the compounds in solution is not changed upon binding (this assumption is true for experiments in a flow system).

2.3.8 Dose Response Filters

Dose response filters are very valuable but require the highest workload. They are based on data recorded for dilution series of compounds (8–10 concentrations per compound). Exclusion criteria are based on the fit of the experimental data points to theoretical curves with respect to curve slope and saturation behavior. Sigmoidal dose response (response versus logarithm of concentration) or hyperbolic (response versus concentration) functions are both used as theoretical fit functions. Due to throughput limitations of most of the presently available SPR systems, complete dose response curves can only be recorded for a limited number of compounds. In case no specificity filter can be applied, a primary screen could deliver several hundreds of positives, as hit rates of 10–20% are observed for some targets. For such projects, a rough dose response filter can be applied by measurement of two concentrations per compound. Based on the theoretical background dose/response, the behavior of a compound can be tested by measuring the response at two different concentrations and comparing the resulting response ratio with the theoretically expected one.

The number and accuracy of filters varies from target to target and strongly depends on the properties of the target and the feasibility of developing the respective assays. Sometimes the application of the selectivity filter is not feasible as the ligand binding site is not well defined (e.g., targets with many allosteric sites) or an appropriate reference ligand is not available to perform competition assays. In such cases selection of positive hits relies on filters such as affinity, promiscuity and dose response, only. Material and measurement time is saved, if several of the filter criteria can be covered by a single assay. With the flexibility offered by modern SPR instruments, many reference proteins can be immobilized in parallel with the target protein making data for selectivity, promiscuity and affinity criteria available in one single assay. The sequence of assays in a screening cascade is guided by efficiency consideration, i.e., assays with lower time demand per tested compound are generally located at the top of the cascade whereas more time consuming assays are at the bottom when filtering has already reduced the number of test compounds. In general, the more complex an assay the more stringent the filter criteria related to it, i.e., the filtering becomes more and more stringent along the screening cascade.

2.4 SPR Based Screening with Pharmaceutically Relevant Targets

In order to illustrate the theoretical considerations of a fragment screening effort we selected two targets, chymase and β-secretase (BACE) for a more detailed discussion of the experimental set-up.

2.4.1 Chymase

Binding experiments were performed with the wt-protein and the zymogen (an inactive proprotein) immobilized via amine coupling on a CM5 sensor chip. Figure 1 depicts typical binding curves monitored for the two proteins during the contact with solutions containing a positive control compound at different concentrations. The figure indicates that this set-up is an ideal filter to distinguish between selective active site binding and nonselective binding of compounds. Nonselective binding would lead to positive response in the channel with the proprotein as well as in the channel with the wt-protein. Based on the saturation response of about 60 RU and the response of about 6,000 RU monitored upon immobilization of the protein, the relative amount of active protein was estimated to be 66% considering the two different molecular weights of the protein (30,000 Da) and the positive control (456 Da). The equilibrium dissociation constant for the positive control compound was determined to be 290 nM. The Z' factor determined for this positive control was 0.83 indicating excellent quality data. For the determination of the robustness of the assay with fragments, one has however to consider a MW corrected Z' factor [39]. For the average molecular weight of 214 Da of the library screened the Z' factor is around 0.73. Figure 2 shows the results from reproducibility testing with the samples of one 96 well plate. The statistical data of the correlation of the responses from each plate indicate that the measurements are highly reproducible. The slope of the correlation is as expected 1.0 and the standard error is about 2.4.

Two thousand two hundred and twenty-six fragments were tested in the assay described above. Figure 3 shows the results of the screen of one 96-well plate in a
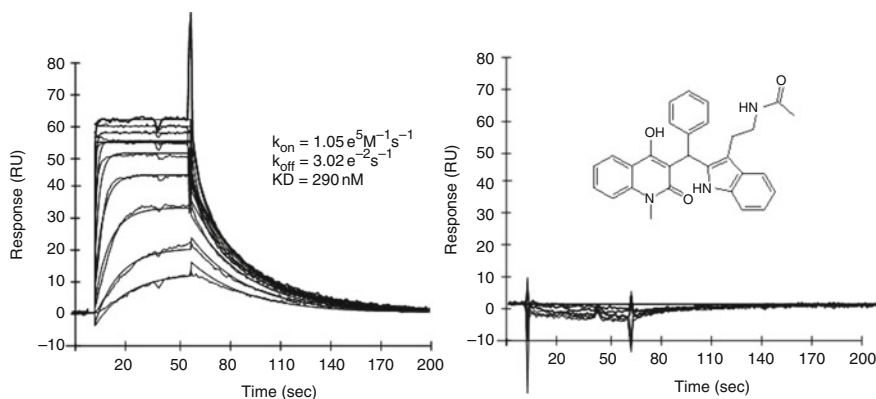
Fig. 1 Sensograms monitored from sensor surface with immobilized active chymase (*left*) and zymogen (*right*) in contact with solutions at different concentrations of the positive control (structure shown in the *inset*). This set-up is highly valuable to differentiate between binders that bind to active site (same pattern as for positive control) or to a different site (no response monitored from the surface with the immobilized zymogen). For the active protein the experimental response curves are overlaid with the theoretical curves obtained by fitting the experimental curves with the mathematical equations for a 1/1 kinetic model. Kinetic (k_{on} and k_{off}) as well as equilibrium binding parameters of the positive control given in the *inset* are extracted using this model
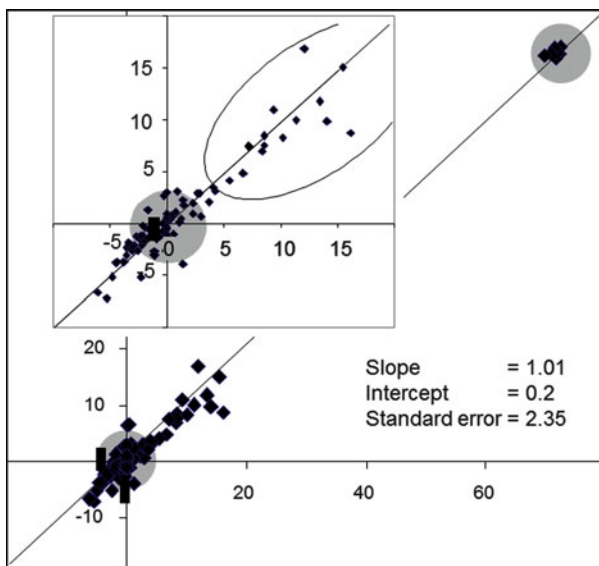
Fig. 2 The graph shows the reproducibility of the assay. 96 compounds are measured twice and the responses correlated with each other. Positives are marked in the *inset*
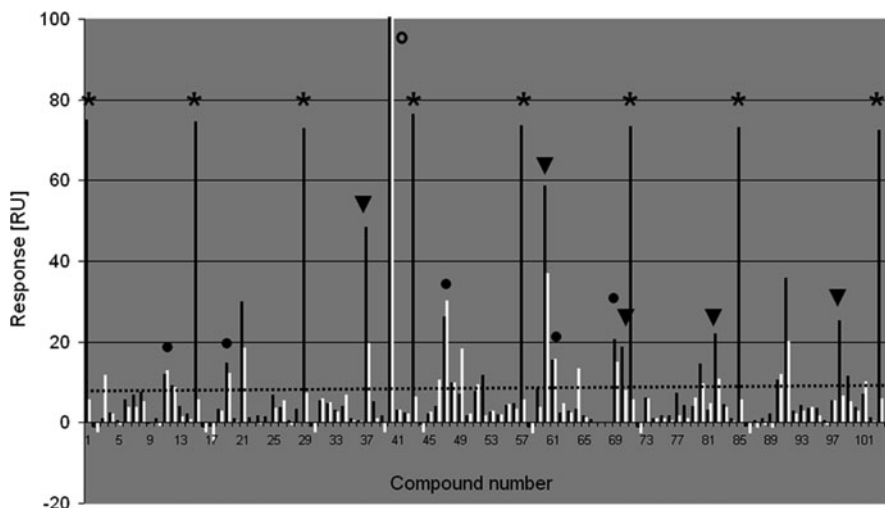
Fig. 3 Responses monitored in a screening set-up for compounds in a 96-well plate from the surface with active chymase (*black bars*) and with zymogen (*white bars*). Responses marked with a *star* are from injections of the positive control. The high quality of the assay is obvious from the amplitude of the signal as well as from the stable ratio of active and zymogen response. Signals from positive compounds showing selectivity are marked with *triangles*, signals from positive nonselective compounds with *filled circles*. The *dotted line* marks three times the standard deviations of the response of the negative control, and corresponds to the threshold for the positive hits

graphical representation with the responses at the report points as vertical bars. The plate contained 96 solutions of test compounds and 8 solutions of the control compound. In addition four negative controls (buffer with DMSO) were injected during the run. Figure 3 shows binding of many compounds to the immobilized chymase but only for a few of them a significant difference in binding between the wt-type protein and the inactive zymogen is observed, indicating selective binding to the active site.

Selection of the primary positives was first based on a promiscuity filter applying the criteria defined by Gianetti et al. [52], and by the affinity and the selectivity filter. Compounds were taken as positives if they exhibit no indication of promiscuity, show a response on the active protein that is higher than three times the standard deviation of the negative control and have a ratio of the responses on the active protein and the zymogen greater than two. One hundred and eighty fragments passed all the filters and were defined as positives.

In addition, these positives where confirmed in a competition assay with a positive control leaving 80 compounds for further characterization. The next validation step consisted of the determination of the K_D's via 10 point dose response experiments. This left 36 substances with well defined dose response in an affinity range from 10 to 60 μM for further characterization in X-ray crystallographic experiments.

2.4.2 BACE

A similar assay set-up was used for the fragment screening of BACE [11]. BACE was immobilized (12,000 RU) by standard amine coupling chemistry on a CM5 sensor. A mutant protein with the essential active site aspartate D39 mutated to alanine was used as a reference protein in a second channel. Figure 4 shows a typical sensogram monitored for the wild-type and mutant protein when contacted with a known high affinity (60 nM) small molecule inhibitor. The set-up is well suited as a selectivity filter, as compounds with selective binding to the active site of BACE show no or a reduced signal on the channel with the mutated protein. Figure 5 shows the screening results obtained from 96 compounds demonstrating the importance of such a selectivity filter for the BACE screen. Application of the affinity filter (response greater than 3 times the standard deviation) and the promiscuity filter alone would lead to a hit rate of about 60%, but the specificity filter that considers the ratio of the responses of wild-type and mutant protein reduces this number to 2.1%. It has to be mentioned in this context that a 60% hit rate without specificity filter is not frequently observed. This hit rate for primary positives depends on the screening concentration and the target protein. Whereas the screening concentration was not exceptionally high (250 μM) the properties of the target protein could be responsible for this high primary positive rate. The protein used for this screening was the full length protein that contains a hydrophobic membrane anchor and this area could be the source of the numerous unspecific positives.

All specific primary positives were confirmed in a competition assay using a known high affinity active site binder as competitor compound followed by dose response experiments to determine the K_D values.
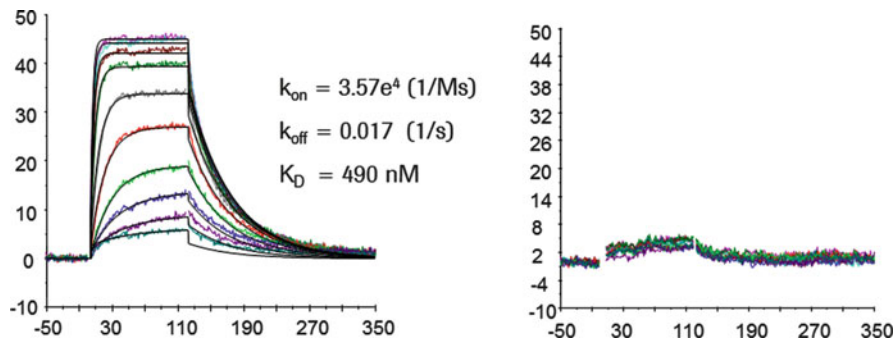
Fig. 4 Sensograms monitored from sensor surface with immobilized active BACE-1 (*left*) and blocked BACE-1 (*right*) in contact with solutions of different concentrations of the positive control (structure shown in the *inset*). This set-up enables differentiation of binders that bind to active site (same pattern as for positive control) or to a different site (no response monitored from the surface with the immobilized zymogen). For the active protein the experimental response curves are overlaid with the theoretical curves obtained by fitting the experimental curves with the mathematical equations for a 1/1 kinetic model. The kinetic and equilibrium binding parameters of the positive control given in the *inset* are extracted using this model
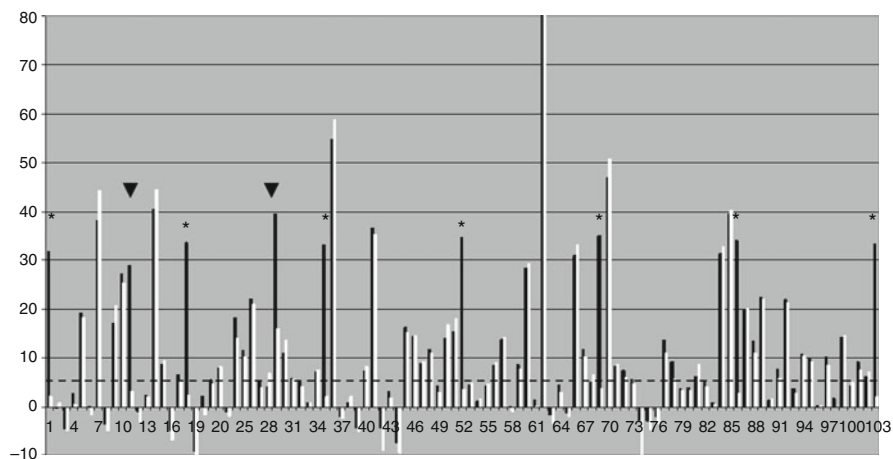
Fig. 5 Responses monitored in a screening set-up for compounds in a 96-well plate from the surface with active BACE-1 (*black bars*) and with active site mutated BACE-1 (*white bars*). Responses marked with a *star* are from injections of the positive control. The high stability of the set-up is obvious from the amplitude of the signal as well as from the stable ratio of active and blocked BACE-1 response. Signals from positive compounds showing selectivity are marked with *triangles*. The *dotted line* marks three times the standard deviations of the response of the negative control, and corresponds to the threshold for the positive hits

3   X-Ray Crystallography

Although direct crystallographic screening can be successfully applied for fragment screening, and offers a number of advantages, it is now less commonly used in this way compared with biophysical or biochemical assays that require less resource [14, 53]. However, most fragment based drug discovery programs that have advanced beyond mere screening have used structural biology [54] to drive hit progression. Indeed, only a few groups have applied the fragment approach to target classes like transmembrane proteins (e.g., GPCRs and ion channels) where protein structures are not easily accessible [55]. The additional information coming from the structures of hits in complex with their target helps to select the most promising candidates for subsequent fragment growth or fragment optimization. Structure based molecular modeling allows more efficient optimization of low affinity fragment hits to leads. Indeed for targets whose 3D structure is not available a fragment screening often is not considered at all. X-ray crystallography is the preferred biostructural technique, because it can be applied to most protein targets and delivers exact structural information for structure based optimization of chemical leads.

3.1   Prerequisites to Generate Fragment Complex Structures

To optimize the resources needed in following up a fragment screening with crystal structures, the setup of an efficient crystallographic workflow is important. This includes a good supply of crystallization grade protein, a reproducible crystal form diffracting to high resolution or robust soaking system, a reliable crystal harvesting procedure and an optimized X-ray data collection and structure determination process.

3.1.1   Protein

Generating suitable protein is often the most labor intensive step on the way to 3D structures. Enough protein to create hundreds of crystals is needed and therefore care needs to be taken with expression and purification procedures. High yield expression and simple and effective purification protocols are beneficial. Optimized protein constructs for crystallization often lack glycosylation sites and carry affinity tags for purification. All standard structural biology protein expression systems are used to produce the proteins for fragment cocrystallization, i.e., *E. coli*, Baculo virus insect cell systems or mammalian cell lines.

3.1.2   Crystallization System

Many crystals will be needed for determining complex structures of the typically 10^2 hits from a fragment screening. Reproducible production of well diffracting

crystals is therefore important for the crystallization system used and care is taken to optimize the crystallization procedure. To obtain cocrystal structures there are two options: soaking fragment hits into existing crystals and cocrystallization. Each has advantages and disadvantages. For soaking, the crystals can be produced in advance in a few crystallization experiments and low amounts of protein are consumed. Also higher ligand concentrations can be used during soaking to shift the binding equilibrium towards full occupancy. DMSO can frequently be added to concentration as high as 20% favoring solubilization of compounds. In contrast, addition of DMSO in ligand cocrystallization experiments often prevents crystal growth.

Disadvantages of soaking could be the target conformation in the crystal that may not be optimal for binding of a specific ligand. The crystal packing may hinder the diffusion of the ligand to the pocket as the fragment hit needs to be able to diffuse through the crystal solvent channels to the binding site. Therefore crystal packing differences are expected to influence the soaking success. The use of several different crystal forms with different packing of the molecules is one way to reduce false negatives.

During cocrystallization the complex is already formed in solution and the target protein is free to assume any conformation necessary to bind the specific ligand. However for cocrystallization the protein consumption may be higher and the maximum concentration of ligand is limited, because ligand and organic solvent may influence crystallization. Both soaking and cocrystallization methods require that the fragment binding site of the target is not blocked by the lattice packing of the protein target crystal and therefore some crystal forms of the target may not be suitable at all.

3.1.3 Diffraction Data Collection and Structure Determination

Following up a fragment screening with X-ray structures can involve collecting hundreds of datasets. Access to state of the art synchrotron beam lines equipped with modern fast detectors such as PILATUS [56] and automated sample changers such as CATS [57] greatly reduces the time needed for data collection. For example at the beam line X10SA at the SLS equipped with the PILATUS detector, about 60 data sets are now routinely collected per shift of 8 h and diffraction data for the hits from a typical fragment screening campaign can be collected in less than a day. Often data from the obtained diffraction images are processed by automated scripts that output difference electron density maps without the need for manual interference. For well behaving crystals the crystallographer's task is reduced to inspection of difference electron density maps, building and refining the model of the complex structure. Tracking the big number of crystallization, soaking and diffraction experiments done in parallel is a challenge in itself and book keeping is best managed with a Lab information management system (LIMS) that supports this comprehensive workflow [58]. In Pharma companies the resulting complex structures are deposited in in-house databases that are similar to the PDB, which however can be accessed by medicinal chemists easily e.g., by querying the ligand properties or generate superpositions based on the ligand binding

pockets (e.g., Proasis2® from http://www.desertsci.com). Last but not least the obtained cocrystal structures are communicated to the drug discovery chemists in front of the computer screen in a modeling session. For this purpose molecular graphics such as Moloc [59] and PyMOL (http://www.pymol.org) are used.

3.2 Determinants for Success in Cocrystallization

The success rate of getting cocrystal structures of fragment screening hits varies greatly. Whereas only very few hit structures were achieved in several independent fragment screenings on BACE (reviewed in [60]), for some other targets (e.g., chymase), cocrystal structures were obtained for about a third of the selected fragment screening hits [39]. What determines the differences in success rate has not yet been well assessed and is somewhat controversial. Here we want to list several factors without the claim for comprehension. (1) ability of the target or the ligand binding site investigated to bind small molecules (drugability) and the resulting potency of the fragment hits (2) the packing environment in the available crystal forms (3) the difference in solubility and binding affinity of ligands between the crystallization or soaking conditions and the assay conditions for the upstream fragment screening (SPR). Whereas little can be done for (1) and (2), the following paragraphs give some considerations of how to optimize the experimental set-up for (3).

3.2.1 Matching of Conditions for SPR-Screening/Cocrystallization

For fragments containing ionizable groups or interacting with acidic or basic groups of the target protein their protonation state greatly influences the K_D and, therefore, fragment binding is pH dependent. Differences in pH between the screening conditions and the crystallization or soaking conditions can lead to reduced or increased affinity of the fragment and failure to get complex structures. Fragment solubility is dependent on buffer pH and buffer composition. The precipitants in protein crystallization experiments are selected to reduce solubility of proteins and are, unfortunately, effective to small molecules as well. During the biophysical screening often organic solvents such as DMSO are present or detergents are added to increase the solubility of organic compounds. Such additives or solvents, however, could prevent growth or even dissolve crystals and are therefore often omitted from the crystallization experiment.

Ideally the conditions from which the cocrystal structures are obtained should be identical to those where the upstream screening experiment was performed. Matching the conditions between the primary screening and the crystallization or soaking experiment as closely as possible is one strategy to increase the yield of structures. If crystals do not grow at such conditions, the search for crystal soaking conditions that match the screening conditions can be tried. Another approach was used by AstaZeneca [44]. They used the surrogate protein endothiapepsin to get complex structures of BACE fragment screening hits as endothiapepsin crystallizes at pH 4.6 which is closer

to the acidic assay conditions. In contrast, BACE crystallization conditions have a neutral pH. If the crystallization conditions cannot be changed, it may be possible to run the primary biophysical screening assay at conditions like pH, buffer and salt concentration closer to those of the one suitable for the X-ray crystal system.

Overall, we have to accept that a perfect match of experimental conditions is not feasible and that a lack of hit confirmation may not result from an issue with a particular biophysical method. Further we need to accept that some valid hits will not be confirmed and, consequently, not considered for follow up work.

3.2.2 Prioritization of Ligands for X-Ray Experiments: K_D and Solubility

The cocrystallization and structure determination needs more time and resources than the primary screening methods like SPR. In order to limit the number of X-ray experiments, prioritization of the experiments is important. This enables a focus on the effort of crystallization experiments with those ligands where chances of complex structures are highest and to deprioritize experiments with ligands yielding less likely structures or not at all. Amongst others, binding affinity and solubility of ligands can be used as criteria to prioritize experiments.

The affinities of fragment screening hits range from a few µM to mM. Most fragment screening hits therefore have lower affinities than compounds from already advanced chemistry series or HTS with affinities in the nM to µM range. It is important to note that there seems to be no minimum affinity required for successful determination of complex structures and even mM compounds have been reported [11]. The experience of many fragment projects suggests that it takes more effort to get complex structures of low affinity fragments. The main reason could be the high compound concentration to be required for the experiment (about >10 times the K_D), which can result in concentrations as high as 10–50 mM which are often at the solubility limit of the compounds. Analysis of Roche fragment screening efforts indicated however that K_D alone was a better indicator of cocrystallization success than compound solubility (data not shown). The pH influences both affinity and solubility of the ligand. Besides matching the crystallization conditions to the assay conditions, the use of two or more independent X-ray systems with different crystal packing and different crystallization conditions and pH could also be expected to increase the yield of cocrystallization efforts and this was indeed the case in our labs (data to be published elsewhere).

3.2.3 Hit Expansion

Another approach to get more structural information from fragment screening hits is to use hit expansion [11, 44]. Hit expansion is a similarity search or in silico screening for potent analogs of the initial fragment screening hits from public or proprietary compound libraries. Application of synthetic chemistry by growing fragments (for example addition of solubilizing groups) or exchange of moieties or

single atoms is more resource intensive, but can quickly generate SAR information for fragments. In addition, further compounds for cocrystallization with lower K_D albeit with higher molecular weight and possibly lower ligand efficiency than the original screening hits are established. In the BACE fragment screening at Roche, additional binders were identified during hit expansion which subsequently delivered several structures that could be used for computer assisted molecular modeling [11].

## 3.3 Making Use of Structural Information in Synthetic Chemistry

Much has been reported about drug discovery facilitated by fragment screening and the transformation of fragments into clinical candidates and there are some excellent reviews on this subject. Here we focus on three examples because of their association with Roche to exemplify such a drug discovery effort.

One of the most intensively characterized targets regarding fragment screening is BACE, and many complex structures of different fragments targeting the active site of this aspartyl protease have been published [11, 44, 61–63]. The primary fragment screening methods included SPR, NMR, crystal soaking and computational methods. The resulting hits belong to different scaffolds. All hits hydrogen bond directly or indirectly to the catalytic aspartates, have hydrophobic interactions at the S1 pocket and often to one of the other subpockets in the BACE substrate binding site (Fig. 6a). Taken together they map the most tractable or drugable part of the BACE substrate binding pocket [60]. From the complex structures two binding hot spots could be identified, which are the side chains of the two catalytic aspartates 32 and 228 and the S1 pocket. Amines or other basic groups are observed binding the aspartates and always a benzyl ring filling S1. Subpocket S3 is frequently occupied, and a variety of hydrophobic groups are accepted there. From the structures of these fragments a common pharmacophore can be derived, which can guide the computer aided molecular modeling of BACE inhibitors with new chemical scaffolds. Further optimization and growth of the fragments by structure guided medicinal chemistry efforts resulted in potent inhibitors that extend to the prime side subpockets S1′ and S2′. One larger and more potent compound even displaces the Tyrosine sidechain of the flap loop opening up a new pocket that does not exist in the unliganded and most of the fragment complex structures (Fig. 6c; [60]).

Many cocrystal structures of fragments were obtained for the serine protease chymase [39]. The common feature of all fragments was an aromatic group binding to the S1 pocket and many fragments had an acidic group or oxygen atom in the oxyanion hole. The observation can be explained with the substrate specificity of chymase, which cleaves after aromatic side chains. The structures of the fragment screening hits highlight the importance of the S1 pocket and the oxyanion hole as hot spots for inhibitor binding. The different binding geometries exemplify the possibilities and limitations for groups fitting S1 and for possible exit vectors from S1 to the rest of the binding site (Fig. 7). The high hit rate suggested a good drugability of the target that was soon be confirmed by rapid progress in drug discovery.
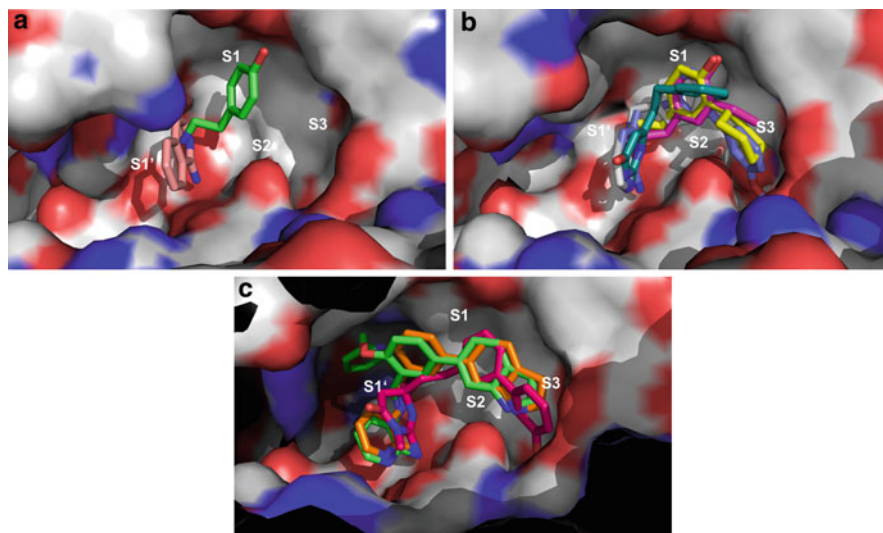
Fig. 6 (**a**) Fragments bound to the BACE active site aspartates (PDB entries 2OHK, 2BRA). (**b**) More structures after fragment hit expansion. Compounds occupy more of the BACE active site. (PDB entries 2OHM, 2OHQ, 3BUG, 3BUH, 2V00). (**c**) BACE inhibitor leads from fragments extend into prime sites (PDB entries 2OHT, 2OHU, 2VA7)
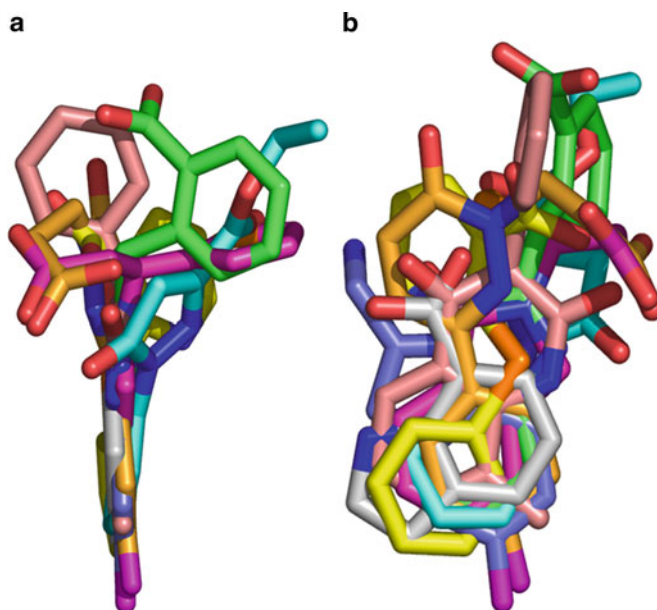
Fig. 7 Fragments bound to Chymase. The S1 pocket is always filled by aromatic rings, although these are not precisely oriented due to the lack of hydrogen bonds. Only the ring plane is very well conserved. Figure 7a relates to Fig. 7b by 90° rotation around the vertical axis

The example with best progress for a fragment screening derived lead compound is the B-Raf protein kinase inhibitor discovered at Plexxikon and shown to be successful in advanced clinical studies for Melanoma at Roche. In a fragment screening at Plexxikon with several kinases a nonspecific kinase binding fragment was found. Subsequently, selectivity was built into this novel lead series during fragment growth taking in the information from X-ray structures with several kinases into consideration. The lead compound PLX4720 binds to a pocket almost unique to the activated B-Raf. It is highly selective and shows nanomolar affinity for the oncogenic B-Raf (V600E) mutant. Studies in animal models have confirmed its therapeutic potential for treating B-Raf(V600E)-driven tumors [64].

4 Discussion and Conclusions

4.1 Combination of Efforts for Fragment Screening in a Seamless Workflow

There are numerous ways to establish a workflow for fragment screening that can successfully be applied in drug discovery projects. Today, most of the fragment screening efforts reported in literature are performed with a combination of biophysical methods. Figure 8 outlines one possible workflow as applied in a number of projects at Roche. A method with the ability of high(er) throughput like SPR is used for screening a fragment library of several thousand compounds, and hit confirmation is carried out with the same assay, as outlined in this review. The filtered and confirmed hits are further characterized by an orthogonal assay in order to improve the confidence that the fragment identified really binds to the target and to the binding site of interest. Here the role of X-ray crystallography is of extreme importance to visualize the binding of the fragment in detail and to facilitate analysis of the binding mode by computational chemistry. This fragment binding information leads to the establishment or refinement of pharmacophore models as well as gives insight into new patterns of interaction of small molecules with their protein targets. At this point, or earlier in the workflow after confirmed hits from the SPR screening are analyzed, a hit expansion can be performed by a similarity or pharmacophore search. Additional compounds from the internal library or purchased from external vendors are in our experience a great source for improving the chance of finding more potent ligands as well as to get X-ray structures before synthetic chemistry efforts are required. Several cycles of screening, hit analysis and characterization can be applied to a target and essential information for the drug discovery project derived.

There is no doubt that the results from a fragment screening effort influence the decision making of a project. In particular, the use of such data to inspire synthetic chemistry by identification of new binding scaffolds for a particular target is well established. Such information can be used to initiate novel lead series or to optimize chemical leads by replacement of a moiety. Another use of the results from a fragment
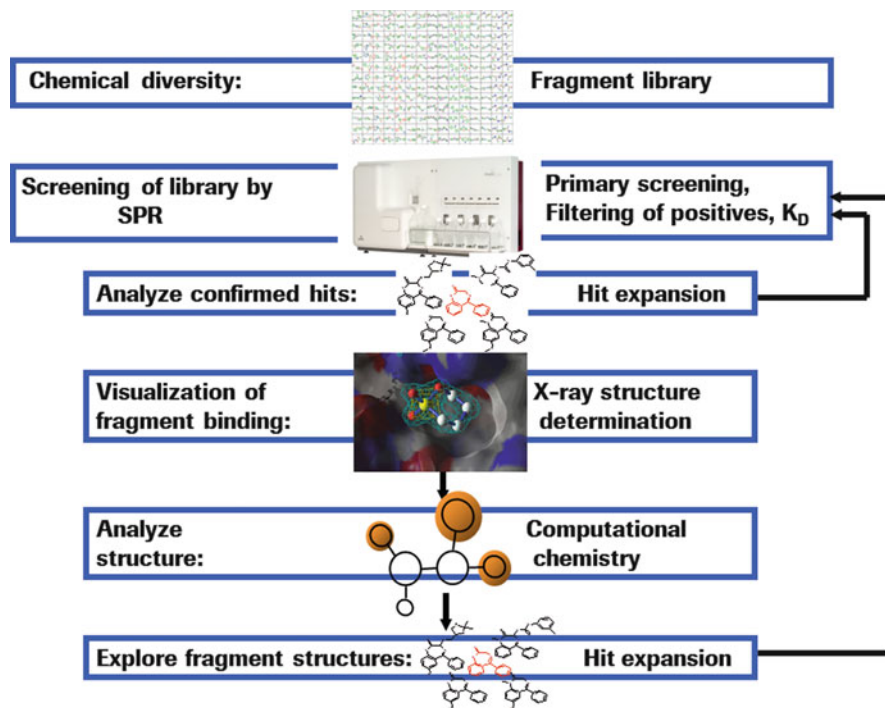
Fig. 8 Workflow for fragment screening used at Roche

screening is the assessment of the small molecule drugability of a target. The hit rate as well as the average ligand efficacy of the best fragments of a respective target gives a good indication about the effort required to identify potent small molecule ligands.

4.2 Outlook

There are still a number of ways to further improve the success of fragment screening efforts as defined as the identification and validation of novel binding motifs for a drug target in order to inspire synthetic chemistry efforts. Improvement of the fragment library (potential areas are solubility, structural diversity, compound purity etc.) and better alignment of assay conditions should both be considered. Addition to the workflow of further assay methods with a protein consumption, throughput and sensitivity profile similar to SPR, but without the need for immobilization (i.e., a homogenous assay) would also be welcomed.

We see the greatest value of this approach for novel drug targets with limited knowledge regarding small molecule ligands, targets with perceived low drugability or in projects with limited chemical space. This includes protein–protein interactions as well as targets like proteases or other enzymes.

Encouraging progress regarding the application of biophysical methods to transmembrane proteins including GPCR's will pave the way for the extension of the application of fragment screening to further targets classes.

References

1. Ringe D (1995) What makes a binding site a binding site? Curr Opin Struct Biol 5:825–829
2. Boehm HJ, Boehringer M, Bur D, Gmuender H, Huber W, Klaus W, Kostrewa D, Kuehne H, Luebbers T, Meunier-Keller N, Mueller F (2000) Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. J Med Chem 43:2664–2674
3. Erlanson DA, McDowell RS, O'Brien T (2004) Fragment-based drug discovery. J Med Chem 47:3463–3482
4. Rees DC, Congreve M, Murray CW, Carr R (2004) Fragment-based lead discovery. Nat Rev Drug Discov 3:660–672
5. Hubbard RE, Davis B, Chen I, Drysdale MJ (2007) The SeeDs approach: integrating fragments into drug discovery. Curr Top Med Chem 7:1568–1581
6. Hopkins AL, Groom CR, Alex A (2004) Ligand efficiency: a useful metric for lead selection. Drug Discov Today 9:430–431
7. Kuntz ID, Chen K, Sharp KA, Kollman PA (1999) The maximal affinity of ligands. Proc Natl Acad Sci USA 96:9997–10002
8. Dalvit D (2009) NMR methods in fragment screening: theory and a comparison with other biophysical techniques. Drug Discov Today 14:1051–1057
9. Bartoli S, Fincham CI, Fattori D (2006) The fragment-approach: an update. Drug Discov Today Technol 3:425–431
10. Barker J, Courtney S, Hesterkamp T, Ullmann D, Whittaker M (2005) Fragment screening by biochemical assay. Exp Opin Drug Discov 1:225–236
11. Kuglstatter A, Stahl M, Peters JU, Huber W, Stihle M, Schlatter D, Benz J, Ruf A, Roth D, Enderle T, Hennig M (2008) Tyramine fragment binding to BACE. Bioorg Med Chem Lett 18:1304–1307
12. Holdgate GA, Anderson M, Edfeldt F, Geschwindner S (2010) Affinity-based, biophysical methods to detect and analyze ligand binding to recombinant proteins: matching high information content with high throughput. J Struct Biol 172:142–157
13. Jhoti H, Cleasby A, Verdonk M, Williams G (2007) Fragment-based screening using X-ray crystallography and NMR spectroscopy. Curr Opin Chem Biol 11:485–493
14. Davies TG, Tickle IJ (2011) Fragment screening using X-ray crystallography. Top Curr Chem. doi:10.1007/128_2011_179
15. Dalvit C, Fogliattob G, Stewart A, Veronesia M, Stockman B (2001) WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability. J Biomol NMR 21:349–359
16. Jahnke W (2007) Perspectives of biomolecular NMR in drug discovery: the blessing and curse of versatility. J Biomol NMR 39:87–90
17. Shuker BS, Hajduk PJ, Meadows RP, Fesik AW (1996) Discovering high-affinity ligands for proteins: SAR by NMR. Science 274:1531–1534
18. Pellecchia M, Bertini I, Cowburn D, Dalvit D, Giralt E, Jahnke W, James TL, Homans SW, Kessler H, Luchinat C, Meyer B, Oschkinat H, Peng J, Schwalbe H, Siegal S (2008) Perspectives on NMR in drug discovery: a technique comes of age. Nat Rev Drug Discov 7:738–745

19. Rathore R, Corr JJ, Lebre DT, Seibel WL, Greis KD (2009) Extending matrix-assisted laser desorption/ionization triple quadrupole mass spectrometry enzyme screening assays to targets with small molecule substrates. Rapid Commun Mass Spectrom 23:3293–3300
20. Annis DA, Nickbarg E, Yang X, Ziebell MR, Whitehurst CE (2007) Affinity selection-mass spectrometry screening techniques for small molecule drug discovery. Curr Opin Chem Biol 11:518–526
21. Ladbury JE, Klebe G, Freire E (2010) Adding calorimetric data to decision making in lead discovery: a hot tip. Nat Rev Drug Discov 9:24–27
22. Freire E (2008) Do enthalpy and entropy distinguish first in class from best in class? Drug Discov Today 13:869–874
23. Lo M-C, Aulabaugh A, Jin G, Cowling R, Bard J, Malamas M, Ellestad G (2004) Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. Anal Biochem 332:153–159
24. Pantoliano MW, Petrella EC, Kwasnoski JD, Lobanov VS, Myslik J, Graf E, Carver T, Asel E, Springer BA, Pamela P, Salemme FR (2001) High-density miniaturized thermal shift assays as a general strategy for drug discovery. J Biomol Screen 6:429–440
25. Cimmperman P, Baranauskiene L, Jachimovičiute S, Jachno J, Torresan J, Michailoviene V, Matuliene J, Sereikaite J, Bumelis V, Matulis D (2008) A quantitative model of thermal stabilization and destabilization of proteins by ligands. Biophys J 95:3222–3231
26. Kussrow A, Enders CS, Morcos EF, Bornhop DJ (2009) Backscattering interferometry for low sample consumption molecular interaction screening. JALA 14:341–347
27. Markov DA, Swinney K, Bornhop DJ (2004) Label-free molecular interaction determinations with nanoscale interferometry. J Am Chem Soc 126:16659–16664
28. Zartler ER, Huaping M (2007) Practical aspects of NMR-based fragment discovery. Curr Top Med Chem 7:1592–1599
29. Klein J, Meinecke R, Mayer M, Meyer B (1999) Detecting binding affinity to immobilized receptor proteins in compound libraries by HR-MAS STD NMR. J Am Chem Soc 121:5336–5337
30. Gossert AD, Henry Ch, Blommers MJJ, Jahnke W, Fernández C (2009) Time efficient detection of protein–ligand interactions with the polarization optimized PO-WaterLOGSY NMR experiment. J Biomol NMR 43:211–217
31. Skinner AL, Laurence JS (2008) High-field solution NMR spectroscopy as a tool for assessing protein interactions with small molecule ligands. J Pharm Sci 97:4670–4695
32. Ross A, Schlotterbeck G, Klaus W, Senn H (2000) Automation of NMR measurements and data evaluation for systematically screening interactions of small molecules with target proteins. J Biomol NMR 16:139–146
33. Damberg ChS, Orekhov VY, Billeter M (2002) Automated analysis of large sets of heteronuclear correlation spectra in NMR-based drug discovery. J Med Chem 45:5649–5654
34. Johnsson B, Löfås S, Lindquist G (1991) Immobilization of proteins to a carboxymethyldextran-modified gold surface for biospecific analysis in surface plasmon resonance. Anal Biochem 198:268–277
35. Cunningham B, Lin B, Qiu J, Li P, Pepper J, Hugh B (2002) A plastic resonant optical biosensor for multiparallel detection of label-free biochemical interactions. Sens Actuators B 85:219–226
36. Huber W, Mueller F (2006) Biomolecular interaction analysis in drug discovery using surface plasmon resonance technology. Curr Pharm Design 12:3999–4021
37. Hämäläinen MD, Zhukov A, Ivarsson M, Fex T, Gottfries J, Karlsson R, Björsne M (2008) Label-free primary screening and affinity ranking of fragment libraries using parallel analysis of protein panels. J Biomol Screen 13:202–209
38. Nordström H, Gossas T, Hämäläinen M, Källblad P, Nyström S, Wallberg H, Danielson UH (2008) Identification of MMP-12 inhibitors by using biosensor-based screening of a fragment Library. J Med Chem 51:3449–3459

39. Perspicace S, Banner D, Benz J, Müller F, Schlatter D, Huber W (2009) Fragment-based screening using surface plasmon resonance technology. J Biomol Screen 14:337–349

40. Antonysamy SS, Aubol B, Blaney J, Browner MF, Giannetti AM, Harris SF, Hébert N, Hendle J, Hopkins S, Jefferson E, Kissinger Ch, Leveque V, Marciano D, McGee E, Nájera I, Nolan B, Tomimoto M, Torresa E, Wrighta T (2008) Fragment-based discovery of hepatitis C virus NS5b RNA polymerase inhibitors. Bioorg Med Chem Lett 18:2990–2995

41. Rich RL, Myszka DG (2010) Kinetic analysis and fragment screening with Fujifilm AP-3000. Anal Biochem 402:170–178

42. Rich RL, Quinn JG, Morton T, Stepp JD, Myszka DG (2010) Biosensor-based fragment screening using FastStep injections. Anal Biochem 407:270–271

43. Huber W (2005) A new strategy for improved secondary screening and lead optimization using high-resolution SPR characterization of compound–target interactions. J Mol Recogn 18:273–281

44. Geschwindner S, Olsson LL, Albert JS, Deinum J, Edwards PD, de BT, Folmer RH (2007) Discovery of a novel warhead against beta-secretase through fragment-based lead generation. J Med Chem 50:5903–5911

45. Johnsson B, Löfås S, Lindquist G, Edström A, Müller Hillgren RM, Hansson A (1995) Comparison of methods for immobilization to carboxymethyl dextran sensor surfaces by analysis of the specific activity of monoclonal antibodies. J Mol Recogn 8:125–131

46. O'Shannessy DJ, O'Donell KC, Martin J, Brigham-Burke M (1995) Detection and quantitation of hexa-His-tagged recombinant proteins on Western blots and by surface plasmon resonance technology. Anal Biochem 229:119–124

47. Wear MA, Patterson A, Malone K, Dunsmore C, Turner NJ, Walkinshaw MD (2005) A surface plasmon resonance-based assay for small molecule inhibitors of human cyclophilin A. Anal Biochem 345:214–226

48. Yoshitani N, Saito K, Saikawa W, Asanuma M, Yokoyama S, Hirotal H (2007) NTA-mediated protein capturing strategy in screening experiments for small organic molecules by surface plasmon resonance. Proteomics 7(494):9

49. Lue RYP, Chen GYJ, Qing Zhu YH, Yao SQ (2003) Versatile protein biotinylation strategies for potential high-throughput proteomics. J Am Chem Soc 126:1055–1062

50. Papalia GA, Giannetti AM, Arora N, Myszka DG (2008) Thermodynamic characterization of pyrazole and azaindole derivatives binding to p38 mitogen-activated protein kinase using Biacore T100 technology and van't Hoff analysis. Anal Biochem 383:255–264

51. Huber W, Perspicace S, Kohler J, Müller F, Schlatter S (2004) SPR-based interaction studies with small molecular weight ligands using hAGT fusion proteins. Anal Biochem 333:280–288

52. Giannetti AM, Koch BD, Browner MF (2008) Surface plasmon resonance based assay for the detection and characterization of promiscuous inhibitors. J Med Chem 51:574–580

53. Murray CW, Blundel TL (2010) Structural biology in fragment-based drug design. Curr Opin Struct Biol 20:497–507

54. Hajduk PJ, Greer J (2007) A decade of fragment-based drug design: strategic advances and lessons learned. Nat Rev Drug Discov 6:211–219

55. Früh V, Zhou Y, Chen D, Loch C, Eiso AB, Grinkova YN, Verheij H, Sligar SG, Bushweller JH, Siegall G (2010) Application of fragment-based drug discovery to membrane proteins: identification of ligands of the integral membrane enzyme DsbB. Chem Biol 17:881–891

56. Broennimann Ch, Eikenberry EF, Henrich B, Horisberger R, Huelsen G, Pohl E, Schmitt B, Schulze-Briese C, Suzuki M, Tomizaki T, Toyokawa H, Wagner A (2006) The PILATUS 1M detector. J Synchrotron Radiat 2006(Pt 2):120–130

57. Jacquamet L, Joly J, Bertoni A, Charrault P, Pirocchi M, Vernede X, Bouis F, Borel F, Périn JP, Denis T, Rechatin JL, Ferrer JL (2009) Upgrade of the CATS sample changer on FIP-BM30A at the ESRF: towards a commercialized standard. J Synchrotron Radiat 16:14–21

58. Haquin S, Oeuillet E, Pajon A, Harris M, Jones AT, van Tilbeurgh H, Markley JL, Zolnai Z, Poupon A (2008) Data management in structural genomics: an overview. Methods Mol Biol 426:49–79

59. Gerber PR (1992) Peptide mechanics: a force field for peptides and proteins working with entire residues as small unites. Biopolymers 32:1003–1017
60. Kuglstatter A, Hennig M (2010) Fragment based approaches for identification of BACE inhibitors. In: Varghese J (ed) BACE lead target for orchestrated therapy of Alzheimer's disease. Wiley, New Jersey, pp 107–121
61. Murray CW, Callaghan O, Chessari G, Cleasby A, Congreve M, Frederickson M, Hartshorn MJ, McMenamin R, Patel S, Wallis N (2007) Application of fragment screening by X-ray crystallography to beta-secretase. J Med Chem 50:1116–1123
62. Edwards PD, Albert JS, Sylvester M, Aharony D, Andisik D, Callaghan O, Campbell JB, Carr RA, Chessari G, Congreve M, Frederickson M, Folmer RH, Geschwindner S, Koether G, Kolmodin K, Krumrine J, Mauger RC, Murray CW, Olsson LL, Patel S, Spear N, Tian G (2007) Application of fragment-based lead generation to the discovery of novel, cyclic amidine beta-secretase inhibitors with nanomolar potency, cellular activity, and high ligand efficiency. J Med Chem 50:5912–5925
63. Congreve M, Aharony D, Albert J, Callaghan O, Campbell J, Carr RA, Chessari G, Cowan S, Edwards PD, Frederickson M, McMenamin R, Murray CW, Patel S, Wallis N (2007) Application of fragment screening by X-ray crystallography to the discovery of aminopyridines as inhibitors of beta-secretase. J Med Chem 50:1124–1132
64. Tsai J, Lee JT, Wang W, Zhang J, Cho H, Mamo S, Bremer R, Gillette S, Kong J, Haass NK, Sproesser K, Li L, Smalley KS, Fong D, Zhu YL, Marimuthu A, Nguyen H, Lam B, Liu J, Cheung I, Rice J, Suzuki Y, Luu C, Settachatgul C, Shellooe R, Cantwell J, Kim SH, Schlessinger J, Zhang KY, West BL, Powell B, Habets G, Zhang C, Ibrahim PN, Hirth P, Artis DR, Herlyn M, Bollag G (2008) Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. Proc Natl Acad Sci USA 105:3041–3046

Targeting Protein–Protein Interactions and Fragment-Based Drug Discovery

Eugene Valkov, Tim Sharpe, May Marsh, Sandra Greive, and Marko Hyvönen

Abstract Protein–protein interactions (PPI) are integral to the majority of biological functions. Targeting these interactions with small molecule inhibitors is of increased interest both in academia as well as in the pharmaceutical industry, both for therapeutic purposes and in the search for chemical tools for basic science. Although the number of well-characterised examples is still relatively modest, it is becoming apparent that many different kinds of interactions can be inhibited using drug-like small molecules. Compared to active site targeting, PPI inhibition suffers from the particular problem of more exposed and less defined binding sites, and this imposes significant experimental challenges to the development of PPI inhibitors. PPI interfaces are large, up to thousands of square angstroms, and there is still debate as to what part of the interface one should target. We will review recent developments in the field of PPI inhibition, with emphasis on fragment-based methods, and discuss various factors one should take into account when developing small molecule inhibitors targeted at PPI interfaces.

Keywords Biophysical screening · Fragment-based drug discovery · Inhibitor · Protein–protein interactions · Structural biology

Contents

E. Valkov, T. Sharpe, M. Marsh, S. Greive, and M. Hyvönen (✉)
Department of Biochemistry, University of Cambridge, 80 Tennis Court Road,
Cambridge CB2 1GA, UK
e-mail: marko@cryst.bioc.cam.ac.uk

1 Introduction

1.1 Motivation for Targeting Protein–Protein Interactions

One of the biggest challenges that the pharmaceutical industry faces is the decreasing rate of approval of novel drugs, despite increased expenditure in research and development [1]. The reasons for this trend are not entirely evident, but it is clear that novel strategies must be found to increase the number of novel drugs coming to the market. Increased use of fragment-based drug discovery (FBDD) methods has already made an impact and has reduced development time compared to more traditional methods [194]. It will also be necessary to re-evaluate the choice of target and, in particular, to explore how target space might be expanded. Analyses of known human target proteins of currently marketed drugs has revealed that, in comparison to the whole proteome, the efforts are centred on a very limited set of targets and target classes [2]. Even if the target space is expanded to include homologous proteins, it still constitutes only a fraction of the approximately 20,000 protein-encoding genes in the human genome, even without taking into account the further divergence of the proteome created by alternative splicing and/ or post-translational modification. The same analysis shows that novel target classes are identified at a reasonably constant rate of two per year [2]. We must therefore be more innovative in targeting novel classes of proteins, especially those without active sites that offer pre-existing small molecule binding sites; and also use novel biological and synthetic approaches to tackle traditional targets.

Protein–protein interactions (PPIs) are crucial for normal cellular function. In the past two decades, developments in high-throughput genetic (such as yeast two-hybrid) and biochemical (TAP-tagging and mass spectrometry) analysis of

protein–protein interactions have increased exponentially the volume of data concerning macromolecular interactions in living cells. Numerous recent investigations in genomics, proteomics and bioinformatics have provided convincing experimental evidence that PPIs participate in networks of interactions that may also involve other biological macromolecules [3–5]. This view of a cell as a massively interlinked, dynamic proteome is supported by the estimation that in the *Saccharomyces cerevisiae* proteome each protein has on average nine interacting partners [6, 7], and in the human proteome there are in the region of 400,000 PPIs [8]. PPIs are increasingly attracting attention as a new frontier in drug development, particularly where they are involved in regulation of cellular function and in disease states. At the same time, structures of larger, multiprotein complexes are being determined by X-ray crystallography, NMR and electron microscopy to provide atomic details of the interaction sites. These structures are complemented by data on the role of individual residues in disease states, provided by high-throughput genomic mapping and next generation sequencing [9].

A limitation of targeting small molecule binding sites in proteins lies in the restraints that a shared natural ligand imposes on evolutionary divergence of the binding site. For example, most kinase active sites share common structural features because they have all evolved to bind a common ligand, ATP. Therefore, development of specific inhibitors that do not cross-react with multiple kinases and other ATP binding enzymes can be challenging [10]. In contrast to protein–small ligand interaction sites, inhibition at protein–protein interaction sites by small molecules has been seen as challenging, if not impossible. The properties of PPI sites (as discussed in more detail in Sect. 2) differ substantially from small molecule binding sites, but with a limited number of well-characterised examples, rules are still being defined and work in this area is in its early stages. However, an increasing number of "biological" drugs, such as antibodies (many of which target protein–protein interaction sites) are being approved as drugs and thereby validate this mechanism of inhibition [11]. Small molecule replacements of these therapeutic agents have been extremely difficult to develop; either the necessary potency has been hard to achieve or the expected biological effect has not been obtained. The diversity of the composition at protein–protein interfaces is one aspect that makes PPIs such challenging targets for small-molecule intervention.

Despite the challenges, PPI targeting has the potential to bring about significant improvement in addressing challenging areas, not least by expanding the target space away from the "traditional" classes of drug targets. It should be stressed that targeting protein–protein interaction sites does not mean that the targets themselves need to be novel; for example, an existing validated target for which active site inhibitors show poor selectivity could also be targeted through inhibition of key PPIs crucial to its regulation.

Another advantage that targeting PPIs may bring is the potential increase in biological specificity of inhibition. Many signalling proteins participate in multiple pathways depending on the upstream signal, the cell type and the biological/physiological context, and distinct protein–protein interactions modulate this. It might, therefore, be possible to limit the adverse effects of the drug by selectively targeting one of the many functions the protein might have.

Although this review focuses on the development of drugs that can be used for the treatment of human diseases, PPI inhibitors hold great promise for basic science as chemical "tools". Epigenetic techniques such as siRNA have revolutionized cellular studies by allowing specific knock-down of target genes with relative ease, but since delivery methods are not sufficiently developed for this to be used in whole organisms, small molecule inhibitors can provide a complementary "chemical genetics" approach [12]. In this review, we will provide an overview of the specific features of protein–protein interactions sites in comparison with small molecule binding sites and known PPI inhibitors. We will discuss which features of the PPI site should be targeted and how this can be done effectively. We will also outline technical challenges specific to the design of protein–protein interaction inhibitors. Where possible we place our discussion of PPIs and their inhibition in the context of fragment-based approaches to drug discovery.

1.2 What is a Protein–Protein Interaction?

For the purpose of this review, we will limit our discussion to direct protein–protein interactions that are well characterized, are relevant for the biological function of the protein, and where structural information is available for the complex and the interactions sites. Much data from high-throughput proteomic analysis is at the level of biochemical observation and cataloguing of the interaction, hence, little or no information is known of the atomic details of many such interactions. However, a thorough structural understanding of the interaction is essential for FBDD.

1.3 Bioinformatic Analysis and Cataloguing of PPIs

Increasingly, there are efforts to classify and catalogue PPIs, using bioinformatic tools, into comprehensive databases [13–16]. However, unlike the repositories that hold, for example, structural or sequence information there is considerable debate about the source, quality and type of data that is assembled [17]. Although litera-ture-curated small-scale datasets are also available, most public databases have opted to catalogue experimental evidence that establishes PPIs without imposing "biased" selection criteria [18]. Potential users of such tools have to invest some time and effort in making an informed choice as to what databases are best for their needs, although the subtle differences between them may not always be immedi-ately apparent. An example of graphical output from the STRING database [19] for interaction partners of B-Raf is shown in Fig. 1; graphical illustration allows a more intuitive overview of the interactions, which can then be analysed in more detail by following the links provided in the graph.

In broad terms, there are three different approaches in the collection and presentation of interaction data: (1) primary databases, which include experimen-tally proven protein interactions coming from either small-scale or large-scale
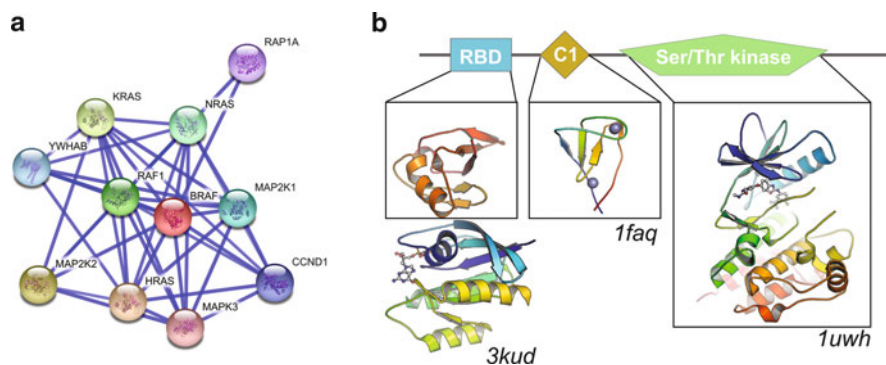
Fig. 1 Interaction network and domain structure of B-Raf. (**a**) Interaction network from the STRING database (http://string-db.org) for human proto-oncogene B-Raf. The *lines* indicate interactions between the proteins, with *thickness* of the lines reflecting confidence of the displayed interaction. (**b**) Domain structure of B-Raf and structures of individual domains, illustrating the different functional units that could be targeted by small molecule inhibitors: B-Raf RBD in complex with Ras (PDB code: *3kud*), diacylglycerol binding C1 domain (*1faq*) and the kinase domain in complex with active site inhibitor Sorafenib (*1uwh*)

published studies that have been manually curated; (2) meta-databases, which include only experimentally proven PPIs obtained by consistent integration of several primary databases (sometimes including small sets of original PPI data); and (3) prediction databases, which include mainly predicted PPIs, combined with experimentally proven PPIs [15, 16]. Thus, there is a clear distinction between the two types of data available for mining: one is experimental and the other is essentially theoretical, based on computational prediction, which may not be based on any evidence for a direct interaction, but instead inferred, for example, from genetic co-expression. Examples of collections of primary, experimental data are DIP [18], IntAct [20], and MINT [21]. APID [22, 23] and PINA [24] are databases with comprehensively integrated PPI experimental data. It is important to note that it has proved challenging to assign error rates to experimental data to limit the impact of false "hits". For example, computational estimates of the binary protein interactome in yeast have yielded 16,000–26,000 interactions [6], whereas some databases currently list more than 50,000 binary interactions in the same organism.

2 Features of Protein–Protein Interactions

2.1 Physical and Chemical Characteristics of PPIs

Protein–protein associations can be subdivided very broadly into two distinct categories: transient, non-obligate and permanent, obligate. The key feature of obligate complexes, which include many oligomeric species, is a stable association

between the individual partners, whereas in transient complexes association is typically a result of some activation process. The associations are further subjected to modulation by changes in temperature, pH and ionic strength as well as regulation by biological processes including post-translational modifications and interactions with cofactors and substrates. The two types of association can generally be distinguished by considering the physico-chemical features of the interfaces involved [25, 26].

The analysis of the buried surface area and composition of the interfaces of permanent and transient interactions reveals that stable, obligate complexes involve a significant proportion of their overall surface area in the association, with each monomer contributing up to 30% of its surface area, and the average buried area being in the region of 800 Å^2 per monomer [26, 27]. There also appears to be a fairly well-defined distribution of secondary structure elements in the interfaces of obligate complexes, with random coil and helical segments predominating [26]. Overall, hydrophobic amino acids tend to be over-represented, much like in the protein core [26, 28, 29]. Hydrophobic residues tend be clustered into patches of 200–400 Å^2 rather than evenly distributed all over the interface [30]. In contrast, the interfaces of transient complexes demonstrate significantly greater variability both in the type of amino acid residues present and their secondary structure [31]. Analysis of the topological features of interfaces demonstrates that shape complementarity plays a significant role in permanent associations, whereas in non-obligate complexes complementarity is significantly reduced [26, 31]. Water molecules may also be found in the cavities of the protein interfaces [32, 33]. The stabilizing effect of the ordered solvent along the protein interface is exerted by various means including the formation of hydrogen bonds, and by increasing the shape and charge complementarity of the interface [34–36].

Many different chemical interactions combine to favour PPI. Covalent bonds are not commonly observed in PPI due to a biological requirement for reversibility. However, covalent modification (i.e. phosphorylation, acetylation) is frequently used in biology to expand the repertoire of moieties available for interaction and to alter the overall shape of interacting proteins (i.e. ubiquitylation, glycosylation) [37]. Also, inhibition by covalent modification has been employed frequently in targeting proteases, and rapid and reversible crosslinking of thiol-containing compounds and cysteine residues in proteins has also been employed in fragment screening (Sect. 4.8).

A range of weaker electronic interactions are observed at PPI interfaces: van der Waal's forces, hydrogen bonding, charge–charge interactions and cation–π interactions [25]. In all cases, the energy of interaction is scaled by the dielectric constant, an important factor in the buried core of PPI interfaces, where the exclusion of bulk solvent can give a lower local dielectric constant (by a factor of as much as 10–20) and thus higher interaction energies than in free solution [38].

Hydrogen bond networks present at protein–protein interfaces are formed predominantly by residue side-chains [25, 39], and tend to be sub-optimal in energy terms due to distorted geometry imposed by the physical association as well as by

the association with surrounding solvent molecules [31, 36]. There is significant variability in the charge distribution for PPI interfaces [36, 39–42], so general features are hard to establish. However, electrostatic interactions are thought to control specificity of association and are also significantly more amenable to external, environmental regulation, which would permit some measure of control over the lifetime of the complex association [43–45]. Van der Waal's forces, hydrogen bonds and cation–π interactions are short-range (<5 Å) due to exponential scaling of interaction energy with the separation of interacting groups. However, the energy of charge–charge interactions scales linearly with distance and so long-range interactions can be important, for example in funnelling one protein partner towards the binding site of another [46–49].

There are several important entropic contributions to the stability of PPIs. Given the predominance and clustering of hydrophobic residues at interface regions, hydrophobic interactions are likely to be a major contributing force to the stabilization of protein interfaces [31, 50, 51], particularly in the case of permanent associations [26]. A significant favourable contribution to the free energy of binding (ΔG_{bind}) is thought to come from the large positive entropy change (ΔS) associated with the release of water from clathrate solvent cages around hydrophobic surfaces when those surfaces bind to each other. This "hydrophobic effect" can be modelled by the free energy of transfer for amino acids from the aqueous to organic phase, and has a complex dependence on temperature and solution conditions (due to the effects of both on the hydrogen bonding, mobility and surface tension of bulk and interfacial water) [52, 53]. Analysis of structural and calorimetric data on nearly 100 unique protein–ligand complexes in the SCORPIO database [54] has allowed assessment of the relative importance of burial of hydrophilic (polar) and hydrophobic (apolar) surface area in determining ligand selectivity and affinity, respectively.

On first inspection, PPIs might appear to involve the docking of pre-formed, rigid interfaces. However, a protein may undergo many complicated changes in translational, rotational and vibrational entropy when participating in a PPI. Conformational changes (with concomitant changes in rotational freedom, flexibility and diffusion coefficient) may occur on a range of scales: from the rotation of individual bonds, to local refolding of the binding partners, the re-positioning of secondary structure elements or domains, and changes in the oligomeric state of a protein [25–27, 55]. Those conformational changes may be key, for example, to regulation of macromolecular assemblies, to creation of a catalytic site, or to a functionally important plasticity of ligand binding. An emerging theme in PPIs is involvement of one or more partners that are wholly or partly intrinsically unstructured (i.e. having a peptide-like heterogeneity of conformation in solution); in some cases a complex ordering process analogous to protein folding may take place during binding. The inherent malleability of such PPIs confers a biologically useful heterogeneity and promiscuity to the binding interactions of proteins involved in multiple pathways in the cell.

2.2   Energetics of PPIs

In biochemical literature, binding affinity is commonly expressed as the dissociation constant K_D (the reciprocal of the association constant K_A), which can be expressed in terms of the equilibrium concentration of interacting species and from which the change in standard Gibbs free energy for binding (G_{bind}^0) may be calculated:

$$A + B \rightleftharpoons AB$$
$$K_D = [A][B]/[AB]$$
$$\Delta G_{bind}^0 = RT\ln K_D = RT\ln(k_{off}/k_{on})$$

Biologically relevant binding affinities are generally much higher than those observed in supramolecular chemistry, having a range of K_D on the order of 10^{-4}–10^{-14} M [56]. For the simple binary association above, assuming that one component (A) of a complex is titrated in considerable excess against the other (B), the K_D is the concentration of A at which [B] = [AB] i.e. 50% saturation of B. K_D is related via rate equations to the second-order rate constant for association (k_{on}) and the first-order rate constant for dissociation (k_{off}). For biological systems, values of k_{on} fall in the range $<10^3$–10^9 M^{-1} s^{-1} [57], and typically in the range 10^5–10^6 M^{-1} s^{-1} in the absence of the electrostatic steering (higher k_{on}) or conformational change upon association (lower k_{on}).

$$\Delta G_{bind} = \Delta H_{bind} - T\Delta S_{bind}$$

The Gibbs equation relates the change in free energy of binding to the changes in enthalpy (ΔH_{bind}) and entropy (ΔS_{bind}) for binding. In most cases, ΔH_{bind} and ΔS_{bind} are temperature-dependent for biomolecular binding due to a non-zero change in heat capacity for binding (ΔC_p). ΔC_p is empirically well-correlated with the change in solvent-accessible surface area upon binding, which can be explained by reference to changes in the water structure around hydrophobic surfaces with temperature.

Binding sites for protein–protein interactions may have a large surface area (600–4,500 Å^2) and involve a large number of residues [25], particularly by comparison with traditional drug targets such as enzyme active sites. Even the highest resolution structural data does not necessarily illuminate which residues or clusters of residues contribute most to binding energy and so which parts of a PPI interface ought to be targeted to ensure the highest likelihood of disrupting complex formation. If an interface presents several potentially druggable pockets, how can one determine which should be the highest priority to achieve the goal of inhibition? The key to answering this question is to analyse the contribution to the free energy of binding made by different interacting moieties in the protein and ligand.

Physical chemical measurements and site-directed mutagenesis evidence from studies on protein folding and PPI have given values for the typical contributions to the free energy of different types of interaction. For example, burial of hydrophobic surface area contributes ~20–25 cal/mol/Å^2 [58–61], each methylene group in a crystalline hydrocarbon contributes ~2 kcal/mol of van der Waals energy, burial of a single methyl group 1–1.8 kcal/mol [62, 63], neutral hydrogen bonds typically contribute 0.5–1.8 kcal/mol, buried hydrogen bonds involving one or more charged groups 3–5 kcal/mol [64], 0.1–1.5 kcal/mol for a solvent exposed salt-bridge [65–68], and 0.4–1.1 kcal/mol for a cation–π interaction involving lysine [69].

To estimate the free energy of binding, one also needs to have an idea of the unfavourable contributions to binding (i.e. it is incorrect to start at a ΔG_{bind} of zero and add up the contributions from favourable interactions). Examples of unfavourable contributions to ΔG_{bind} are loss of conformational and translational entropy, along with endothermic contributions arising from the need to desolvate polar functionalities on the protein and ligand. The loss of entropy upon binding of a small molecule to a protein has been estimated to contribute 5–25 kcal/mol to ΔG_{bind} [70–72] and the loss of entropy upon dimer formation results in a free energy penalty of approximately 15 kcal/mol [58].

2.3 Alanine-Scanning Mutagenesis and Binding Hotspots

The most common experimental approach to characterisation of the energetics of a PPI interface is alanine scanning [73]. Site-directed mutagenesis is used to alter the gene encoding a target protein to introduce single-point mutations in the protein sequence. By mutating a residue at a given position to alanine and measuring the change in the free energy of binding ($\Delta\Delta G_{\text{bind}}$), the energetic contribution of the original side-chain to binding can be calculated from $\Delta\Delta G_{\text{Bind}}^{\text{Mut-WT}} = RT\ln (K_D^{\text{Mut}}/K_D^{\text{WT}})$[74]. Alanine is chosen rather than glycine (the simplest amino acid) because mutation to glycine introduces an unusually large degree of freedom into peptide backbone conformation, which can greatly destabilise protein structure. It is also worth noting that alanine scanning may stabilise or destabilise the bound and unbound states of a protein in a variety of ways that may compromise clear interpretation of the data [75]. For mutants that show a particularly large $\Delta\Delta G_{\text{bind}}$ it is advisable to check for gross changes in structure or stability that might affect binding. This can be assessed by measuring the change in the thermal denaturation midpoint with respect to the wild-type protein, or comparing the far-UV circular dichroism spectrum or the one-dimensional homonuclear NMR spectrum with that of the wild-type protein.

Techniques used to measure $\Delta\Delta G_{\text{bind}}$ include, but are not limited to, radio-activity-based assays, fluorescence emission, fluorescence anisotropy, fluorescence resonance energy transfer, isothermal titration calorimetry, surface plasmon resonance, NMR titration, analytical ultracentrifugation and multi-angle laser light scattering. In each case, a signal that changes upon binding is measured as one

partner is titrated against the other. The data can be fitted to a Langmuir isotherm (or a modified version thereof) to extract the K_D.

The first systematic application of alanine scanning to a PPI was for the characterisation of the interface between the human growth hormone hGH and its receptor hGHbp [73, 74, 76]. In an interface comprising 30 side chains from each protein, two tryptophan residues had $\Delta\Delta G_{bind} > 4.5$ kcal/mol, four hydrophobic residues had $\Delta\Delta G_{bind}$ of 1.5–3.5 kcal/mol, and five charged residues had $\Delta\Delta G_{bind}$ of 1–2 kcal/mol. The tryptophans denote a binding "hotspot", occupying a central hydrophobic region, and surrounded by more hydrophilic residues with water molecules trapped in cavities in the interface such that a cross-section of the interface resembled the cross-section of a folded globular protein [59].

Systematic reviews of data from >3,000 alanine mutants from a range of PPI interfaces [77, 78] in the ASedb (alanine scanning energetics) database [77] show that many of the features of the hGH–hGHbp interface are common to other PPI interfaces. Contributions to ΔG_{bind} are not equally distributed across interfaces, but are concentrated in hotspots ($\Delta\Delta G_{bind} > 2$ kcal/mol) that comprise approximately 10% of residues in an interface. There is little correlation at side-chain level with accessible surface area (ΔASA) and $\Delta\Delta G_{bind}$ but the distribution is skewed towards higher $\Delta\Delta G_{bind}$ at lower ΔASA, reflecting an increased burial of hydrophobic surface area seen for binding interfaces at a whole-surface level. In most cases, hotspot residues are located in complemented protrusions and pockets on opposing surfaces [79, 80]. Hotspots are enriched in tryptophan, tyrosine and arginine residues, which may reflect the pluripotency of interaction types made by the polar and hydrophobic moieties of each of those residues: for example, a single tryptophan to alanine mutation could remove substantial buried hydrophobic area, disrupt van der Waals interactions, delete hydrogen bonds to the indole nitrogen, and remove cation–π interactions. Leucine, valine, serine and threonine are disfavoured in hotspots, which may simply reflect the relatively minor chemical changes upon mutation to alanine by comparison with other residue types, which might impose a lower maximum achievable $\Delta\Delta G_{bind}$. It may be that a more accurate assessment of the energetic effect of a mutation would be obtained by normalizing $\Delta\Delta G_{bind}$ by the number of heavy atoms deleted, in a manner analogous to the calculation of ligand efficiency for small-molecule binding thermodynamics.

The hotspot residues are frequently surrounded by a ring of more polar residues that acts as an "O-ring" to exclude bulk water from the immediate environment of the hotspot [77]. Exclusion of water from the binding interface is thought to be entropically favourable, and exclusion of solvent dipoles lowers the local dielectric constant for the hotspot, increasing the energetic contribution of electrostatic interactions. The 600 Å^2 lower limit on PPI interface area [26] may represent the smallest interface required to exclude bulk solvent. Mutation of residues in O-ring regions may not result in large values of $\Delta\Delta G_{bind}$ since the Cβ and backbone atoms of alanine may still be sufficient to exclude solvent, especially if there is some local repacking. Hence, solvent exclusion from the surface can be viewed as a cooperative function and even the mutation of multiple residues to alanine (alanine shaving) [81] may have little effect on $\Delta\Delta G_{bind}$.
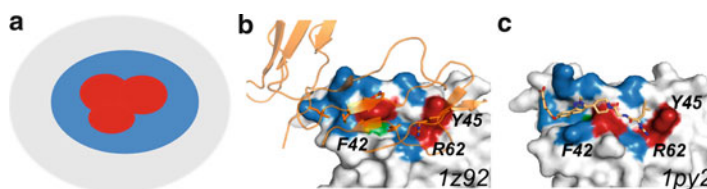
Fig. 2 Hotspot on protein binding sites and the hotspot residues on IL-2 binding site for its receptor and its inhibitor. (**a**) Protein surface (in *light gray*) with PPI site in colour. Hotspot residues (*red*) are typically found in the middle of the interface, surrounded by a ring of typically more polar residues (*blue*) that provide less energy to the binding, but help to create a low dielectric environment for the more hydrophobic central interactions. (**b**) Surface representation of IL-2 bound to its receptor (*orange cartoon*, PDB code: *1z92*) with hot spot residues coloured in *red* and labelled. Those interfacial residues that contribute less to the binding energy are coloured in *blue*. Residue 69, which showed marginal increase of affinity to both the receptor and inhibitor when mutated to alanine, is in *green*. (**c**) Complex of IL-2 with its inhibitor SP4206 (*1py2*). Colouring as in (**b**)

The potential for hotspots in natural ligand interactions to be targeted as sites for small molecules is shown by studies of the high-affinity small molecule SP4206 [82] that inhibits binding of interleukin 2 (IL-2) to the interleukin 2 α receptor (IL2-αR) (Fig. 2). Alanine-scanning mutagenesis of IL-2 [83] showed that the hotspots of binding affinity were similar for the interaction with IL2-αR and SP4206, although SP4206 was designed without reference to the IL-2/IL2-αR complex structure. However, structural analysis shows that although the regions of contact overlap, the nature of the interactions is different. The moieties of IL2-αR and SP4206 that interact with IL-2 share only very limited chemical similarity, and sterically incompatible conformations of the IL-2 interface are stabilised by each binding partner, in line with the known plasticity of the binding interface (Fig. 2b, c). Although hotspot analysis of IL-2 identified a suitable small-molecule binding site, structural analysis of the natural ligand interaction and attempts precisely to mimic that interaction might not have led to a potent compound. Perhaps hotspot analysis and structural characterisation of natural ligand–receptor interactions would best serve to guide, but not to limit, small-molecule development.

3 Modulating Protein–Protein Interactions

3.1 Protein–Protein Interaction Mimetics

Competitive inhibition of protein function is traditionally achieved at active sites using molecules that masquerade as enzyme substrates, or allosterically by small molecules that bind to cavities with recognition characteristics similar to active sites. It is not clear how to effectively inhibit PPIs with high affinity and selectivity using small molecules. In terms of competitive inhibition, a small molecule must

cover 800–1,100 Å^2 of a protein surface and must complement the poorly defined projection of hydrophobic and charged domains on a flat or moderately curved surface. The principal theme of the "proteomimetic" approaches is to develop a scaffold capable of mimicking the three-dimensional display of side chains from a structural element, typically an α-helix. Although this strategy was conceived and implemented initially due to the general sense that protein–protein interactions might not be amenable to inhibition by small molecules, it is now viewed in the field as synergistic with small-molecule drug discovery. Although the potency of the proteomimetics is not usually as high as has subsequently been obtained with the most effective small molecule antagonists, especially in vivo, the principles developed in the course of their study and development may yet prove useful in other contexts. Recent developments suggest a dichotomy among protein–protein interactions, with one class amenable to disruption by traditional medicinal chemistry methods and another class requiring alternative strategies [84, 85]. A number of proteomimetic strategies together with their biological applications are summarised in Table 1.

3.2   Features of Small Molecule PPI Inhibitors

Traditional medicinal chemistry efforts are increasingly being shown to be successful in generating low molecular weight compounds that target PPIs and that appear to show the expected effects in vivo. Virtually all the PPIs that have been successfully disrupted by small molecules have crystallographic or NMR structural data for the protein–protein complex (IL-2/IL2-Rα, Bcl-XL/Bad, MDM2/p53, CD80/CD28, S100B/p53, TNF trimer, XIAP/Smac, B-catenin/Tcf4, ZipA/FtsZ, cMyc/Mac, E1/E2, iNOS dimer and UL42/HSV-Pol) or for one of the complex components (CMR1) [104–109]. Some of the best structurally characterised PPI inhibitors and targeted PPIs are illustrated comparatively in Fig. 3. There is a strong correlation between the "druggability", a term frequently used to define success in finding small molecules binding to the interface, and the structural complexity of the PPIs. Complexes where one of the partners is a short, linear epitope appear to be more amenable for effective targeting. The obvious conclusion to be derived from this is that such interfaces are more druggable than the interfaces from globular constituents, as the existence of one partner that becomes ordered on binding allows a larger interaction surface between ligand and protein and often better formed pockets [110]. In addition, these complexes may be more amenable to the development of scalable competitive binding assays to identify small molecule inhibitors.

PPI modulators tend to be large, lipophilic molecules, which have more rings and less rotatable bonds than average for drugs and ligands from the Protein Data Bank (PDB) [111]. Molecules in the TIMBAL database of PPI inhibitors are on average the same size as drugs, but have on average fewer hydrogen bond donors and acceptors, and make more hydrophobic interactions than other small

**Table 1** Mimetic strategies employed in modulating protein–protein interactions

Proteomimetic class	Design strategy	Biological application	Additional notes	Reference
Constrained peptides	α,ω-Diaminoalkanes connecting glutamates spaced at the i; $i + 7$ positions with aminoisobutyric acid constraint	Disruption of the hexameric gp41 core assembly that leads to HIV-1 viral fusion with host cells	Micromolar IC_{50} values achieved in cell–cell fusion assay and measurable inhibitory activity in a luciferase-based viral infectivity assays	[86]
	Constrained peptide through olefin metathesis	Activation of apoptosis by inhibition of Bcl-x_L	Increased affinity over unstapled peptides, proteolytic stability, cell-permeability and activity in cellular assays	[87]
	Hydrogen-bond surrogate approach	Activation of apoptosis by inhibition of Bcl-x_L/Bak interaction	Metabolically stable peptides	[88]
	Dibenzofuran scaffold-based β-turn mimetic	Inhibition of WW domain of PIN1, a mitosis cell cycle regulator	Improved solubility and resistance to aggregation without compromising thermodynamic stability	[89]
	D-Pro-L-Pro and 2-aminocyclohexanoic acid-based β-turn mimetic	Incorporated into RNase A using expressed protein ligation	Semi-synthetic enzyme retains full catalytic activity and gains conformational stability	[90]
	Cyclic peptide predisposed to the β-turn/sheet conformation using a D-Pro-L-Pro hairpin template	Antagonist of the interaction between p53 and hMD2	Inhibition of the p53/hDM2 interaction with the IC_{50} values in the low nanomolar range	[91]
Foldamers (β-peptides)	3-Trifluoromethylphenyl or a 6-chlorotryptophan in place of the tryptophan in the recognition epitope	Inhibition of p53/hMD2 and p53/hDMX interactions	Nanomolar inhibitors of hDM2 binding, a tenfold improvement over our previously reported β3-peptide ligands	[92]
	Binding epitope presented on one face of a short 14-helix stabilized by side chain–side chain salt bridges	Inhibitors of HIV viral fusion that act by blocking gp41 assembly	Molecules bind in vitro to a validated gp41 model and inhibit syncytia formation in cell culture	[93]
	Mixed α-β peptides	Inhibitors of the Bcl-2/Bak interaction	All-β-peptide oligomers are poor inhibitors, possibly due to imperfect matching of side chains to the binding cleft of the target Bcl-2	[94]

(continued)

Table 1 (continued)

Proteomimetic class	Design strategy	Biological application	Additional notes	Reference
Secondary structure mimetics	Pyrrolinone-based low-molecular weight β-sheet mimetics	β-Strand proteomimetic inhibitor of HIV-1 protease	High proteolytic stability and good pharmacokinetic properties	[95]
	Terphenyl scaffold to present functionalities of the i, $i + 3(4)$ and $i + 7$ residues of an α-helix	Inhibitor of calmodulin and small muscle myosin light chain kinase interaction	Inhibited the calmodulin-mediated activation of 3′,5′-phosphodiesterase with IC_{50} values in the low nanomolar range	[96, 97]
	Oligoamide-, terephthalamide-, oligourea- and biphenyldicarboxamide-based α-helical mimetics	Bcl-x$_L$/Bak interaction inhibition	Mimetics could successfully displace the BH3 domain of the Bak peptide from the Bcl-X$_L$/Bak BH3 complex, with K_i values in the low micromolar range	[98, 99]
Surface mimetics	β-Cyclodextrin dimers linked on the secondary face by a pyridine-2,6-dicarboxylic group	Selective disruption of protein oligomerisation	Citrate synthase and L-lactate dehydrogenase catalytic activities inhibited with low micromolar potency	[100]
	Dipicolylamines linked by a substituted bipyridine to orient the metal–ligand arms	Recognition of phosphorylated peptides	Increases in helicity (stability) of the peptides were observed for phosphorylated targets, whereas no changes observed for non-phosphorylated peptides	[101]
	Central calixarene scaffold with four peptide loop domains based on a cyclic hexapeptide	Blocking the binding of platelet-derived growth factor to its receptor tyrosine kinase	Good selectivity and micromolar potency and significant inhibition of tumour growth in xenograft model	[96]
	Cyclophane-based resorcinarene trimer bearing a dansyl moiety and 28 carboxylate residues	Fluorescent probes capable of sensing post-translational modifications of histone	Exhibited potent discrimination capability between histone and chemically acetylated histone due to the electrostatic interactions	[102]
Allosteric inhibitors	Cell-based screening of a library of pyrimidine imidazoles based on known haem ligand	Inhibition of dimerisation of inducible nitric oxide synthase (iNOS)	Low nanomolar inhibitor of monomeric iNOS, but inactive against dimeric iNOS in enzymatic assays, with allosteric role confirmed crystallographically	[103]
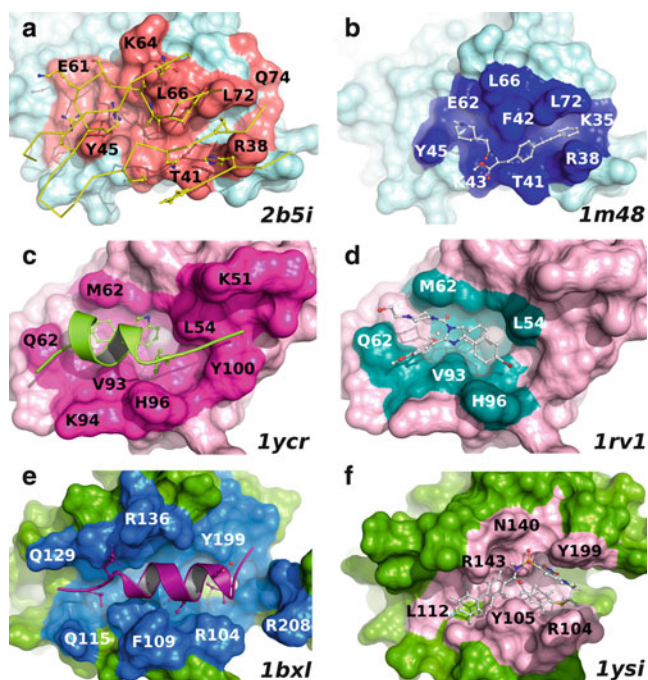
Fig. 3 Comparison of protein–protein and inhibitor interaction interfaces of IL-2, MDM2 and Bcl-X$_L$. Some of the better validated PPI inhibitors are compared to their natural ligand partners to highlight the extent of interaction surfaces and structural changes that take place on binding different ligands. (**a**) Structure of IL-2 (*cyan surface*) bound to the IL-2 α receptor (*yellow ribbon trace*). The residues of IL-2 that form interactions with the receptor are coloured *red* on the surface, with some of the key residues labelled. (**b**) IL-2 bound to inhibitor of receptor interaction. IL-2 residues that form interactions with the receptor are coloured *blue* on the surface. (**c**) Structure of MDM2 (*pink surface*) bound to a peptide from the transactivation domain of p53 (*green ribbon*) with interaction site coloured in *purple* on MDM2 and key interacting residues labelled on the surface. (**d**) Structure of the MDM2 inhibitor Nutlin-2 (in ball-and-stick) bound to MDM2 with interaction surface coloured in *cyan*. (**e**) Solution structure of Bcl-x$_L$ (*green surface*) complexed with a 16-residue peptide derived from the BH3 region of Bak (*magenta ribbon*) with interaction surface coloured in *light blue*. (**f**) Complex of ABT-737 bound to Bcl-x$_L$ (*green surface*) with interaction surface coloured *light purple*. In this particular example significant conformational rearrangement of the target protein is observed. PDB codes for the structures are shown under each figure

molecule–protein complexes [111]. Although the surface buried by PPI modulators is higher on average than for other small molecules, this is due to their larger overall size, and the ratio between buried and molecular surface is very similar. In a drug discovery context, molecules with such a profile could be considered potentially promiscuous and therefore unattractive. In terms of chemical functionality, small molecule PPI modulators contain more carboxylic acids and sulphonamides and fewer ether groups than drugs. One surprising observation is the high nitro-group content of PPI modulators. In general, aromatic nitro groups are avoided in

drug development because of the potential for toxicity when the nitro group is reduced within the body [112]. The ligand efficiency (LE) threshold described by Wells and McClendon for the most optimized small molecule inhibitors of protein–protein interactions complemented by structural analysis is 0.24 kcal/mol per non-hydrogen atom. For the 76 targets currently held in the TIMBAL database, most reside within the range of 0.15–0.35, with an average of 0.27. This average is reached for the PPI modulators with an average of 30 atoms, which indicates that they are slightly less efficient than typical small molecule lead compounds with the same number of atoms [113].

3.3   Modulating PPIs at Allosteric Binding Sites

Peptidic mimetics and small molecules that act directly at the site of interaction provide the most direct approach to modulating PPIs. However, allosteric regulation with small molecules shows increasing promise as a mechanism for modulating PPIs [114]. Allosteric modulation of a PPI site has been shown with brefeldin A, a lactone antibiotic that is frequently used as a tool compound in cell biology. It binds to the interface of the small G-protein Arf and its nucleotide exchange factor Sec7 and inhibits the exchange reaction by locking the complex in a non-productive conformation [115]. This site has been targeted with small molecule LM11, which binds both Arf1/GDP and Arf1/Sec7 complexes producing an similarly inactive conformation for the proteins [116]. Taxol, an anticancer drug, affects the polymerisation and stability of microtubules and results in blockage of mitosis and cell division. Protein kinases, an increasingly important class of drug targets, are highly regulated enzymes and are often under allosteric control by interacting proteins [117]. Modulation of these regulatory interactions could be targeted using allosteric inhibitors to gain specificity over active site inhibitors [10].

A recent example of the discovery of specific allosteric inhibitors against farnesyl pyrophosphate synthase (FPPS) through fragment-based methods is a good illustration of the discovery and validation of an allosteric binding pocket and the development of inhibitors targeting this site [118]. An initial screen of a small library of 400 fragments against FPPS using NMR led to the identification of four low affinity binders. Competition studies revealed that most fragments were not competitive with a known active site inhibitor, zoledronic acid, but bound to FPPS independently. Crystallographic analysis confirmed that all fragments bound outside the active site, and away from the dimer interface. Iterative evolution of fragments produced novel allosteric non-bisphosphonate FPPS inhibitors with IC_{50} values of 200 and 80 nM. It is worth mentioning that in this particular case, high-throughput screening (HTS) had failed to identify any potential leads with greater than micromolar potency, which underscores the relevance and potential of fragment-based methods in tackling challenging targets.

4 Screening and Validation Techniques

As described above, there are several small molecule inhibitors of protein–protein interactions that showed promising results in early clinical trials; however, these have largely evolved out of HTS rather than from screening of fragment libraries [119–122]. Although FBDD for active site inhibitors (ASIs) is a well-validated approach ([123–126, 195], application of this type of drug discovery process to PPI is in its infancy, with limited information on the design and success of fragment library screens available in the public domain [127–130]. However, as with all drug discovery programs, FBDD for PPI requires an efficient and accurate first-round screening procedure that is followed by a hit validation phase during which the binding interaction between individual hit compounds and the target protein is rigorously characterised with biophysical and structural approaches. This information is then used to design molecules with increased binding affinity to the target during the second, lead optimisation, phase of the program. The techniques and experimental design approaches that have been previously used in first-round screening and hit validation are discussed in Sect. 4.1 and in more detail elsewhere in this issue [196, 197].

4.1 General Considerations in Assay Design

Assay development should take into consideration the biophysical and thermodynamic properties of the interaction partners, for instance these characteristics vary significantly for enzyme active site interactions and protein–protein interaction surfaces. The more enclosed nature and relatively small contact area of ASI binding sites has enabled successful screening of fragment libraries by direct binding approaches or structural approaches [131–134]. However, even for these relatively ideal situations, first-round hits from fragment libraries tend to bind within the active site with low affinity (K_D in the micromolar to millimolar range). Small molecules would be expected to bind at PPI sites with very low affinity, due to their solvent exposure and limited contact area. These weak interactions can only be reliably detected by assays that are both sensitive and have significant dynamic range over the optimal concentration of components used. Careful design is required to target specifically the surface features involved in the contact of interest when screening for potential inhibitors of PPIs that are also enzymes and therefore contain an active site. Assays that screen for potential inhibitors can be either direct, where binding between library members and the target is measured; or indirect where inhibition of target function (such as the PPI itself) is assayed in a competition-in-solution format.

4.2 Direct Binding Assays

Direct binding activity is quantified and compared using methods that include the structural approaches of NMR spectroscopy and X-ray crystallography, in addition to biophysical techniques such as non-dissociating mass spectrometry, surface plasmon resonance (SPR), isothermal titration calorimetry (ITC) and thermal stability assays. Since any catalytic site which may also be present is likely to contain binding sites for fragment compounds with equal or higher affinity than those that might inhibit the target interaction, it is essential that access to such sites are blocked with a tight-binding or irreversible inhibitor [135]. This will ensure that any hits (defined by one standard deviation in signal above the baseline) are more likely to specifically target the interaction surface of interest. Finally, some studies have identified hits from "slow" binders, fragments likely to require a conformational change in the target for binding to occur, and care must be taken to not to miss this category of positive hits [136, 137].

4.3 Competition-in-Solution Assays

Competition-in-solution assays can quantify the effect of library components on the interaction between components of a protein complex. These assays include traditional biochemical assays that measure interactions through antibody- or tag-mediated "pull-down" or label displacement assays, along with biophysical methods that measure changes in the property of a component upon binding to its partner protein, such as fluorescence spectroscopic assays and SPR. It is essential for assay development that the requirements (affinity and binding site) for assembly of the interaction partners be at least partially defined such that a tractable model system for the target interaction can be developed, as was the case for MDM2/p53, IL-2/IL2-αR and Bcl/Bak [138–140].

4.4 Model Interaction Systems

Since many of the proteins with PPIs identified as potential therapeutic targets are difficult to purify and/or are poorly soluble in physiological buffers, model interaction systems are often stripped down to the minimal constructs required for the binary interaction. These model systems may be the domain(s) or subdomain(s) that constitute the target interaction surfaces of the human binding partner proteins. For interaction surfaces that are confined to small contiguous regions of the protein, peptides may be used as the binding partner in these model systems, such as BH3 and p53 peptides binding to Bcl and MDM2, respectively [141, 142] (Fig. 3c, e). Some peptide sequences are insoluble in isolation and this property may be

improved by expressing the sequence in a scaffold display system. Thioredoxin, antibodies or recombinant phage virions are routinely used for peptide display in screening for sequences that improve or disrupt the binding interaction [143], and a thioredoxin "mini-protein" was used to study the interaction between MDM2 and p53 [144].

Model system design must also consider the method used to detect the interaction and ensure that the proteins or peptides can be tagged with fluorophores or attached to a surface without interfering with the PPI of interest. Since any protein surface included in the model will provide a potential binding surface for fragments, approaches using protein solubility tags, coupling tags or signal tags (e.g. MBP, GST, antibody epitopes, 6XHis, GFP or luciferase) need to include controls to identify fragment–tag interactions and eliminate corresponding hits [145]. Whatever the design or source, the model system components need to be soluble, stable, mimic the thermodynamic properties of the natural interaction and be available in large quantities of pure product.

4.5   Nuclear Magnetic Resonance Spectroscopy

Various NMR-based approaches have proven highly successful in both screening and hit validation for fragment-based inhibitors of PPIs (reviewed in [146, 147]). These may measure either changes in the chemical shift of labelled target protein upon ligand binding (so called protein-observe methods), or changes in nuclear Overhauser effects (NOEs) between the ligand and target upon binding (ligand-observe methods), Automated sampling, along with compound pooling and deconvolution methods, has allowed large libraries to be rapidly screened for hits. Small protein targets (<50 kDa) can be labelled with ^{15}N and screened for changes in ^{15}N/^{1}H chemical shift in the two-dimensional heteronuclear single-quantum coherence (2D HSQC) spectra upon fragment binding, providing detailed structural information on the binding site and affinity of the interaction, and has been routinely used for identification of active site inhibitors from fragment libraries [132]. This approach was used to identify hits from a fragment library that inhibited the FKBP/FK506 interaction [148], binding of IL-2 to IL2-αR [129], MDM2/p53 interaction [149], and Bcl/BH3 interactions [127]. Hits from the IL-2/IL2-αR and Bcl/BH3 interaction screens were validated and structurally characterised by 2D HSQC [127, 150]. Ligand-observe methods (e.g. saturation transfer difference and water-LOGSY [151]), are also widely used, as described for an HTS screen for inhibitors of the β-catenin/Tcf4 interaction [121]. The principal limitation of NMR methods is the requirement for high concentrations of pure protein and fragment, which may be difficult to make and/or insoluble; however, this characteristic also means that compounds with very weak affinity to the target protein are also identified.

4.6 Thermal Stability Screens

Thermal stability screens measure the increase in fluorescence intensity of a hydrophobic fluorescent probe that binds to proteins as they unfold during thermal denaturation. Automation of this assay in plate format has allowed rapid screening of libraries for compounds that improve the thermal stability profile of the target protein and has been particularly successful in identifying hits against active site pockets [152, 153], and has recently been demonstrated in screening for inhibitors of Bcl interactions and human MDM2/p53 [142, 154].

4.7 Surface Plasmon Resonance

SPR essentially measures the change in mass at the surface of a gold-coated "chip" as the components in the mixture flow across the surface and interact with the moieties covalently attached to this surface. PPIs have long been analysed this way [155] and recent advances in the sensitivity of this technology has allowed fragment libraries to be screened for direct binding to the active site in the protein of interest immobilised on the chip surface [134, 135, 156, 157]. SPR is a particularly useful method for hit validation and characterisation of the kinetic parameters for binding and dissociation events during hit-target interactions for PPIs against IL-2/IL2-αR, Tcf/β-catenin and Cbl/phosphopeptide [120, 128, 136, 158]. Thermodynamic parameters for these interactions can also be determined by van't Hoff analysis of SPR measurements taken at several different temperatures [159], an approach used to validate hits from an FBDD study against the IL-2/IL2-αR interaction [158]. A particularly nice demonstration of SPR-based kinetic and thermodynamic characterisation of fragment–target binding was used to rank compounds "grown" from the original fragment hit in the ligand binding active site of acetylcholine-binding proteins (AchBP) [160]. Competition-in-solution assays are possible with SPR technology; however, the relative molecular weights of the immobilised and solution components of the interaction, as well as the potential inhibitor, must be considered in the experimental design. Generally these assays are best performed with the lower molecular weight binding partner (peptides or ligands) immobilised and the target protein in solution with/without potential inhibitors [136, 137]. However, it is possible to reverse this orientation if the smaller binding partner is displayed on a larger scaffold, such as thioredoxin.

The target protein or peptide may be immobilised on the surface by direct covalent coupling or via a stable capture method (biotin–streptavidin, antibody capture or His-tag); however, the relative success of these will depend largely on the properties of the target protein. For instance, if the protein is very stable then covalent coupling will generate a chip that can be reused many times to give highly consistent data. However, unstable proteins may be better immobilised via a capture system that allows easy regeneration. Given that this is a highly sensitive kinetic measurement, any change in

refractive index will be recorded and may be misinterpreted as a small molecule binding event, including changes in the refractive index induced by buffer components such as DMSO or aggregates of poorly soluble compounds or proteins. Most false positive hits in SPR-based screening result from the latter situation and are identified in the hit validation phase by the superstoichiometric mode of binding [161].

4.8 Isothermal Titration Calorimetry

ITC measures the change in heating power required to keep the temperature of the sample and reference cells constant as the binding partner is added incrementally, or titrated, into the sample cell containing the target protein or vice versa. These measurements allow calculation of the enthalpy change, binding affinity and stoichiometry of the interaction, and, through the laws of thermodynamics, the change in entropy and the free energy. These precise measurements provide important information about the binding event, and along with detailed structural information, can be very useful for hit validation and lead optimisation, as was used for understanding MDM2/p53, Tcf/β-catenin and Human Papilloma virus (HPV) E1/E2 interactions [120, 121, 149, 162]. This approach was also used to confirm the SPR-based ranking of "grown" fragments for binding to AchBPs [160]. Although ITC is amenable to automation, this is a relatively slow-throughput method, and when run in a traditional format requires high concentrations and quantities of protein and ligand. It is therefore currently a less suitable method for screening large libraries of compounds compared with the other techniques described here [163].

4.9 Fluorescence Spectroscopy

The quantum properties of different fluorescent probes to enter a higher electronic energy state when excited by light in a particular wavelength range, followed by release of this energy as light with a longer wavelength, has been exploited to develop several sensitive and rapid biophysical assays [164]. This sensitivity requires low reagent concentrations, hence these methods have long been used to study the assembly and dissociation of protein complexes and are easily adapted to the process of screening for inhibitors of PPIs since these measurements can be made in plate format. Caveats to the development of these assays are that the assay buffer conditions and small molecules in the library are characterised spectroscopically for absorption and fluorescence properties that may decrease the sensitivity of the assay or skew the data acquired. Recent advances in fluorophore technology have increased quantum yields and extended the available excitation and emission spectra into the infrared spectrum, significantly improving signal-to-noise ratios for many applications of fluorescence spectroscopy [164]. Finally, the target protein(s) usually need to be chemically or biologically labelled with a fluorophore.

Fluorescence polarisation (FP) or anisotropy measures the decrease in rotational diffusion as the small target protein or peptide, labelled with a fluorophore, increases in size upon association with its binding partner. This method is routinely used to screen libraries for drugs that disrupt protein–protein interactions and to validate hits from other screens, including HPV E1/E2, c-myc/max dimerization, MDM2/p53, Bcl/BH3 and ZipA/FtsZ [119, 127, 130, 141, 162, 165–169], although not all these libraries were fragment-based. Unfortunately, fluorescence polarization assays are particularly sensitive to false negatives since light scattered by aggregates in the assay produces a signal that is similar to that of the fully bound complex [164].

Fluorescence intensity/quenching assays measure the change in the intensity of the light emitted by a fluorescent probe upon alteration of its local environment, such as occurs during assembly or dissociation of the PPI complex, or binding of the fluorophore, such as native tryptophan. Decreases in fluorescence intensity were used to validate the interaction between HPV E2 and small molecule inhibitors [162]. Although popular for the study of protein–nucleic acid interactions, this approach is more difficult in the case of PPIs since the probes need to be close to the interaction site and thus may alter target folding or target interaction. This assay is most useful in the rare cases where tryptophan residues are involved in the PPI, and the signal is not masked by other aromatic amino acids in the target complex or by the spectroscopic properties of the fragment library components.

Fluorescence correlation spectroscopy was recently used to screen a fragment library for several inhibitors of the interaction of c-Cbl with fluorescently labelled phosphotyrosine target peptides [136]. This method measures the increase in diffusion time for a fluorescent protein or peptide after binding to its partner and is less sensitive to signal artefacts introduced by contaminating fluorescence, quenchers or compound aggregates, but requires specialist equipment and training, in addition to advanced mathematical data-fitting algorithms.

Fluorescence resonance energy transfer (FRET) measures the efficiency of quantum energy transferred from one fluorescent probe to another, where it is released as light that is much longer in wavelength than the original excitation light. This efficiency is directly related to the distance and relative orientation of the transition dipoles of the two probes and is particularly useful for the study of PPIs because the probes are only bought close together upon binding of the interaction partners. FRET was used in HTS screens for inhibitors of HIV-1 fusion and PCKι/ Par6 interactions [170, 171]. A variation on this approach [Luminescene (L)-RET] was used to screen for natural product inhibitors of the interaction between bacterial sigma70 and RNA polymerase β subunit [172].

4.10 Tethering

This approach uses a native or introduced cysteine residue close to the target binding site to "tether" hits from a disulphide-containing fragment library, which are identified by mass spectrometry or NMR spectroscopy [173]. Tethering has
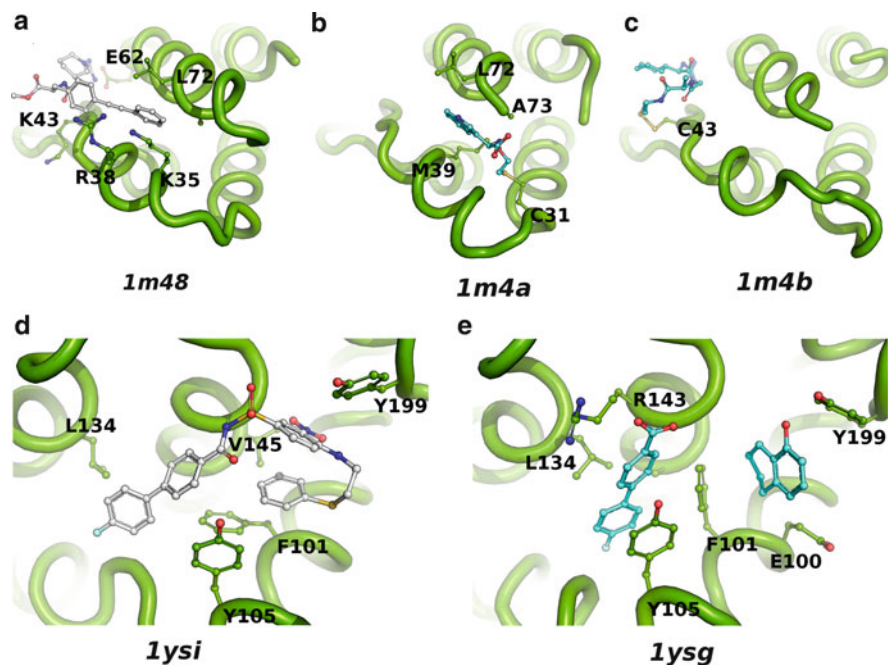
Fig. 4 Fragment optimization through tethering and linkage. (**a**) 8.2 μM inhibitor (compound 1, [158]) bound to IL-2. (**b**) Disulphide-tethered indole glyoxylate fragment bound to IL-2 on residue Y31C. (**c**) Disulphide-tethered guanidine fragment bound to IL-2 on residue K43C. (**d**) Bcl-X_L in complex with high affinity ligand ABT-737. (**e**) Fragment hits targeting BH3-binding in the groove of Bcl-X_L. 4′-Fluoro-biphenyl-4-carboxylic acid ($K_d = 0.30$ mM) on the *left hand side* of the groove and 5,6,7,8-tetrahydro-naphthalen-1-ol ($K_d = 4.3$ mM) on the *right side*. Key interacting residues in the target proteins are labelled and PDB codes for the structures are shown under each figure

mostly been used for active site screening [131, 174]; however, identification of PPIs for IL-2/IL2-αR have been successful with this method [82, 128, 158] (Fig. 4). Tethering is, however, limited compared to solution-based methods as it requires a suitably positioned cysteine in the protein, restricting the screening to a predefined area of the target site, while limitations in flexibility of the tethered ligand will influence the binding affinity.

4.11 X-Ray Crystallography

X-ray crystallography offers an unparalleled level of detail about the chemical environment of PPI sites and affords the opportunity to visualise compound binding to a target protein in atomistic detail. As such, it is a widely used technique in FBDD [175–179] and has become an integral part of the workflow of many

pharmaceutical companies e.g. Astex Therapeutics [180], Vernalis [177], and Johnson and Johnson [181].

However, its use as a tool in the screening and validation steps of a fragment-based project offers many technical challenges. A crystallisation strategy needs to be developed whereby structures of initial fragments and developed compounds in complex with the target protein can routinely be obtained. To achieve this, both co-crystallisation and crystal soaking techniques could in principle be used. Crystal soaking is preferable, especially if X-ray crystallography is to be used at the screening stage. Not only does this have practical advantages in terms of efficiency – the crystals can be grown in identical conditions and in advance – but it also allows higher concentrations of compounds to be used. This is especially advantageous for the characterisation of weak binding interactions often encountered with fragments [182]. For a crystal system to be suitable for soaking experiments, the binding site of interest needs to be both unoccupied and unobstructed by crystal contacts; the crystals need to be readily reproducible in large numbers; and the crystals need to be tolerant to the compound and co-solvent used. The maximal compound concentration that can be used in a soaking experiment is limited only by the solubility of the compound and the tolerance of the crystals.

Although X-ray crystallography can in principle be used as a tool in the initial fragment screening step, in practise it is more likely to be implemented following pre-screening with another technique [181, 183]. This offers advantages in terms of efficiency, but screening purely with X-ray crystallography offers the advantage of evaluating fragment hits based solely on their binding mode and not on potency. For example, a weaker potency fragment whose orientation offers more opportunity for fragment growing could be preferred over a tighter binding fragment that would be more difficult to develop. If X-ray crystallography is to be used primarily as a validation tool, the use of co-crystallisation is more tractable than if used for screening. The lower throughput requirements of a validation technique means the disadvantages of co-crystallisation, i.e. screening for new crystallisation conditions with each different compound are less problematic.

Hassell and coworkers present an overview of some of the strategies that can be utilised in obtaining the crystal structures of protein–ligand complexes [184]. One such strategy is "replacement soaking". When a particular target protein can only crystallise in the presence of a natural ligand or other inhibitor, the original ligand can be replaced by the compound of interest using soaking [185]. Although this technique is useful for small molecule binding sites it is likely to be more difficult for protein–protein interactions because a protein binding partner cannot easily be soaked out. A technique that we have found particularly useful is the use of the microseed matrix screening method combined with cross-seeding between different mutants [186] (Marsh and Hyvönen, manuscript in preparation). This has enabled us both to crystallise previously un-crystallisable mutants and to engineer novel crystal forms that are optimal for soaking (Fig. 5).

The targeting of protein–protein interactions offers additional challenges to the use of X-ray crystallographic screening compared to active site inhibition. It might be considered easier to soak a surface site as it is more accessible, but paradoxically
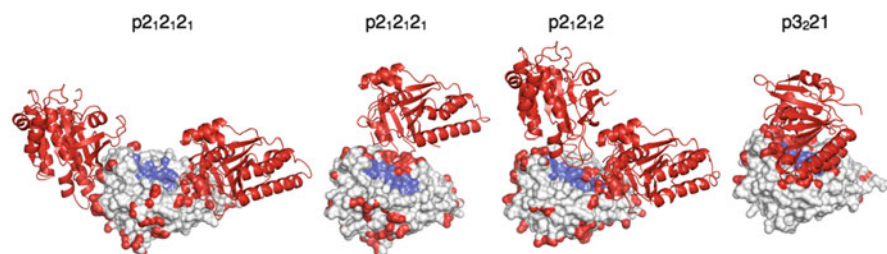
Fig. 5 Benefit of obtaining multiple crystal forms to aid with protein–ligand crystallographic work. Different crystal forms of mutant RadA arranged from the most open (*left*) to the most occluded (*right*) binding site (Marsh and Hyvönen, manuscript in preparation). Atoms involved in crystal contacts are coloured *red* and atoms that comprise the ligand binding site are shown in *blue*. The crystal symmetry mates closest to the binding sites are shown as cartoons. Space groups are indicated

the nature of crystallisation means that on average approximately 30% of the protein surface is involved in crystal contacts [187, 188]. The average size of a protein–protein interaction site is also much larger than that of an average protein small molecule binding site, 1,500–3,000 $\mathring{A}^2$ compared with 300–1,000 $\mathring{A}^2$ [108]. These facts decrease the probability of the protein–protein interaction site of interest being entirely free in a particular crystal form. Combined with the difficulty of obtaining soluble unliganded protein for PPI targets, this can make obtaining a suitable crystal form for screening particularly challenging. This problem is ameliorated somewhat if a hotspot (See Sect. 2.3) can be identified and targeted rather than the whole interface. Furthermore, in the early stages of an FBDD project the soaked fragments are (by definition) small in size and hence less affected by the crystal environment. As the fragments grow and cover a larger interface area, the limitation will become more apparent. But, at this stage co-crystallisation will become a more practical alternative due to increased affinity and reduced number of compounds compared to initial library.

Although crystal contacts can complicate fragment screening, the information gleaned from the interaction of molecules in a crystal lattice can also be exploited in the discovery of new binding sites. This was exemplified by identification of an activating interaction in an asymmetric epidermal growth factor receptor kinase dimer and by extension of an existing binding site in the polo-box domain (PBD) of polo-like kinase 1 [189, 190]. Sledz and coworkers analysed the crystal packing of number of different crystal forms of PBD and an unreported binding site was predicted based on crystal contacts in some of the structures. Through the use of biophysical techniques, mutagenesis and X-ray crystallography, this new binding site was shown to participate in binding to a natural ligand peptide. While this example shows the potential to identify cryptic binding sites on proteins, considerable effort has been made to distinguish genuine sites of PPI from artefacts of crystallization [191, 192]; i.e. crystal contacts represent a weak form of protein–protein interaction manifest only at the high concentrations present in crystal lattices (Fig. 5).

Even if it is not possible to use X-ray crystallography as a tool either through soaking or co-crystallisation during the screening and validation stages of a project, knowledge of the nature of the natural interface could still provide inspiration for the initial stages of drug design, i.e. through the development of peptide or peptide-like lead compounds.

5   Summary

Although there are still only a limited number of examples of small molecules that are well characterised and can effectively inhibit PPIs, it is clear that this can be achieved with relatively ligand-efficient molecules. Some of these inhibitors, such as Bcl-targeted ABT-263, have already reached clinical trials [193].

It is still difficult to draw clear conclusions as to which kind of interactions are easiest to inhibit and whether any special kind of chemistry is required for PPI inhibition in comparison to active site inhibition, and we have highlighted in this review some of the challenges and considerations for PPI inhibition – from choosing the target site to methods of screening and evaluating binders. As more examples of successful inhibition of PPIs are published, our understanding of the requirements for a good target and efficient ligand will be enhanced.

Fragment-based methods were employed against PPI targets very early on, and they are likely to be increasingly employed for this purpose. They are also increasingly used in academic settings and applied to ever more challenging targets, both for therapeutic intervention and for chemical tool development. With more examples accumulating, the chemical space we need to sample for this class of targets will be better defined, and focused fragment libraries with increased hit rates will be developed. Increased structural information will expand the target space further, and continually developing automation for crystallography, expedited data collection, and structure determination will make fragment-based approaches ever more appealing.

References

1. Pammolli F, Magazzini L, Riccaboni M (2011) The productivity crisis in pharmaceutical R&D. Nat Rev Drug Discov 10(6):428–438
2. Overington JP, Al-Lazikani B, Hopkins AL (2006) How many drug targets are there? Nat Rev Drug Discov 5(12):993–996
3. Schwikowski B, Uetz P, Fields S (2000) A network of protein-protein interactions in yeast. Nat Biotechnol 18(12):1257–1261
4. Schachter V (2002) Bioinformatics of large-scale protein interaction networks. Biotechniques 32 (3 suppl):S16–S27

5. Schachter V (2002) Protein-interaction networks: from experiments to analysis. Drug Discov Today 7(11):S48–S54
6. Grigoriev A (2003) On the number of protein-protein interactions in the yeast proteome. Nucleic Acids Res 31(14):4157–4161
7. Kumar A et al (2002) Subcellular localization of the yeast proteome. Genes Dev 16 (6):707–719
8. Ramani AK et al (2005) Consolidating the set of known human protein-protein interactions in preparation for large-scale mapping of the human interactome. Genome Biol 6(5):R40
9. Voisey J, Morris CP (2008) SNP technologies for drug discovery: a current review. Curr Drug Discov Technol 5(3):230–235
10. Fabian MA et al (2005) A small molecule-kinase interaction map for clinical kinase inhibitors. Nat Biotechnol 23(3):329–336
11. Hughes B (2011) Large drugs outdo small. Nat Biotechnol 29(4):296
12. Carlson SM, White FM (2011) Using small molecules and chemical genetics to interrogate signaling networks. ACS Chem Biol 6(1):75–85
13. Bossi A, Lehner B (2009) Tissue specificity and the human protein interaction network. Mol Syst Biol 5:260
14. Lehne B, Schlitt T (2009) Protein-protein interaction databases: keeping up with growing interactomes. Hum Genomics 3(3):291–297
15. Shoemaker BA, Panchenko AR (2007) Deciphering protein-protein interactions. Part I. Experimental techniques and databases. PLoS Comput Biol 3(3):e42
16. Shoemaker BA, Panchenko AR (2007) Deciphering protein-protein interactions. Part II. Computational methods to predict protein and domain interaction partners. PLoS Comput Biol 3(4):e43
17. Cusick ME et al (2009) Literature-curated protein interaction datasets. Nat Methods 6 (1):39–46
18. Salwinski L et al (2009) Recurated protein interaction datasets. Nat Methods 6(12):860–861
19. Szklarczyk D et al (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 39(Database issue): D561–D568
20. Aranda B et al (2010) The IntAct molecular interaction database in 2010. Nucleic Acids Res 38(suppl 1): D525–D531
21. Ceol A et al (2010) MINT, the molecular interaction database: 2009 update. Nucleic Acids Res 38(suppl 1):D532–D539
22. Hernandez-Toro J, Prieto C, De las Rivas J (2007) APID2NET: unified interactome graphic analyzer. Bioinformatics 23(18):2495–2497
23. Prieto C et al (2008) Human gene coexpression landscape: confident network derived from tissue transcriptomic profiles. PLoS One 3(12):e3911
24. Wu J et al (2009) Integrated network analysis platform for protein-protein interactions. Nat Methods 6(1):75–77
25. Lo Conte L, Chothia C, Janin J (1999) The atomic structure of protein-protein recognition sites. J Mol Biol 285(5):2177–2198
26. Jones S, Thornton JM (1996) Principles of protein-protein interactions. Proc Natl Acad Sci USA 93(1):13–20
27. Stites WE (1997) Protein-protein interactions: interface structure, binding thermodynamics, and mutational analysis. Chem Rev 97(5):1233–1250
28. Ivanov YD, Kanaeva IP, Archakov AI (2000) Optical biosensor study of ternary complex formation in a cytochrome P4502B4 system. Biochem Biophys Res Commun 273 (2):750–752
29. Ivanov YD et al (1999) The optical biosensor studies on the role of hydrophobic tails of NADPH-cytochrome P450 reductase and cytochromes P450 2B4 and b5 upon productive complex formation within a monomeric reconstituted system. Arch Biochem Biophys 362 (1):87–93

30. Lijnzaad P, Argos P (1997) Hydrophobic patches on protein subunit interfaces: characteristics and prediction. Proteins 28(3):333–343
31. Tsai CJ et al (1997) Studies of protein-protein interfaces: a statistical analysis of the hydrophobic effect. Protein Sci 6(1):53–64
32. Dall'Acqua W et al (1998) A mutational analysis of binding interactions in an antigen-antibody protein-protein complex. Biochemistry 37(22):7981–7991
33. Vaughan CK, Buckle AM, Fersht AR (1999) Structural response to mutation at a protein-protein interface. J Mol Biol 286(5):1487–1506
34. Janin J (1999) Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition. Structure 7(12):R277–R279
35. Larsen TA, Olson AJ, Goodsell DS (1998) Morphology of protein-protein interfaces. Structure 6(4):421–427
36. Xu D, Lin SL, Nussinov R (1997) Protein binding versus protein folding: the role of hydrophilic bridges in protein associations. J Mol Biol 265(1):68–84
37. Perkins JR et al (2010) Transient protein-protein interactions: structural, functional, and network properties. Structure 18(10):1233–1243
38. Fersht A (1999) Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. W.H. Freeman, New York, xxi, p 631
39. Xu D, Tsai CJ, Nussinov R (1997) Hydrogen bonds and salt bridges across protein-protein interfaces. Protein Eng 10(9):999–1012
40. Ivanov AS et al (2007) Protein-protein interactions as new targets for drug design: virtual and experimental approaches. J Bioinform Comput Biol 5(2B):579–592
41. Sheinerman FB, Norel R, Honig B (2000) Electrostatic aspects of protein-protein interactions. Curr Opin Struct Biol 10(2):153–159
42. Stevens JM, Armstrong RN, Dirr HW (2000) Electrostatic interactions affecting the active site of class sigma glutathione S-transferase. Biochem J 347(Pt 1):193–197
43. Vijayakumar M et al (1998) Electrostatic enhancement of diffusion-controlled protein-protein association: comparison of theory and experiment on barnase and barstar. J Mol Biol 278(5):1015–1024
44. McCoy AJ, Chandana Epa V, Colman PM (1997) Electrostatic complementarity at protein/protein interfaces. J Mol Biol 268(2):570–584
45. Camacho CJ et al (1999) Free energy landscapes of encounter complexes in protein-protein association. Biophys J 76(3):1166–1178
46. Neira JL, Vazquez E, Fersht AR (2000) Stability and folding of the protein complexes of barnase. Eur J Biochem 267(10):2859–2870
47. Buckle AM, Schreiber G, Fersht AR (1994) Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-A resolution. Biochemistry 33(30):8878–8889
48. Serrano L, Day AG, Fersht AR (1993) Step-wise mutation of barnase to binase. A procedure for engineering increased stability of proteins and an experimental analysis of the evolution of protein stability. J Mol Biol 233(2):305–312
49. Schreiber G, Fersht AR (1993) Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. Biochemistry 32(19):5145–5150
50. Wells JA (1996) Binding in the growth hormone receptor complex. Proc Natl Acad Sci USA 93(1):1–6
51. Tsai CJ et al (1996) Protein-protein interfaces: architectures and interactions in protein-protein interfaces and in protein cores. Their similarities and differences. Crit Rev Biochem Mol Biol 31(2):127–152
52. Kauzmann W (1959) Some factors in the interpretation of protein denaturation. Adv Protein Chem 14:1–63
53. Tanford C (1980) The hydrophobic effect: formation of micelles and biological membranes, 2nd edn. Wiley, New York, ix, p 233
54. Olsson TS et al (2008) The thermodynamics of protein-ligand interaction and solvation: insights for ligand design. J Mol Biol 384(4):1002–1017

55. Betts MJ, Sternberg MJ (1999) An analysis of conformational changes on protein-protein association: implications for predictive docking. Protein Eng 12(4):271–283
56. Li W et al (2004) Highly discriminating protein-protein interaction specificities in the context of a conserved binding energy hotspot. J Mol Biol 337(3):743–759
57. Schreiber G, Haran G, Zhou HX (2009) Fundamental aspects of protein-protein association kinetics. Chem Rev 109(3):839–860
58. Finkelstein AV, Janin J (1989) The price of lost freedom: entropy of bimolecular complex formation. Protein Eng 3(1):1–3
59. Chothia C (1974) Hydrophobic bonding and accessible surface area in proteins. Nature 248 (446):338–339
60. Schellman JA (1997) Temperature, stability, and the hydrophobic interaction. Biophys J 73 (6):2960–2964
61. Baldwin RL (1996) How Hofmeister ion interactions affect protein stability. Biophys J 71 (4):2056–2063
62. Jackson SE, Fersht AR (1993) Contribution of long-range electrostatic interactions to the stabilization of the catalytic transition state of the serine protease subtilisin BPN'. Biochemistry 32(50):13909–13916
63. Lorch M et al (1999) Effects of core mutations on the folding of a beta-sheet protein: implications for backbone organization in the I-state. Biochemistry 38(4):1377–1385
64. Fersht AR et al (1985) Hydrogen bonding and biological specificity analysed by protein engineering. Nature 314(6008):235–238
65. Strop P, Mayo SL (2000) Contribution of surface salt bridges to protein stability. Biochemistry 39(6):1251–1255
66. Luisi DL et al (2003) Surface salt bridges, double-mutant cycles, and protein stability: an experimental and computational analysis of the interaction of the Asp 23 side chain with the N-terminus of the N-terminal domain of the ribosomal protein l9. Biochemistry 42(23):7050–7060
67. Daopin S et al (1991) Structural and thermodynamic analysis of the packing of two alpha-helices in bacteriophage T4 lysozyme. J Mol Biol 221(2):647–667
68. Sali D, Bycroft M, Fersht AR (1991) Surface electrostatic interactions contribute little of stability of barnase. J Mol Biol 220(3):779–788
69. Prajapati RS et al (2006) Contribution of cation-pi interactions to protein stability. Biochemistry 45(50):15000–15010
70. Murray CW, Verdonk ML (2002) The consequences of translational and rotational entropy lost by small molecules on binding to proteins. J Comput Aided Mol Des 16(10):741–753
71. Chang CE, Gilson MK (2004) Free energy, entropy, and induced fit in host-guest recognition: calculations with the second-generation mining minima algorithm. J Am Chem Soc 126(40):13156–13164
72. Chang CE, Chen W, Gilson MK (2007) Ligand configurational entropy and protein binding. Proc Natl Acad Sci USA 104(5):1534–1539
73. Cunningham BC, Wells JA (1997) Minimized proteins. Curr Opin Struct Biol 7(4):457–462
74. Clackson T, Wells JA (1995) A hot spot of binding energy in a hormone-receptor interface. Science 267(5196):383–386
75. DeLano WL (2002) Unraveling hot spots in binding interfaces: progress and challenges. Curr Opin Struct Biol 12(1):14–20
76. Bass SH, Mulkerrin MG, Wells JA (1991) A systematic mutational analysis of hormone-binding determinants in the human growth hormone receptor. Proc Natl Acad Sci USA 88(10):4498–4502
77. Thorn KS, Bogan AA (2001) ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. Bioinformatics 17(3):284–285
78. Moreira IS, Fernandes PA, Ramos MJ (2007) Hot spots–a review of the protein-protein interface determinant amino-acid residues. Proteins 68(4):803–812

79. Keskin O, Ma B, Nussinov R (2005) Hot regions in protein–protein interactions: the organization and contribution of structurally conserved hot spot residues. J Mol Biol 345 (5):1281–1294

80. Li X et al (2004) Protein-protein interactions: hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: implications for docking. J Mol Biol 344(3):781–795

81. Jin L, Wells JA (1994) Dissecting the energetics of an antibody-antigen interface by alanine shaving and molecular grafting. Protein Sci 3(12):2351–2357

82. Braisted AC et al (2003) Discovery of a potent small molecule IL-2 inhibitor through fragment assembly. J Am Chem Soc 125(13):3714–3715

83. Thanos CD, DeLano WL, Wells JA (2006) Hot-spot mimicry of a cytokine receptor by a small molecule. Proc Natl Acad Sci USA 103(42):15422–15427

84. Hajduk PJ, Huth JR, Fesik SW (2005) Druggability indices for protein targets derived from NMR-based screening data. J Med Chem 48(7):2518–2525

85. Hajduk PJ, Huth JR, Tse C (2005) Predicting protein druggability. Drug Discov Today 10(23–24):1675–1682

86. Sia SK et al (2002) Short constrained peptides that inhibit HIV-1 entry. Proc Natl Acad Sci USA 99(23):14664–14669

87. Walensky LD et al (2004) Activation of apoptosis in vivo by a hydrocarbon-stapled BH3 helix. Science 305(5689):1466–1470

88. Wang D, Liao W, Arora PS (2005) Enhanced metabolic stability and protein-binding properties of artificial alpha helices derived from a hydrogen-bond surrogate: application to Bcl-xL. Angew Chem Int Ed Engl 44(40):6525–6529

89. Kaul R, Deechongkit S, Kelly JW (2002) Synthesis of a negatively charged dibenzofuran-based beta-turn mimetic and its incorporation into the WW miniprotein-enhanced solubility without a loss of thermodynamic stability. J Am Chem Soc 124(40):11900–11907

90. Arnold U et al (2002) Protein prosthesis: a semisynthetic enzyme with a beta-peptide reverse turn. J Am Chem Soc 124(29):8522–8523

91. Fasan R et al (2006) Structure-activity studies in a family of beta-hairpin protein epitope mimetic inhibitors of the p53-HDM2 protein-protein interaction. Chembiochem 7(3):515–526

92. Harker EA et al (2009) Beta-peptides with improved affinity for hDM2 and hDMX. Bioorg Med Chem 17(5):2038–2046

93. Stephens OM et al (2005) Inhibiting HIV fusion with a beta-peptide foldamer. J Am Chem Soc 127(38):13126–13127

94. Sadowsky JD et al (2007) (alpha/beta + alpha)-peptide antagonists of BH3 domain/Bcl-x(L) recognition: toward general strategies for foldamer-based inhibition of protein-protein interactions. J Am Chem Soc 129(1):139–154

95. Smith AB 3rd et al (1997) An orally bioavailable pyrrolinone inhibitor of HIV-1 protease: computational analysis and X-ray crystal structure of the enzyme complex. J Med Chem 40 (16):2440–2444

96. Blaskovich MA et al (2000) Design of GFB-111, a platelet-derived growth factor binding molecule with antiangiogenic and anticancer activity against human tumors in mice. Nat Biotechnol 18(10):1065–1070

97. Orner BP, Ernst JT, Hamilton AD (2001) Toward proteomimetics: terphenyl derivatives as structural and functional mimics of extended regions of an alpha-helix. J Am Chem Soc 123(22):5382–5383

98. Saraogi I, Hamilton AD (2008) alpha-Helix mimetics as inhibitors of protein-protein interactions. Biochem Soc Trans 36(Pt 6):1414–1417

99. Rodriguez JM et al (2009) Synthetic inhibitors of extended helix-protein interactions based on a biphenyl 4,4'-dicarboxamide scaffold. Chembiochem 10(5):829–833

100. Leung DK, Yang Z, Breslow R (2000) Selective disruption of protein aggregation by cyclodextrin dimers. Proc Natl Acad Sci USA 97(10):5050–5053

101. Ojida A et al (2003) Cross-linking strategy for molecular recognition and fluorescent sensing of a multi-phosphorylated peptide in aqueous solution. J Am Chem Soc 125(34):10184–10185
102. Hayashida O, Ogawa N, Uchiyama M (2007) Surface recognition and fluorescence sensing of histone by dansyl-appended cyclophane-based resorcinarene trimer. J Am Chem Soc 129(44):13698–13705
103. McMillan K et al (2000) Allosteric inhibitors of inducible nitric oxide synthase dimerization discovered via combinatorial chemistry. Proc Natl Acad Sci USA 97(4):1506–1511
104. Berg T (2008) Small-molecule inhibitors of protein-protein interactions. Curr Opin Drug Discov Devel 11(5):666–674
105. Fischer U, Schulze-Osthoff K (2005) Apoptosis-based therapies and drug targets. Cell Death Differ 12(Suppl 1):942–961
106. Fry DC (2008) Drug-like inhibitors of protein-protein interactions: a structural examination of effective protein mimicry. Curr Protein Pept Sci 9(3):240–247
107. Gonzalez-Ruiz D, Gohlke H (2006) Targeting protein-protein interactions with small molecules: challenges and perspectives for computational binding epitope detection and ligand finding. Curr Med Chem 13(22):2607–2625
108. Wells JA, McClendon CL (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature 450(7172):1001–1009
109. Whitty A, Kumaravel G (2006) Between a rock and a hard place? Nat Chem Biol 2(3):112–118
110. Blundell TL et al (2006) Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. Philos Trans R Soc Lond B Biol Sci 361(1467):413–423
111. Higueruelo AP et al (2009) Atomic interactions and profile of small molecules disrupting protein-protein interfaces: the TIMBAL database. Chem Biol Drug Des 74(5):457–467
112. Boelsterli UA et al (2006) Bioactivation and hepatotoxicity of nitroaromatic drugs. Curr Drug Metab 7(7):715–727
113. Hopkins AL, Groom CR, Alex A (2004) Ligand efficiency: a useful metric for lead selection. Drug Discov Today 9(10):430–431
114. Christopoulos A (2002) Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. Nat Rev Drug Discov 1(3):198–210
115. Mossessova E, Corpina RA, Goldberg J (2003) Crystal structure of ARF1*Sec7 complexed with Brefeldin A and its implications for the guanine nucleotide exchange mechanism. Mol Cell 12(6):1403–1411
116. Viaud J et al (2007) Structure-based discovery of an inhibitor of Arf activation by Sec7 domains through targeting of protein-protein complexes. Proc Natl Acad Sci USA 104(25):10370–10375
117. Pellicena P, Kuriyan J (2006) Protein-protein interactions in the allosteric regulation of protein kinases. Curr Opin Struct Biol 16(6):702–709
118. Jahnke W et al (2010) Allosteric non-bisphosphonate FPPS inhibitors identified by fragment-based discovery. Nat Chem Biol 6(9):660–666
119. Kiessling A et al (2006) Selective inhibition of c-Myc/Max dimerization and DNA binding by small molecules. Chem Biol 13(7):745–751
120. Lepourcelet M et al (2004) Small-molecule antagonists of the oncogenic Tcf/beta-catenin protein complex. Cancer Cell 5(1):91–102
121. Trosset JY et al (2006) Inhibition of protein-protein interactions: the discovery of druglike beta-catenin inhibitors by combining virtual and biophysical screening. Proteins 64(1):60–67
122. Vassilev LT et al (2004) In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. Science 303(5659):844–848
123. Buchwald P (2010) Small-molecule protein-protein interaction inhibitors: therapeutic potential in light of molecular size, chemical space, and ligand binding efficiency considerations. IUBMB Life 62(10):724–731

124. Congreve M, Murray CW, Blundell TL (2005) Structural biology and drug discovery. Drug Discov Today 10(13):895–907
125. Coyne AG, Scott DE, Abell C (2010) Drugging challenging targets using fragment-based approaches. Curr Opin Chem Biol 14(3):299–307
126. Rees DC et al (2004) Fragment-based lead discovery. Nat Rev Drug Discov 3(8):660–672
127. Lugovskoy AA et al (2002) A novel approach for characterizing protein ligand complexes: molecular basis for specificity of small-molecule Bcl-2 inhibitors. J Am Chem Soc 124 (7):1234–1240
128. Raimundo BC et al (2004) Integrating fragment assembly and biophysical methods in the chemical advancement of small-molecule antagonists of IL-2: an approach for inhibiting protein-protein interactions. J Med Chem 47(12):3111–3130
129. Tilley JW et al (1997) Identification of a Small Molecule Inhibitor of the IL-2/IL-2Rα Receptor Interaction Which Binds to IL-2. J Am Chem Soc 119(32):7589–7590
130. Tsao DH et al (2006) Discovery of novel inhibitors of the ZipA/FtsZ complex by NMR fragment screening coupled with structure-based design. Bioorg Med Chem 14(23):7953–7961
131. Erlanson DA et al (2003) In situ assembly of enzyme inhibitors using extended tethering. Nat Biotechnol 21(3):308–314
132. Hajduk PJ et al (1997) Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. J Am Chem Soc 119(25):5818–5827
133. Howard N et al (2006) Application of fragment screening and fragment linking to the discovery of novel thrombin inhibitors. J Med Chem 49(4):1346–1355
134. Nordstrom H et al (2008) Identification of MMP-12 inhibitors by using biosensor-based screening of a fragment library. J Med Chem 51(12):3449–3459
135. Hamalainen MD et al (2008) Label-free primary screening and affinity ranking of fragment libraries using parallel analysis of protein panels. J Biomol Screen 13(3):202–209
136. Mikuni J et al (2010) A fluorescence correlation spectroscopy-based assay for fragment screening of slowly inhibiting protein-peptide interaction inhibitors. Anal Biochem 402(1):26–31
137. Sullivan JE et al (2005) Prevention of MKK6-dependent activation by binding to p38alpha MAP kinase. Biochemistry 44(50):16475–16490
138. Haines DS et al (1994) Physical and functional interaction between wild-type p53 and mdm2 proteins. Mol Cell Biol 14(2):1171–1178
139. Sattler M et al (1997) Structure of Bcl-xL-Bak peptide complex: recognition between regulators of apoptosis. Science 275(5302):983–986
140. Voss SD et al (1993) Identification of a direct interaction between interleukin 2 and the p64 interleukin 2 receptor gamma chain. Proc Natl Acad Sci USA 90(6):2428–2432
141. Knight SM et al (2002) A fluorescence polarization assay for the identification of inhibitors of the p53-DM2 protein-protein interaction. Anal Biochem 300(2):230–236
142. Wan KF et al (2009) Differential scanning fluorimetry as secondary screening platform for small molecule inhibitors of Bcl-XL. Cell Cycle 8(23):3943–3952
143. Rothe A, Hosse RJ, Power BE (2006) In vitro display technologies reveal novel biopharmaceutics. FASEB J 20(10):1599–1610
144. Bottger A et al (1997) Design of a synthetic Mdm2-binding mini protein that activates the p53 response in vivo. Curr Biol 7(11):860–869
145. Ladbury JE et al (1995) Measurement of the binding of tyrosyl phosphopeptides to SH2 domains: a reappraisal. Proc Natl Acad Sci USA 92(8):3199–3203
146. Hajduk PJ, Meadows RP, Fesik SW (1999) NMR-based screening in drug discovery. Q Rev Biophys 32:211–240
147. Meyer B, Peters T (2003) NMR spectroscopy of proteins NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. Angew Chem Int Ed Engl 42:864–890
148. Shuker SB et al (1996) Discovering high-affinity ligands for proteins: SAR by NMR. Science 274(5292):1531–1534

149. D'Silva L et al (2005) Monitoring the effects of antagonists on protein-protein interactions with NMR spectroscopy. J Am Chem Soc 127(38):13220–13226

150. Emerson SD et al (2003) NMR characterization of interleukin-2 in complexes with the IL-2Ralpha receptor component, and with low molecular weight compounds that inhibit the IL-2/IL-Ralpha interaction. Protein Sci 12(4):811–822

151. Dalvit C et al (2001) WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability. J Biomol NMR 21(4):349–359

152. Major LL, Smith TK (2011) Screening the MayBridge Rule of 3 Fragment Library for Compounds That Interact with the Trypanosoma brucei myo-Inositol-3-Phosphate Synthase and/or Show Trypanocidal Activity. Molecular Biology International 2011:1–14

153. Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc 2(9):2212–2221

154. Koblish HK et al (2006) Benzodiazepinedione inhibitors of the Hdm2:p53 complex suppress human tumor cell proliferation in vitro and sensitize tumors to doxorubicin in vivo. Mol Cancer Ther 5(1):160–169

155. Karlsson R, Falt A (1997) Experimental design for kinetic analysis of protein-protein interactions with surface plasmon resonance biosensors. J Immunol Methods 200(1–2):121–133

156. de Kloe GE et al (2010) Surface plasmon resonance biosensor based fragment screening using acetylcholine binding protein identifies ligand efficiency hot spots (LE hot spots) by deconstruction of nicotinic acetylcholine receptor alpha7 ligands. J Med Chem 53(19):7192–7201

157. Perspicace S et al (2009) Fragment-based screening using surface plasmon resonance technology. J Biomol Screen 14(4):337–349

158. Arkin MR et al (2003) Binding of small molecules to an adaptive protein-protein interface. Proc Natl Acad Sci USA 100(4):1603–1608

159. Day YS et al (2002) Direct comparison of binding equilibrium, thermodynamic, and rate constants determined by surface- and solution-based biophysical methods. Protein Sci 11(5):1017–1025

160. Edink E et al (2011) Fragment growing induces conformational changes in acetylcholine-binding protein: a structural and thermodynamic analysis. J Am Chem Soc 133(14):5363–5371

161. Giannetti AM, Koch BD, Browner MF (2008) Surface plasmon resonance based assay for the detection and characterization of promiscuous inhibitors. J Med Chem 51(3):574–580

162. White PW et al (2003) Inhibition of human papillomavirus DNA replication by small molecule antagonists of the E1-E2 protein interaction. J Biol Chem 278(29):26765–26772

163. Torres FE et al (2010) Higher throughput calorimetry: opportunities, approaches and challenges. Curr Opin Struct Biol 20(5):598–605

164. Lakowicz JR (2006) Principles of fluorescence spectroscopy, 3rd edn. Springer, New York, xxvi, p 954

165. Degterev A et al (2001) Identification of small-molecule inhibitors of interaction between the BH3 domain and Bcl-xL. Nat Cell Biol 3(2):173–182

166. Ding K et al (2005) Structure-based design of potent non-peptide MDM2 inhibitors. J Am Chem Soc 127(29):10130–10131

167. Ding K et al (2006) Structure-based design of spiro-oxindoles as potent, specific small-molecule inhibitors of the MDM2-p53 interaction. J Med Chem 49(12):3432–3435

168. Petros AM et al (2006) Discovery of a potent inhibitor of the antiapoptotic protein Bcl-xL from NMR and parallel synthesis. J Med Chem 49(2):656–663

169. Titolo S et al (2003) Characterization of the minimal DNA binding domain of the human papillomavirus e1 helicase: fluorescence anisotropy studies and characterization of a dimerization-defective mutant protein. J Virol 77(9):5178–5191

170. Dams G et al (2007) A time-resolved fluorescence assay to identify small-molecule inhibitors of HIV-1 fusion. J Biomol Screen 12(6):865–874

171. Stallings-Mann M et al (2006) A novel small-molecule inhibitor of protein kinase Ciota blocks transformed growth of non-small-cell lung cancer cells. Cancer Res 66(3):1767–1774

172. Bergendahl V, Heyduk T, Burgess RR (2003) Luminescence resonance energy transfer-based high-throughput screening assay for inhibitors of essential protein-protein interactions in bacterial RNA polymerase. Appl Environ Microbiol 69(3):1492–1498

173. Erlanson DA, Wells JA, Braisted AC (2004) Tethering: fragment-based drug discovery. Annu Rev Biophys Biomol Struct 33:199–223

174. Choong IC et al (2002) Identification of potent and selective small-molecule inhibitors of caspase-3 through the use of extended tethering and structure-based drug design. J Med Chem 45(23):5005–5022

175. Blundell TL, Jhoti H, Abell C (2002) High-throughput crystallography for lead discovery in drug design. Nat Rev Drug Discov 1(1):45–54

176. Verlinde CLMJ et al (1977) Anti-trypanosomiasis drug development based on structures of glycolytic enzymes. In: Veerapandian P (ed) Structure-based drug design. Marcel Dekker, New York, pp 365–394

177. Hubbard RE et al (2007) The SeeDs approach: integrating fragments into drug discovery. Curr Top Med Chem 7(16):1568–1581

178. Hartshorn MJ et al (2005) Fragment-based lead discovery using X-ray crystallography. J Med Chem 48(2):403–413

179. Nienaber VL et al (2000) Discovering novel ligands for macromolecules using X-ray crystallographic screening. Nat Biotechnol 18(10):1105–1108

180. Davies TG et al (2006) Pyramid: an integrated platform for fragment-based drug discovery. In: Fragment-based approaches in drug discovery. Wiley-VCH Verlag GmbH & Co. KGaA. p 193–214

181. Spurlino JC (2011) Fragment screening purely with protein crystallography. Methods Enzymol 493:321–356

182. Carr R, Jhoti H (2002) Structure-based screening of low-affinity compounds. Drug Discov Today 7(9):522–527

183. Murray CW, Blundell TL (2010) Structural biology in fragment-based drug design. Curr Opin Struct Biol 20(4):497–507

184. Hassell AM et al (2007) Crystallization of protein-ligand complexes. Acta Crystallogr D Biol Crystallogr 63(Pt 1):72–79

185. Skarzynski T, Thorpe J (2006) Industrial perspective on X-ray data collection and analysis. Acta Crystallogr D Biol Crystallogr 62(Pt 1):102–107

186. D'Arcy A, Villard F, Marsh M (2007) An automated microseed matrix-screening method for protein crystallization. Acta Crystallogr D Biol Crystallogr 63(Pt 4):550–554

187. Carugo O, Argos P (1997) Protein-protein crystal-packing contacts. Protein Sci 6(10): 2261–2263

188. Janin J, Rodier F (1995) Protein-protein interaction at crystal contacts. Proteins 23(4):580–587

189. Sledz P et al (2011) From crystal packing to molecular recognition: prediction and discovery of a binding site on the surface of polo-like kinase 1. Angew Chem Int Ed Engl 50(17):4003–4006

190. Zhang X et al (2006) An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. Cell 125(6):1137–1149

191. Janin J (1997) Specific versus non-specific contacts in protein crystals. Nat Struct Biol 4(12):973–974

192. Janin J et al (2007) Macromolecular recognition in the Protein Data Bank. Acta Crystallogr D Biol Crystallogr 63(Pt 1):1–8

193. Gandhi L et al (2011) Phase I study of Navitoclax (ABT-263), a novel Bcl-2 family inhibitor, in patients with small-cell lung cancer and other solid tumors. J Clin Oncol 29(7):909–916

194. Erlanson DA (2011) Introduction to fragment-based drug discovery. Top Curr Chem DOI:10.1007/128_2011_180

195. 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_2011_183
196. Davies TG, Tickle IJ (2011) Fragment screening using X-ray crystallography. Top Curr Chem DOI:10.1007/128_2011_179
197. Hennig M, Ruf A, Huber W (2011) Combining biophysical screening and X-ray crystallography for fragment-based drug discovery. Top Curr Chem DOI:10.1007/128_2011_225

Fragment Screening and HIV Therapeutics

Joseph D. Bauman, Disha Patel, and Eddy Arnold

Abstract Fragment screening has proven to be a powerful alternative to traditional methods for drug discovery. Biophysical methods, such as X-ray crystallography, NMR spectroscopy, and surface plasmon resonance, are used to screen a diverse library of small molecule compounds. Although compounds identified via this approach have relatively weak affinity, they provide a good platform for lead development and are highly efficient binders with respect to their size. Fragment screening has been utilized for a wide range of targets, including HIV-1 proteins. Here, we review the fragment screening studies targeting HIV-1 proteins using X-ray crystallography or surface plasmon resonance. These studies have successfully detected binding of novel fragments to either previously established or new sites on HIV-1 protease and reverse transcriptase. In addition, fragment screening against HIV-1 reverse transcriptase has been used as a tool to better understand the complex nature of ligand binding to a flexible target.

Keywords Drug design · Fragment screening · HIV · Protease · Reverse transcriptase · Surface plasmon resonance · X-ray crystallography

Contents

J.D. Bauman, D. Patel, and E. Arnold (✉)
Center for Advanced Biotechnology and Medicine, Department of Chemistry and Chemical Biology, Department of Medicinal Chemistry, Rutgers University, Piscataway, NJ, USA
e-mail: arnold@cabm.rutgers.edu

Abbreviations

AIDS	Acquired immune deficiency syndrome
ClogP	Calculated logarithm of octanol–water partition coefficient
HAART	Highly active antiretroviral therapy
HIV	Human immunodeficiency virus
IC_{50}	Concentration of a compound leading to 50% enzyme inhibition
IN	Integrase
ITC	Isothermal calorimetry
K_D	Dissociation constant
LE	Ligand efficiency
MS	Mass spectroscopy
NMR	Nuclear magnetic resonance
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NRTI	Nucleoside/nucleotide reverse transcriptase inhibitor
PR	Protease
RT	Reverse transcriptase
SPR	Surface plasmon resonance

1   Introduction

Human immunodeficiency virus (HIV), the causative agent of acquired immune deficiency syndrome (AIDS), remains a medical challenge with more than 33 million people currently infected worldwide [1]. HIV, like other related retroviruses, relies on the replication of its RNA genome using the host's cellular machinery. Upon infection, reverse transcriptase (RT) copies the viral single-stranded (ss)RNA genome to double-stranded proviral DNA, which is then transported into the nucleus for integration into the host cell's genome. The provirus then exploits host cellular machinery to produce new infectious viral particles via normal cellular transcription and translation.

The elucidation of the viral replication cycle has identified key viral enzymatic targets – HIV-1 RT, integrase (IN), and protease (PR) – for anti-retroviral drug discovery and design. For the most part, HIV-1 RT and HIV-1 PR have been the focus of extensive drug therapy efforts. It is only recently that drugs targeting viral entry and HIV-1 IN have been approved. Highly active antiretroviral therapy (HAART), consisting of combination therapy usually with RT and PR inhibitors, has been found to cause a dramatic decrease in HIV-viral load within a few months. Despite such progress, treatment failure stemming from noncompliance, drug–drug

interactions, and long-term drug toxicity continues to be a persistent problem [2]. Emergence of multiple drug-resistant strains of the virus due to prolonged chemotherapy continues to press the need for novel, highly efficacious drugs.

2 Fragment-Based Drug Discovery

In recent years, the fragment-based approach has greatly facilitated the discovery and optimization of novel leads. Prior to its introduction in 1996, a commonly used drug discovery paradigm involved high-throughput screening of several hundred thousand drug-like compounds using in vivo assays for the detection of relatively strong inhibiting compounds. Although this approach has been successfully utilized in the development of numerous drugs, the drug design efforts were routinely plagued with challenges due to low hit rates, false positives, and substantial labor-intensive lead optimization. To circumvent these problems, the fragment-based approach was introduced as an alternative tool for drug discovery and design. Here, chemically diverse libraries of small molecule compounds or fragments are screened against a target protein to find relatively weak binding compounds. The promiscuous nature of fragments allows for higher hit rates while enabling efficient search of diverse chemical space [3–5]. Additionally, the small size of the fragments allows for higher ligand efficiency, which is a measure of the atomic contribution to the overall binding energy of a ligand. Ligand efficiency (LE) is typically defined as the free energy of dissociation (K_D) divided by the number of non-hydrogen atoms (n_{HA}) [3]:

$$LE = \frac{-RT \ln K_D}{n_{HA}}$$

where $-RT\ln K_D$ is the free energy of binding. Currently, there are three approaches that can guide lead optimization when utilizing fragment-based drug discovery: fragment evolution, fragment linking/merging, and fragment self-assembly. Fragment evolution involves the addition of functional groups to the original fragment hit to improve potency and binding. Here, the original hit acts as an "anchor" and often maintains its binding mode during the evolution process [6]. Typically, this process is guided by structural information provided by either X-ray crystallography or NMR spectroscopy.

If the target of interest has multiple fragment binding sites, a fragment linking approach can be utilized. Here, two fragment hits found binding to proximal sites within a target protein are joined using a linking group. This generally results in an improvement in the potency since the expected binding free energy of the linked molecule is greater than the sum of the binding energy of the individual fragments. Alternatively, if two or more fragments bind to overlapping sites, fragment merging can be used to join the fragments without the aid of a linker. It is important to note that stereochemical requirements for linking the two fragments can be restrictive because both of the fragments have to retain their original binding mode [6].

Fragment self-assembly also involves the fusion of two fragment hits into one larger molecule. In one popular approach, "click chemistry" is utilized. Chemically reactive fragments are screened such that upon binding to proximal sites within the target protein, the fragments react with each other to produce a larger inhibitor [6].

2.1 Fragment Library Design

One key to the fragment-based approach lies in the design of the fragment library. Typically, fragment libraries are relatively small in size, consisting of 500–1,000 commercially available molecules. These compounds are selected such that a high degree of chemical diversity and synthetic tractability is achieved [3–5]. In addition, guidelines, such as the Astex "Rule of Three," are used to ensure that compounds are indeed "fragment-like." The Rule of Three states that fragments should have mass ≤ 300 Da, ≤ 3 hydrogen bond acceptors, ≤ 3 hydrogen bond donors, a ClogP of ≤ 3, rotatable bonds of ≤ 3, and a polar surface area of ≤ 60 Å^2 [7, 8].

To facilitate screening efforts, fragment libraries are often grouped into cocktails of four to ten fragments. Compound solubility, toxicity, and potential chemical reactivity of the species within the cocktail itself are taken into consideration during the cocktail design process. Additionally, fragments are selected to minimize the chance of having more than one compound within a cocktail binding to the protein.

2.2 Fragment Screening Techniques

Due to the low binding affinities of fragment hits, the fragment-based approach heavily relies on sensitive biophysical methods such as nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography for efficient screening. NMR techniques are a favorite in the field because they allow for the reliable detection of very weak binders using a ligand-based as well as a target-based approach for screening of a wide range of targets. Protein-based NMR methods rely on changes in the protein resonances upon ligand binding. In addition to identifying high-affinity fragment hits, protein-based NMR methods can provide information regarding the actual binding pocket. However, these methods require long sample stability and large amounts of protein. Ligand-based NMR methods such as saturation transfer difference (STD) and water LOGSY take advantage of the differences in the ligand resonances between bound and unbound states. Ligand-based detection methods allow rapid hit identification using relatively little protein. However, tightly binding ligands can be false negatives since the rate of dissociation is not large enough to distinguish between the bound and unbound states [9–12].

X-ray crystallography provides a powerful method for fragment screening when the drug target can form suitable crystals. For crystals to be amenable to a fragment screening campaign they must meet the following criteria:

1. The crystals must be highly reproducible and diffract X-rays to high resolution (ideally better than 2.5 Å).
2. The protein must be in a biologically relevant conformation.
3. The druggable sites must not be occluded by protein–protein crystal contacts, by a natural ligand, or by a chemical used for crystallization or cryoprotection.
4. The crystals must be robust enough to survive the soaking of fragments.
5. The pH and ionic strength of the crystallization solution should optimally be near physiological.

Typically, crystals of the target protein are grown and then soaked in solutions of either individual fragments or cocktails. Soaking is conducted at relatively high fragment concentration, ranging from 10 to 100 mM, because the fragments are weak binders and the protein concentration within a crystal is relatively high. Cocktails are designed such that the fragments within a particular cocktail are diverse with respect to shape to allow for easy detection and deconvolution. At times, the crystal form may not be suitable for soaking, thus other crystal forms may need to be generated. An alternative to a soaking experiment is co-crystallization of either cocktails or individual fragments with the protein itself. However, this approach may require optimization of the crystallization conditions for each cocktail or fragment [13–16]. An X-ray crystallographic approach is considered advantageous because it allows for the visualization of multiple binding sites and, specifically, the binding mode to facilitate structure-based lead optimization. The availability of crystallization robotics and advancements in data collection make X-ray crystallography an attractive means for screening. However, it is still an extremely labor-intensive technique that is limited by the need for highly reproducible crystals that diffract X-rays to a reasonable resolution. The use of X-ray crystallography for fragment screening is also discussed in detail by Davies and Tickle [17].

Surface plasmon resonance (SPR) has recently become a common method of primary screening. Here, ligand binding is detected by changes in the refractive index of the solid support onto which the target protein is immobilized. The analytes (fragments) are injected in a continuous flow, and a real time sensogram is recorded. The availability of multiple biosensor channels allows for rapid, high-throughput screening of multiple proteins or protein complexes in parallel. SPR consumes as little as 25–50 μg of protein while retaining a high level of sensitivity to fragments with molecular weight as low as 100 Da [18–20]. In addition to primary hit detection, SPR also provides thermodynamic and kinetic information for ligand binding. Despite such advantages, it is important to note that careful assay design and data analysis are needed, and this is discussed in more detail by Hennig et al. [21].

Mass spectroscopy (MS) and isothermal calorimetry (ITC) also have been utilized as screening tools. MS techniques, such as non-denaturing electrospray ionization MS (ESI-MS), use mass identification as the means for the detection of reversible binding events. MS analysis allows for simultaneous binding of multiple fragments, and hence direct stoichiometric detection of the binding events. Despite such advantages, application is at times limited because the protein of interest may

not be stable in the presence of a volatile buffer necessary for analysis [22]. ITC has widely been used to determine the thermodynamics and stoichiometry of ligand binding in solution. It has not been routinely used in fragment-based drug discovery because it is a low-throughput technique that requires relatively large amounts of protein and time, and also lacks the sensitivity for the relatively weak binders needed for fragment screening [23, 24]. However, with recent improvements in instrumentation, ITC is slowly gaining ground as a screening tool.

Fragment-based drug discovery has been successfully applied to numerous targets. The success of this method depends heavily upon the design of the fragment library as well as on the wide range of techniques available for efficient screening. Thus, it has become an attractive approach for the design of novel HIV therapeutics to circumvent the drug resistance and adherence problems being faced today. Fragment screening against validated viral targets, specifically HIV-1 PR and HIV-1 RT, has been reported thus far.

3   HIV-1 Protease

HIV-1 PR plays a crucial role in the late stage of viral replication. It is responsible for the formation of viral proteins from the cleavage of the *gag-pol* polypeptide produced from the proviral transcription. Site-directed mutagenesis studies showed that a single point mutation can sufficiently inactivate the enzyme and stop viral infectivity, thus making HIV-1 PR an attractive target for antiretroviral therapy [25].

HIV-1 PR is a symmetrical homodimer consisting of two identical subunits of 99 amino acids. Its active site is formed at the dimer interface and contains two conserved, catalytic aspartic acid residues. A water molecule bound to the enzyme between the two aspartates acts as the nucleophile for catalysis. Each monomer contains a prominent β-hairpin loop, known as the "flap," that projects over the substrate-binding cleft. These flaps are highly flexible and can undergo large localized conformational changes upon substrate and inhibitor binding [25–27].

The first series of HIV-1 PR inhibitors, referred to as peptidomimetic inhibitors, are transition state mimics that resemble peptide substrates. They are relatively flexible, linear molecules with a well-defined backbone from which hydrophobic groups are projected into four or more of the subsites of the HIV-1 PR active site. The inhibitors function by creating a hydrogen bond network with a tetra-coordinated structural water molecule that is tightly bound between the inhibitor and the flaps. Inhibition is also depended upon critical interactions between the catalytic aspartates and the inhibitor [27].

Poor pharmacokinetic profiles and complex syntheses of peptidomimetic inhibitors led to a second class of HIV-1 PR inhibitors, loosely termed non-peptidic PR inhibitors. These inhibitors typically consist of a rigid, cyclic core with groups projected into the central subsites of the enzyme. Interestingly, the structural water was found to be absent in the crystal structures of non-peptidic inhibitors bound to

HIV-1 PR. Hydrogen bond accepting groups were found capable of retaining the hydrogen bonding interactions with the flap amide nitrogen directly, without the presence of the water [27, 28]. The ability of non-peptidic inhibitors to retain the necessary interactions for inhibition has made them attractive for drug design efforts. The availability of structural data has greatly facilitated the design of novel, highly efficacious peptidic and non-peptidic HIV-1 PR inhibitors and the search for possible allosteric HIV-1 PR inhibitors.

3.1 Fragment Screening

Examination of several different ligand-bound complexes of HIV-1 PR suggests that some of the drug-resistance mutations observed may alter the equilibrium between the closed and open states of the protein, thereby possibly decreasing drug binding affinity. Based on molecular dynamic simulations comparing the wild-type HIV-1 PR to the V82F/I84V drug-resistant mutant, the mutant was found capable of opening the flaps much farther with a greater degree of flexibility than the wild-type HIV-1 PR. In addition, the simulations revealed the presence of a solvent-exposed cleft, referred to as the "exo site," for both the wild-type and mutant HIV-1 PR in the closed conformation. This suggested a potential allosteric pocket that could inhibit HIV-1 PR by suppressing the motions of the flaps [29, 30].

An X-ray crystallography-based fragment screening was undertaken by Perryman et al., [30] to identify potential molecules targeting the newly discovered site. The Active Sight fragment library (San Diego, CA), consisting of 384 compounds dissolved in DMSO, was screened against HIV-1 PR with and without an active site inhibitor, TL-3. The library itself consisted predominantly of compounds with a single rigid core with one to three small substituents. It was subdivided into a total of 96 cocktails, with each cocktail consisting of four highly shape-diverse fragments. A combination of soaking and co-crystallization approach was utilized against five different crystal forms. Altogether, 808 crystals were screened and 378 datasets were collected and refined.

Individual fragment soaks using the apo $C222_1$ crystal form at a 10 mM fragment concentration yielded no hits. Similar results were also observed for soaking experiments for both the apo and TL-3 bound $P2_12_12$ crystal form. Co-crystallization of the $P6_122$ crystal form with an active site inhibitor, TL-3, and fragment cocktails or individual fragments at concentrations of 2.5 mM and 10 mM, respectively, was undertaken. For the large part, no fragment binding was observed. However, three cocktails (D9, F1, and F4) produced two new crystal forms.
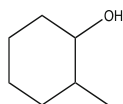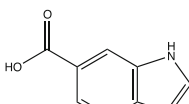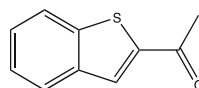
4D9   **1F1**   **2F4**

As with any X-ray crystallography-based experiment, the solvent was found to play a crucial role in the screening process. This was also true in the case of HIV-1 PR. Initial soaking experiments revealed a sensitivity of the $P4_1$ crystal form of the unliganded protein towards the DMSO concentration. To circumvent this problem, a similar concentration of DMSO was used for both drop preparation and soaking to retain diffraction quality. New crystallization conditions were sought to produce crystals that could handle the DMSO concentration and avoid the original precipitant, PEG 8000, which was found bound to HIV-1 PR.

Despite optimizing the crystal form, a large percentage of soaking experiments using the apo protease crystal form, $P2_12_12$, revealed that the exo site was occupied with acetate and water molecules from the buffer. Similarly, the co-crystallization of HIV-1 PR with TL-3 and fragments revealed that DMSO and water molecules occupied the exo site. This is a common occurrence with an X-ray crystallographic approach. Both the solvent and fragment compete for the same binding sites; however, a high solvent-to-fragment ratio in the soaking solution can permit solvent binding to occur despite the relatively weak interactions observed.

Co-crystallization experiments revealed interesting electron density at the exo site for the D9 cocktail, which changed the expected crystal form of the HIV-1–PR-TL-3 complex from $P6_122$ to $P2_12_12$. Subsequent cocktail deconvolution led to the identification of 4D9, 2-methylcyclohexanol, as the bound fragment (Fig. 1). This was further confirmed by soaking the $P2_12_12$ crystal form of the HIV-1 PR–TL-3 complex in a 20 mM solution of 4D9. Interestingly, the complementary experiment using the $P6_122$ form did not reveal fragment binding. This result was attributed to the packing interactions observed between the two different crystal forms.
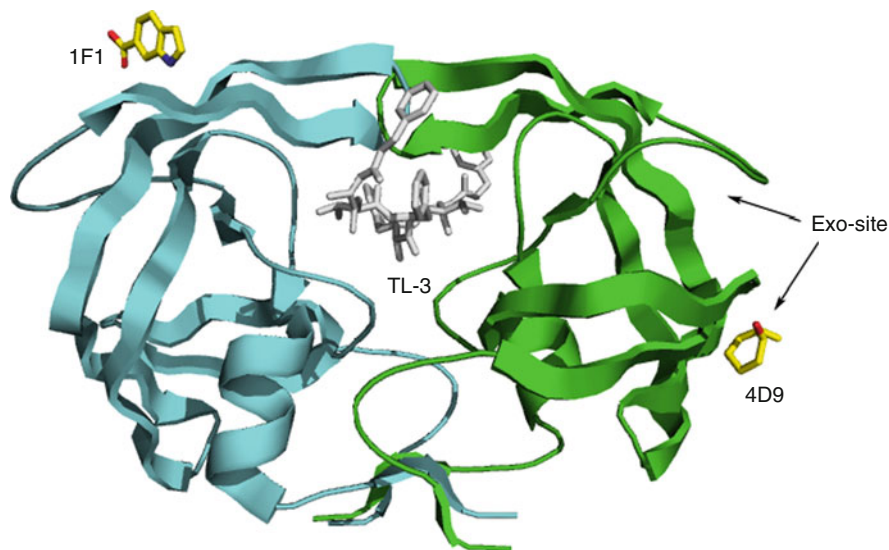
Fig. 1 HIV-1 PR with TL-3 bound in the active site. 4D9 and 1F1 represent fragment binding to the novel sites identified through screening efforts

Comparison of the structures revealed that in the $P2_12_12$ form, Ile72 from subunit A faces the solvent to allow the side chain of Leu63 to flip up to accommodate binding of 4D9. In the $P6_122$ form, Ile72 interacts with itself through a crystallographic twofold axis, thereby preventing the rearrangement of Leu63 needed for fragment binding.

This highlights one key disadvantage of using X-ray crystallography as a screening tool. Compared to solution-based screening tools, protein flexibility within a crystal is greatly limited. Additionally, the crystal packing can hinder fragment binding or create artificial fragment binding sites through nonphysiological crystal subunit interactions. Tight packing may require longer soaking times, whereas, in the case with HIV-1 PR, crystal packing can also affect the availability of pockets for binding. It is also important to note that fragment binding reported in this study was observed for only one monomer, despite the fact that HIV-1 PR is a symmetrical homodimer. This is due to the different crystallographic environment around each monomer. Thus, effective screening may require the use of multiple crystal forms of the protein to take into consideration protein flexibility and accessibility.

Co-crystallization of cocktails F1 and F4 with HIV-1 PR and TL-3, respectively, also revealed two additional fragments, 1F1 (indole-6-carboxylic acid) and 2F4 (2-acetylbenzothiophene), which bind in a similar manner to a pocket on the surface of the flap in monomer B (Fig. 1). Fragment binding induces significant conformational changes with a lateral shift in the anti-parallel β-strand segments at residues 45–47, 53–56, and 78–81 and a rearrangement at residues 35–41. Structural comparison of HIV-1 PR monomers with the flaps open, closed, and with and without 1F1, specifically at the base of the flap, revealed that 1F1 binds in a part open form despite having TL-3 occupying the active site and the flaps being closed. This suggests that the allosteric pocket might be functionally relevant for PR activity; however, further experiments need to be conducted for validation.

X-ray crystallography-based fragment screening was successfully utilized for the validation of the exo site suggested by molecular dynamics studies. Additionally, a novel binding pocket was discovered during the process. Although further studies are needed to demonstrate the inhibitory potential of these sites, the availability of structural data can facilitate further design. The screening process highlighted several key advantages and disadvantages to consider when utilizing an X-ray crystallography-based approach.

4 HIV-1 Reverse Transcriptase

As mentioned previously, RT is a crucial enzyme for viral replication. It is responsible for the conversion of the (+) ssRNA viral genome into double-stranded DNA. At the same time, it is also responsible for the degradation of the RNA genome after it is transcribed into DNA. X-ray crystal structures have revealed that

HIV-1 RT is a heterodimer consisting of p66 and p51 subunits. Both p66 and p51 have the same sequence, except that the p51 subunits lacks the C-terminal RNase H domain. Despite sequence commonality, both subunits vary greatly with respect to conformation. The p66 subunit is arranged such that its N-terminal region resembles an open right hand containing three subdomains, aptly referred to as fingers, palm, and thumb. Following the thumb subdomain, there is a connection domain, which leads to the C-terminal RNase H domain. In the p66 subunit, three catalytic residues are exposed in the nucleic acid binding cleft. However, these three residues are buried in the p51 subunit, which lacks this cleft [31, 32].

Currently, there are two broad classes of drugs targeting RT activity: nucleoside/nucleotide RT inhibitors (NRTIs) and non-nucleoside RT inhibitors (NNRTIs). NRTIs are analogs of endogenous 2′-deoxy-nucleosides that lack the 3′-hydroxyl needed for polymerization. Upon incorporation by RT, they acts as chain-terminators of the viral reverse transcripts. NNRTIs bind to a normally closed allosteric binding pocket in the palm subdomain. This stabilizes a single conformation of the palm/thumb subdomains, which is not sufficient for polymerization to occur. Additional classes of inhibitors in active development are nucleotide-competing RT inhibitors (NcRTIs) [33], p66/p51 dimerization inhibitors [34], and RNase H inhibitors [35].

4.1 Fragment Screening Against HIV-1 Reverse Transcriptase by X-Ray Crystallography

The flexible nature of HIV-1 RT is a crucial feature both from a biological and a drug design perspective [36, 37]. Biologically, the flexibility of the enzyme plays a crucial role in the catalytic activity of the protein by allowing the exact coordination of several domains in the protein to occur. From a drug design perspective, protein flexibility combined with various interdomain hinges present throughout the protein suggests the possible existence of novel druggable sites.

We utilized fragment screening by X-ray crystallography to investigate potential new inhibitory sites in HIV-1 RT. Initially, the apo crystal form of HIV-1 RT was utilized for screening. However, it was found to be unsuitable for soaking experiments due to sensitivity towards the soaking condition and methodology despite extensive optimization. As a result, an alternate crystal form, HIV-1 RT in complex with TMC278, a potent NNRTI, was used for screening. An engineered form of HIV-1 RT was used; it had been developed because crystallization trials of wild-type HIV-1 RT complexed with TMC278 had failed to give crystals of X-ray diffraction quality.

Taking advantage of the tremendous amount of available structural data, crystal engineering was used to improve the crystal quality and subsequent X-ray diffraction. The termini of p66 and p51 were truncated to remove residues found to be disordered in previous crystal structures. In addition, common crystal contacts were

mutated to increase the likelihood of obtaining new crystal forms, and high-entropy residues (lysines and glutamic acids) were mutated to alanine to lessen the entropic penalty of forming crystal contacts. After several rounds of mutagenesis and testing, an HIV-1 RT variant (RT52A) was made to diffract X-rays to 1.8 Å resolution when crystallized in complex with TMC278 and other NNRTIs [38]. These crystals were found to be highly reproducible and robust for fragment screening but could not be used to find new NNRTIs because the NNRTI binding pocket is occupied.

A fragment library consisting of 775 commercially available compounds was assembled in-house. This library consisted of 500 compounds purchased from Maybridge (Cornwall, UK), 175 individual compounds purchased on the basis of recommendations by Christophe Verlinde and Wim Hol [39], and an additional 100 compounds that were generously gifted by James Williamson (The Scripps Research Institute, La Jolla, CA). The fragments were divided into cocktails containing an average of five compounds. The cocktails were designed such that structural diversity and cocktail solubility was maximized and the chemical reactivity between fragments of the same cocktail was not an issue. To ensure maximum structural diversity within the cocktail, a program, FROCIVANTO, was developed (Eck and Arnold, unpublished results from this laboratory). FROCIVANTO produces shape fingerprints for all fragments in the library and then generates the distance matrix for the fingerprints, utilizing the Euclidean-style Ultrafast Shape Recognition distance metric [40]. The fragments are then ordered on the basis of shape similarity and a simple partitioning technique is used whereby cocktails are generated sequentially, with every Nth fragment being selected from the shape-based ordering. Typically, cocktails with an average of five compounds were designed and prepared such that the final concentration of each fragment was 100 mM in d_6-DMSO.

Initial screening experiments pressed the need for optimization of the soaking and freezing conditions with respect to fragment concentration and solubility, soaking time, and crystal stability. The RT52A–TMC278 crystals were found to survive soaking conditions with a final fragment concentration of 20 mM. Importantly, 20% (v/v) d_6-DMSO was also found to serve a dual purpose, both as a solvent for the fragment as well as a cryo-protectant during crystal freezing. This facilitated the soaking process since only one solution was needed for both soaking and freezing. Fragment solubility in the soaking solution proved to be a major obstacle to overcome. The crystal soaking time was increased from several seconds to approximately 1–2 min to counter the lower fragment concentration in solution and allow for the crystal to equilibrate in the soaking solution. Also, the addition of 80 mM L-arginine was found to improve fragment solubility in the soaking solution. This was found to have a tremendous impact on the screening results because many of the fragment hits were found to bind the RT52A–TMC278 complex only in the presence of L-arginine. It is important to note that the L-arginine served only as an additive to improve solubility and was not found present in the electron density.
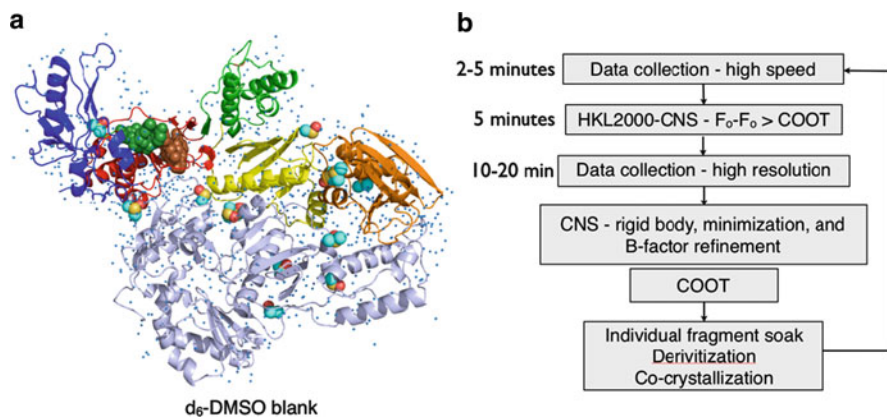
Fig. 2 (**a**) X-ray crystal structure of HIV-1 RT without any fragments soaked, but frozen with all other solution components present including 20% d_6-DMSO. The fingers, palm, thumb, connection, and RNase H subdomains are color coded *blue*, *red*, *green*, *yellow*, and *orange*, respectively. The NNRTI-binding pocket is shown in *brown* and the polymerase active site is colored *dark green*. The positions of the bound d_6-DMSO molecules are shown as *cyan spheres*. (**b**) Scheme of data collection and initial processing

In addition to optimizing experimental conditions, a protocol was designed to efficiently screen and process the wealth of data collected during the course of the screening. Due to the high concentration of d_6-DMSO present in the soaking solution, a reference structure for a crystal soaked in 20% (v/v) d_6-DMSO soaking condition without any fragment present was determined in order to identify the d_6-DMSO binding sites. Multiple binding sites for d_6-DMSO were observed (Fig. 2a). To avoid misinterpretation of the solvent molecules as fragment binding, this structure served as a "blank" for subsequent analyses. As shown in Fig. 2b, a high-speed pass was initially performed for a crystal to collect X-ray diffraction to no better than 2.1 Å. The high-speed pass maximizes the quantity of fragments screened during a time-limited X-ray data collection trip (often five to ten datasets per hour are possible).

The diffraction calculated in the CNS program system was then immediately processed at the beamline using HKL2000 and a map using F_o–F_o coefficients comparing the fragment dataset to the "blank" reference dataset. The maps were then evaluated for changes in electron density for the presence of any strong positive density that could be indicative of fragment binding or the movement of the residue side chains on the protein backbone. A high-resolution pass was conducted when the F_o–F_o maps suggested fragment binding. These datasets were further refined using CNS to clearly delineate fragment binding. Binding detected from cocktail soaks was subsequently verified by soaking of individual fragments. These compounds were then tested for polymerization and RNase H inhibition using activity assays. In addition, fragments were co-crystallized with the apo protein to see whether or not binding was maintained without the NNRTI-pocket being occupied and to verify that inhibition was not due to fragment binding to the
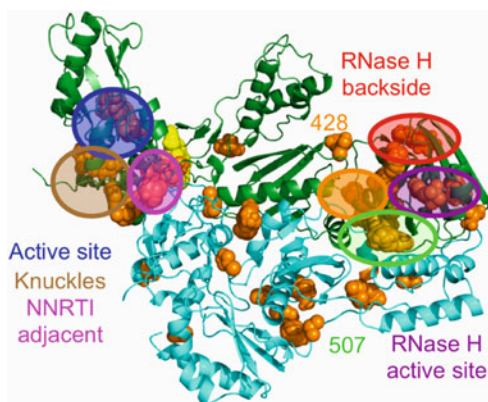
NNRTI pocket. Chemical derivatives of the most promising fragments were purchased and tested for binding and inhibition using X-ray crystallography and activity assays, respectively.

A total of 705 datasets with an average resolution of 2.1 Å were collected for 742 compounds out of the 775 screened. From these datasets, 34 compounds were found to bind to the protein complex, giving a hit rate of 4.4%. Fifty binding events were observed, with many of these compounds binding to multiple locations throughout the protein complex, including the hinge regions as well as the polymerase and RNase H active sites (shown in Fig. 3). In addition, two new allosteric pockets – the knuckles and the NNRTI adjacent – were discovered from the screening efforts (Fig. 4). The knuckles pocket had not been observed in any of the HIV-1 RT structures thus far and was found to be present only in conjunction with fragment binding.

Prior to fragment binding, the knuckles pocket is a non-solvent-exposed cavity near the incoming dNTP substrate-binding site (Fig. 4a). Upon fragment binding to this cavity, the polypeptide backbone rearranges (Ser117 Cα is displaced 2.8 Å) to create a solvent-exposed pocket. The formation of the pocket results in a 3.2 Å movement in the backbone for active site residues Tyr115 and Phe116, which are involved in incoming nucleotide binding during polymerization. The pocket opens also at residue Ser163 creating a potential direction for fragment growth. The original fragment hits and derivatives found binding to this site were found to have inhibitory activity with the top derivative having an IC_{50} of 600 μM, with a ligand efficiency of 0.37 kcal/mol NHA.

In a typical NNRTI-bound structure without fragment binding, the NNRTI adjacent binding site is a solvent-accessible pocket at the p66–p51 interface that is separated from the NNRTI-binding pocket by Val179 and Ile180. Bordered by Thr139, Pro140, Thr165, Leu168, Lys172, and Ile180, this pocket expands upon fragment binding (Fig. 4b). The fragments binding to this pocket were found to be inhibitory in the absence of an NNRTI. The best primary hit was found to have an IC_{50} value of 350 μM and a ligand efficiency of 0.34 kcal/mol NHA. The close
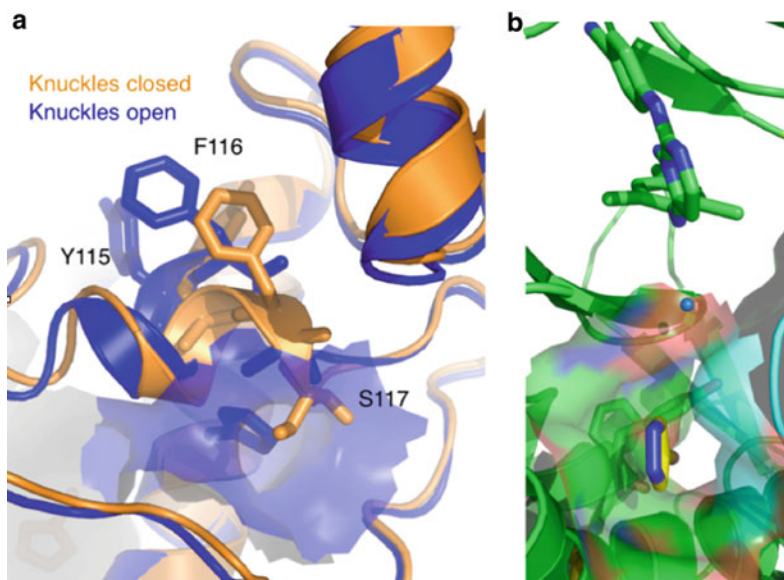
Fig. 4 Novel inhibitory binding sites. (**a**) The knuckles binding site at the fingers/palm of p66 junction. Unbound is colored *orange* and bound is colored *blue*. The *transparent* representation shows the molecular surface of the open pocket. The 4-bromopyrazole fragment is shown in the pocket. (**b**) The NNRTI adjacent binding pocket is shown with a *transparent* surface representation. At the *top* is the NNRTI-binding pocket with TMC278 shown in *green* and *blue*. A conserved water molecule is shown between the two pockets as a *blue sphere*. 4-Bromopyrazole is shown bound in the lower NNRTI adjacent pocket

proximity to the NNRTI-binding pocket may allow for expansion of known NNRTIs by fragment linking. Alternatively, compounds targeting solely the NNRTI adjacent pocket can be potentially promising inhibitors against both wild-type and NNRTI drug-resistant forms of the protein.

Interestingly, during the course of the screening it became apparent that halogenated compounds were frequently found to be hits. In fact, four out of the 17 brominated compounds screened and seven out of the 29 fluorinated compounds screened were found to bind throughout the protein giving a hit rate of 23.5% and 24.1%, respectively. Although this preference for halogen-containing compounds is not fully understood, the higher hit rates suggest that use of halogenated compounds may be advantageous. One of the compounds, 4-bromopyrazole, was found bound to 11 sites through out the protein, including the two new sites described (Fig. 5). Burley and coworkers at SGX Pharmaceuticals (San Diego, CA) described using a fragment library containing a large fraction of brominated compounds, in part to permit Br anomalous scattering measurements for rapid assessment of fragment binding [41].

X-ray crystallography was successfully used for primary fragment screening despite the challenges faced. Fragment screening was surprisingly effective in

Fig. 5 HIV-1 RT with p66
colored *green* and
p51 colored *blue*.
4-Bromopyrazole molecules
found are shown as *colored
spheres*
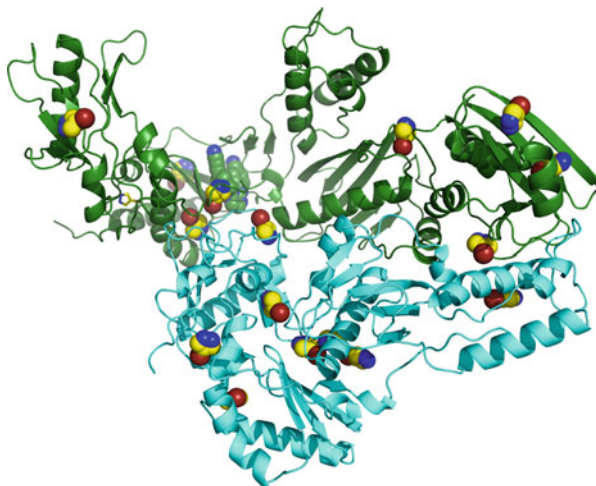
discovering new inhibitory sites on HIV-1 RT when hundreds of drug discovery projects using other methods had failed to detect them. This demonstrates the utility of fragment screening by X-ray crystallography to probe a protein for new allosteric sites.

4.1.1 Fragment Screening and Surface Plasmon Resonance

Recently, Geitmann et al. [42], has reported the use of SPR as a screening tool for identifying novel NNRTI scaffolds from a fragment library consisting of 1,040 compounds screened against HIV-1 RT. Prior to screening, the fragment library was first evaluated for chemical diversity using a Tversky similarity analysis to compare the fragments to each of 826 published NNRTIs extracted from the BindingDB. The analysis revealed that the majority of the compounds in the library were significantly different from the published NNRTIs. The results also indicated that 28 out of 1,040 fragments were substructures of NNRTIs with respect to simple atom connectivity, with many of them having different functionality and polarity compared to the established NNRTI. Overall, the use of the Tversky similarity analysis provided a means for ensuring the exploration of novel chemical scaffolds.

A primary screen was conducted against the wild-type HIV-1 RT with the fragment concentration ranging from 50 to 400 μM and nevirapine, a potent NNRTI, as the positive control. The use of multiple concentrations enabled the identification of false positives, a common problem with SPR-based assays. A total of 165 compounds were selected with apparent K_D values of less than 1 mM and a stoichiometry of 0.75–5 times the value obtained with nevirapine. The sensograms

of the 165 compounds were then subjectively evaluated for basic interaction characteristics, such as rate of dissociation, to eliminate any false positives. In addition, sensograms were evaluated for secondary effects stemming from poor compound solubility, conformational changes within the bound protein, and clearance of the biosensor surface for false positives. A total of 69 compounds were eliminated due to strong secondary effects or slow dissociation, thereby leaving only 96 hits from the primary screen.

The 96 compounds identified as primary hits were then subjected to two independent experiments in parallel. First, an-SPR based competition assay was used to screen each of the 96 fragments and nevirapine at concentrations of 200 and 20 μM, respectively. The sensograms were also evaluated to remove any false positives based on the previously mentioned criteria. Out of the 96 compounds originally identified from the primary screen, only 20 were found to compete with nevirapine. At the same time, the hits from the primary screen were also evaluated in an enzyme inhibition assay. Of the 96 compounds assayed, 27 compounds were considered hits with IC_{50} values lower than 1 mM.
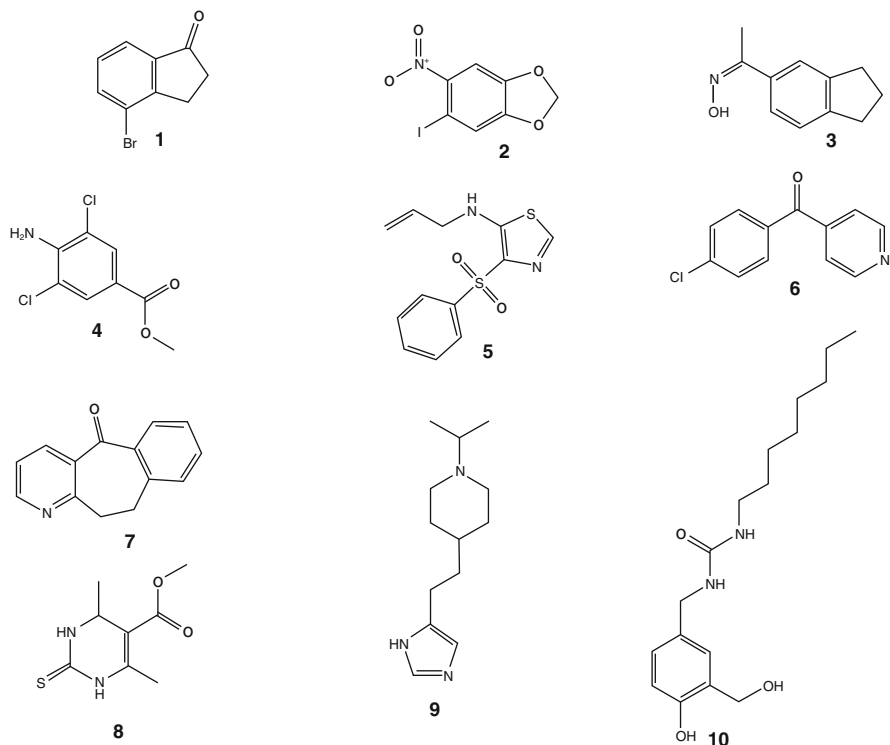
Ten compounds from the set of 27 were found to compete with nevirapine as well as inhibit wild-type HIV-1 RT activity with submillimolar IC_{50} values. The compounds were then re-evaluated under more fully optimized conditions against the wild-type HIV-1 RT for further validation. Fragments **9** and **10** failed to reproducibly show inhibition and were dropped from further analysis. The remaining eight fragments were then screened against drug-resistant mutants of HIV-1 RT: K103N, L100I, and Y181C. Only fragment **1**, bromoindanone, was found to have an IC_{50} value lower than 25 μM against all four enzyme variants. This gives bromoindanone a very high ligand efficiency of greater than 0.57 kcal/mol NHA, making it a potentially promising starting point for lead development for a novel NNRTI.

4.1.2 Fragment Screening and Pocket Flexibility

Interestingly, only two of the 28 fragments initially identified as substructures of published NNRTIs from the similarity analysis were found to be hits among the 20 compounds found through the competition screen, but failed to show enzymatic inhibition. Although the lack of inhibition of the two compounds could be explained by the change in polarity and conformation preference compared to the parent NNRTI, the poor hit rate from the pool of NNRTI-based substructures suggests a lack of an efficient binding hot spot. Brandt et al. [43] utilized fragment screening and SPR as a tool to understand the nature of fragment binding to flexible pockets, specifically the NNRTI pocket in HIV-1 RT.

Twenty-one fragments stemming from three NNRTIs were purchased and assayed in an SPR-based biosensor assay. Based on the ligand efficiency of the parent compound, the deconstructed fragments were expected to have high ligand efficiency and IC_{50} values below 3 mM. However, only nine of the 21 compounds tested were found to be binders. Three of the 21 fragments were found to have reproducible K_D values in the micromolar range. The six remaining compounds were also found to bind, but K_D values were above the highest concentration tested and were estimated despite a large linear component in the equation or by a less good fit to data. A closer look at the binders and the non-binders from the deconstruction experiment revealed that the substructures found to bind were relatively large compounds. In fact, only three (**11**, **14**, **15**) out of the nine that bound could be considered as fragments (i.e., with a molecular weight of less than 300 Da) and the average number of heavy atoms for the binders and non-binders was 21 and 14, respectively.
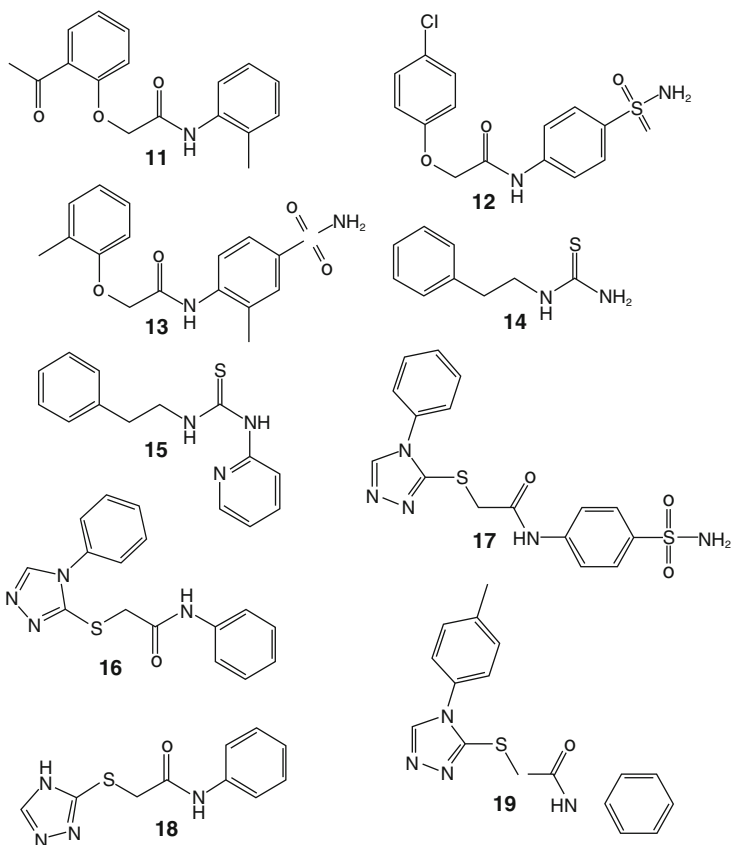
To account for these observations, a corrected ligand efficiency (LE*), which incorporates a ligand-independent free energy fee, ΔG_{ind}, was proposed. Here, ΔG_{ind} is defined as the energy required to create the NNRTI-binding site ($\Delta G_{opening}$) as well as the change in free energy caused by the loss in translational and rotational entropy ($T\Delta S_{tr}$):

$$LE^* = \frac{\Delta G_{opening} - T\Delta S_{tr} - RT \ln K_D}{n_{HA}}$$

A strong correlation between experimental and predicted results was observed when using a value of 7.0 kcal/mol for ΔG_{ind}. However, this model assumes that the protein–ligand interaction energy is uniformly distributed over the ligand–protein interface, thereby ignoring the existence of hot spots, which in turn underestimates the potential maximum affinity of fragments.

The modified ligand efficiency will have a greater preference for smaller fragments, thereby rendering it inefficient as a measure for prioritizing compounds for lead development. To circumvent the observed size dependency on LE*, Brandt

and coworkers proposed the use of a modified fit quality (FQ*), which is a function of the number of heavy atoms:

$$FQ_* = \frac{LE^*}{LE\_Scale^*}$$

$$\text{where } LE\_Scale^* = 0.0975 + \frac{17.3}{n_{HA}} + \frac{35.1}{n_{HA}^2} - \frac{493}{n_{HA}^3}$$

FQ* analyses of the experimental library as well as the 946 NNRTIs published in the database were found to be suboptimal. Additionally, a dramatic drop in the FQ* was observed for compounds with less than 20 heavy atoms, thereby suggesting either that larger concentrations are needed for fragment screening or that small fragments are not ideal for screening. Although FQ*, like LE*, cannot help prioritize hits for lead development, it can provide information about the binding site landscape, which can in turn facilitate the course of lead design.

Fragment screening has been used as a tool to identify novel scaffolds targeting the NNRTI-binding pocket of HIV-1 RT. However, only one compound out of the 1,040 screened was found to be active against both the wild-type and mutant variants of HIV-1 RT. Additionally, a poor hit rate for NNRTI-like fragments in primary screens was observed. A deconstruction analysis, which involves screening of fragments based on known NNRTIs, was used to assess the amenability of the NNRTI pocket towards fragment screening.

5 Closing Remarks

Fragment screening is breathing new life into drug discovery against HIV-1 targets. The relatively low cost of purchasing and curating a collection of 500–1,000 compounds makes it an attractive method of drug discovery for academic as well as industrial groups. Fragment screening has been used to discover new druggable sites as well as novel scaffolds against well-established binding pockets. The novel nature of the compounds discovered by fragment screening will feed drug design efforts against HIV-1 for years to come.

References

1. World Health Organization (2008) World Health Statistics http://www.who.int/whosis/whostat/2008/en/index.html
2. Mehellou Y, De Clercq E (2010) J Med Chem 53:521
3. Hadjuk PJ, Greer J (2007) Nat Rev Drug Discov 6:211–219
4. Hesterkamp T, Whittaker M (2008) Curr Opin Chem Biol 12:260–268
5. Erlanson DA, Wells JA, Braisted AC (2004) Annu Rev Biophys Biomol Struct 33:199–223
6. Rees DC, Congreve M, Murray CW, Carr R (2004) Nat Rev Drug Discov 3:660–672

7. Congreve M, Chessari G, Tisi D, Woodhead AJ (2008) J Med Chem 51:3661–3680
8. Congreve M, Carr R, Murray C, Jhoti H (2003) Drug Discov Today 8:876–877
9. Zartler ER, Shapiro MJ (2005) Curr Opin Chem Biol 9:366–370
10. Siegal G, Ab E, Schultz J (2007) Drug Discov Today 12:1032–1039
11. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Science 274:1531–1534
12. Hadjuk PJ, Meadows RP, Fesik SW (1999) Quart Rev Biophys 32:211–240
13. Jhoti H, Cleasby A, Vedonk M, Williams G (2007) Curr Opin Chem Biol 11:485–493
14. Carr R, Jhoti H (2002) Drug Discov Today 7:522–527
15. Hartshorn MJ, Murray CW, Cleabsy A, Frederickson M, Tickle IJ, Jhoti H (2005) J Med Chem 48:403–413
16. Murray CW, Blundell TL (2010) Curr Opin Struct Biol 20:497–507
17. Davies TG, Tickle IJ (2011) Fragment screening using X-ray crystallography. Top Curr Chem. doi:10.1007/128_179
18. Navratilova I, Hopkins A (2010) ACS Med Chem Lett 1:44–48
19. Neumann T, Junker HD, Schmidt K, Sekul R (2007) Curr Top Med Chem 7:1630–42
20. Huber W, Mueller F (2006) Curr Pharm Des 12:3999–4021
21. Hennig M, Ruf A, Huber W (2011) Combining biophysical screening and X-ray crystallography for fragment-based drug discovery. Top Curr Chem. doi:10.1007/128_225
22. Zartler ER, Shapiro MJ (2008) In: Zartler ER, Shapiro MJ (eds) Fragment-based drug discovery: a practical approach. Wiley, United Kingdom
23. Torres FE, Recht MI, Coyle JE, Bruce RH, Williams G (2010) Curr Opin Struct Biol 20:598–605
24. Ladbury JE, Klebe G, Freire E (2010) Nat Rev Drug Discov 9:23–27
25. Robins T, Plattner J (1993) J Acq Immun Def Syn 6:162–170
26. Kempf DJ, Sham HL (1996) Curr Pharm Des 2:225–246
27. Abdel-Rahman HM, Al-karamany GS, El-Koussi NA, Youssef AF, Kiso Y (2002) Curr Med Chem 9:1905–1922
28. Huff JR (1991) J Med Chem 34:2305–2314
29. Perryman AL, Lin JH, McCammon JA (2004) Prot Sci 13:1108–1123
30. Perryman A, Zhang Q, Soutter HH, Rosenfeld R, McRee DE, Olson AJ, Elder JE, Stout CD (2010) Chem Biol Drug Des 75:257–268
31. Kohlstaedt LA, Wang J, Friedman JM, Rice PA, Steitz TA (1992) Science 256:1783–1790
32. Jacobo-Molina A, Ding J, Nanni RG, Clark AD, Lu X, Tantillo C, Williams RL, Kamer G, Ferris AL, Clark P (1993) Proc Natl Acad Sci USA 90:6320–6324
33. Jochmans D, Deval J, Kesteleyn B, Van Marck H, Bettens E, De Baere I, Dehertogh P, Ivens T, Van Ginderen M, Van Schoubroeck B, Ehteshami M, Wigerinck P, Götte M, Hertogs K (2006) J Virol 80:12283–12292
34. Camarasa M-J, Velázquez S, San-Félix A, Pérez-Pérez MJ (2005) Antivir Chem Chemother 16:147–153
35. Tramontano E, Di Santo R (2010) Curr Med Chem 17:2837–2853
36. Götte M, Rausch JW, Marchand B, Sarafianos S, Le Grice SFJ (2010) Biochim Biophys Acta 1804:1202–1212
37. Sarafianos SG, Marchand B, Das K, Himmel DM, Parniak MA, Hughes SH, Arnold E (2009) J Mol Biol 385:693–713
38. Bauman JD, Das K, Ho WC, Baweja M, Himmel DM, Clark AD, Oren DA, Boyer PL, Hughes SH, Shatkin AJ, Arnold E (2008) Nucleic Acids Res 36:5083–5092
39. Verlinde CLMJ, Fan E, Shibata S, Zhang Z, Sun Z, Deng W, Ross J, Kim J, Xiao L, Arakaki T, Bosch J, Caruthers JM, Larson ET, LeTrong I, Napuli A, Kelley A, Mueller N, Zucker F, Van Voorhis WC, Buckner FS, Merritt EA, Hol WGJ (2009) Curr Top Med Chem 9:1678–1687
40. Ballester PJ, Richards WG (2007) J Comput Chem 28:1711–1723
41. Blaney J, Nienaber V, Burley SK (2006) In: Jahnke W, Erlanson DA (eds) Fragment-based Approaches in Drug Discovery. Wiley-VCH Verlag GmbH & Co. KGaA
42. Gietmann M, Elinder M, Seeger C, Brandt P, de Esch IJP, Danielson UH (2011) J Med Chem 54:699–708
43. Brandt P, Geitmann M, Danielson UH (2011) J Med Chem 54:709–718

Fragment-Based Approaches and Computer-Aided Drug Discovery

Didier Rognan

Abstract Fragment-based design has significantly modified drug discovery strategies and paradigms in the last decade. Besides technological advances and novel therapeutic avenues, one of the most significant changes brought by this new discipline has occurred in the minds of drug designers. Fragment-based approaches have markedly impacted rational computer-aided design both in method development and in applications. The present review illustrates the importance of molecular fragments in many aspects of rational ligand design, and discusses how thinking in "fragment space" has boosted computational biology and chemistry.

Keywords Docking · Drug design · Fragment · Library

Contents

D. Rognan
Structural Chemogenomics, UMR 7200 CNRS-University of Strasbourg, 67400 Illkirch, France
e-mail: rognan@unistra.fr

1 Introduction

Fragment-based drug discovery (FBDD) was introduced almost 15 years ago with the "SAR by NMR method" developed by Fesik and coworkers at Abbott Laboratories [1]. This remarkable technical achievement happened at a time when the pharmaceutical industry was seriously concerned with the problem of high attrition rates in clinical development, notably arising from compounds with unfavorable pharmacokinetic properties [2]. Contrary to initial expectations, neither the sequencing of the human genome [3] nor the advent of miniaturized high-throughput technologies (ranging from synthetic organic chemistry [4] to biochemical screening [5]) increased the productivity of the pharmaceutical industry. The seminal concept of drug-likeness [6] progressively shifted to lead-likeness [7] and even fragment-likeness [8], illustrating the necessity to change paradigms in discriminating good from bad physicochemical properties of a drug precursor. Various biophysical screening methods (X-ray diffraction, NMR, surface plasmon resonance, and mass spectrometry) have evolved to detect small molecular weight fragments, and medicinal chemists have been progressively convinced of the necessity to optimize fragments with high efficiency in binding [9].

From a method of last resort, the fragment approach has evolved into an amazingly successful screening method [10], notably for targets previously considered intractable (e.g., protein–protein interfaces). There are currently 15 compounds in phase I or II clinical trials [11] and 24 biotechnological companies [12] originating from FBDD. As such, it has profoundly impacted rational approaches to drug design at many steps (library design; hit identification, triage, and optimization). Many excellent reviews have partly addressed these issues recently [9, 10, 13–15]. We will focus here on how FBDD has changed rational drug design methods, either in promoting novel areas of research or in boosting previously known approaches.

2 Dissecting the Energetics of Fragment Binding and Fragment Optimization

2.1 Basic Concepts

Two basic concepts arising from computational chemistry have played an important role in analyzing fragment-based screening (FBS) data. The first one, proposed by Jencks almost 30 years ago [16], argues than the Gibbs free energy changes upon

ligand (L) binding to a protein (P) may be decomposed into intrinsic binding energies of both partners and a connection Gibbs energy ΔG_S^0) related to the loss of rotational/translational entropy upon binding:

$$\Delta G_{PL}^0 = \Delta G_P^0 + \Delta G_L^0 + \Delta G_S^0. \tag{1}$$

The additivity principle of binding free energies may be applied to the fragment-to-lead optimization, in which the lead is the sum of two fragments tethered by a linker, with an affinity greater than the sum of affinities of the respective fragments.

The second important concept, developed by Kuntz et al. [17], proposes that the maximal binding free energy (affinity) of a ligand to a protein linearly increases with the number of heavy atoms and levels off once the ligand has reached ca. 15 heavy atoms. It is therefore easier to optimize the affinity of a low molecular-weight fragment than that of bigger drug-like compounds. Along the same line, Hann et al. demonstrated that the probability of a binding event was inversely related to the molecular complexity of the ligand [7].

2.2 Ligand Efficiency Indices

Acknowledging these two paradigms, computational chemists have significantly helped structural biologists and medicinal chemists to analyze FBS data. Following Kuntz's principle, the ligand efficiency (LE) index was formulated to discriminate promising from non-promising fragments [18]. The LE index is a simple normalization of the binding free energy ΔG by the number of non-hydrogen heavy atoms (NHA):

$$LE = -\frac{\Delta G}{NHA}. \tag{2}$$

Although this measure of ligand efficiency is the most widely used, it has been proposed that the LE index could be replaced by size-independent indices like the fit quality score (FQ) [19] or the percentage LE [20], which allow the comparison of fragments irrespective of their molecular weight. Many other related indices (listed in Table 1) have also been proposed recently to take into account properties other than molecular weight in the fragment-to-lead optimization, such as polar surface area or lipophilicity [20].

The increasing availability of FBDD screening data suggests potential relationships between certain properties of fragment hits and their corresponding optimized leads [20]. Decomposing a lead into a scaffold (conserved molecular skeleton) and an "evolution part" (i.e., added substituents) revealed the following relationships for a set of 30 follow-up FBDD programs:

Table 1 Ligand efficiency indices [21]

Index	Abbreviation	Definition
Ligand efficiency	LE	pIC_{50}, pK_i, or $-\Delta G$/NHA
Binding efficiency index	BEI	pIC_{50}, pK_i, or $-\Delta G$/molecular weight
Percent efficiency index	PEI	% Inhibition/molecular weight
Surface efficiency index	SEI	pIC_{50}, pK_i, or ΔG/(PSA/100 Å^2)
Ligand lipophilicity efficiency	LLE	pK_i – clogP
Group efficiency	GE	$-\Delta\Delta G$/NHA
Fit quality	FQ	LE/LE_scale[a]
Percentage ligand efficiency	%LE	(LE/LE$_{max}$[b]) × 100

[a]LE_scale = $0.0715 + (7.5328/\text{NHA}) + (25.7079/\text{NHA}^2) - (361.4722/\text{NHA}^3)$

[b]LE$_{max}$ = $1.614^{\log_2(10/\text{NHA})}$

clogP computed partition coefficient, *NHA* number of non-hydrogen heavy atoms, *PSA* polar surface area

$$\frac{\text{NHA}^{\text{lead}}}{\text{NHA}^{\text{sca}}} = \frac{\text{NHA}^{\text{sca}}}{\text{NHA}^{\text{evo}}} = \frac{(pK_d)_{\text{lead}}}{(pK_d)_{\text{fragment}}} = 1.6. \tag{3}$$

This relationship could potentially be used either to design fragment libraries according to properties of known leads (for example by fixing boundaries in heavy atom counts), or in the fragment-to-lead process to anticipate the maximal affinity of the optimized lead from that of the fragment hit. Reporting such indices in 2D plots (e.g., binding efficiency index versus surface efficiency index) has recently been proposed in order to better integrate ligand and target properties and propose optimal paths for fragment design [21].

2.3 Fragment Tethering

Although less common than fragment growing, linking two fragments that are bound independently to adjacent but proximal subsites, is a potent fragment optimization method, with the potential to turn high micromolar fragments into low nanomolar drug-like compounds [9]. The rationale behind this observation is apparent by applying (1) to the tethering of two molecular fragments A and B into a single ligand AB. The connection Gibbs energy arising from the loss of rotational and translation entropy upon binding is relatively independent of molecular weight and was proposed to account for a substantial energy (up to 45 kJ/mol, or ca. eight orders of magnitude in affinity) upon enzyme–substrate complex formation [16]. For drug–target interactions, which are almost always noncovalent, this energy loss was estimated to be around 15–20 kJ/mol (three orders of magnitude) from a few test cases where both binding free energies of fragments and final lead, as well as corresponding high-resolution 3D structures, were available [22]. Linking two fragments into a single lead thus results in a net gain of one connection Gibbs energy and explains why joining two micromolar fragments could give a

nanomolar lead. This estimate of the connection energy was validated by a recent report [23] in which the effect of joining two fragments by the simplest possible linker (one covalent bond, no atoms) without modifying the separate interactions of both fragments, was measured by isothermal titration calorimetry to be −15.23 kJ/mol (2.1×10^{-3} M). This energy gain can only met when: (1) the linker does not induce strain in the interaction of itself and both fragments with the target, and (2) the bound location of both fragments alone recapitulate that in the final lead. These conditions are not always met, as demonstrated by the unproductive fragmentation of a β-lactamase inhibitor into its constituent fragments [24]. In other exceptional cases, the energy gain is even larger than that expected (e.g., avidin) if the binding of one fragment creates a conformational change at a remote subsite favoring the binding of the second fragment [22].

The medicinal chemistry of linkers should thus be seriously taken into consideration when proposing efficient lead compounds from fragments. Strain and conformational flexibility are of course very important criteria for linker design, although their effect may be quite difficult to predict, even in the presence of high-resolution X-ray structures. This was recently demonstrated for a series of uracil DNA glycosylase inhibitors in which decreasing linker strain and flexibility was only beneficial if close to the fragment exhibiting the weakest affinity [25].

3 Fragment Library Design

Computational chemists play an important role in designing smart fragment libraries for FBDD for the simple reason that the total fragment-like chemical space, although far smaller than the drug-like chemical space, is still too large to be screened exhaustively [26]. Hence, a simple enumeration of all chemically stable and synthesizable compounds of less than 13 atoms (C, N, O, S, and Cl atoms) yields 970 million compounds, of which 45% have been estimated to be fragment-like on the basis of simple property ranges [26]. Diversity analysis and prioritization of high-value scaffolds is therefore necessary to design the best possible library according to its potential usage (target family-oriented or general purpose). There are two possible approaches to enumerate fragments, each corresponding to a different design strategy. The forward approach relies on collections of existing or virtual ready-to-screen fragments, whereas the backward approach aims at fragmenting existing or virtual drugs/drug-like compounds into suitable fragments.

3.1 Forward Approach to Library Design

The forward design approach is mainly based on experience accumulated from previous screens to select the most suitable fragments for what are usually fairly

small-sized libraries (<20,000 compounds) [27]. The "rule-of-3" was proposed quite early by Astex as a simple criterion [28], by analogy to the famous "rule-of-5" for selecting soluble and membrane-permeable drug-like compounds [6]. The rule-of-3 stipulates that high-efficiency fragments should fulfill the following properties: molecular weight ≤300, clogP ≤ 3, H-bond donor count ≤3, and H-bond acceptor count ≤3 [28]. Additional property ranges could be used to refine the selection (topological surface area ≤60 $Å^2$, number of rotatable bonds ≤3) [6]. This core definition is currently well accepted but has been refined by various groups, notably for downsizing library sizes and enhancing their practical suitability. For example, Vernalis set up a series of physicochemical filters to remove unwanted moieties (e.g., toxic, chemically reactive) and favor drug-like features [29]. To avoid screening chemically reactive building blocks, Novartis masked potential linking/growing groups [30]. Siegal added a solubility threshold (1 mM) to ensure the selection of water-soluble fragments [27] and a lower limit for molecular weight (> 150) to avoid fragment reorientation upon modification [24]. Matching the resulting fragment space to existing drug-like space is also important to get the best balance between drug-likeness and uniqueness. Venhorst et al. [31] recently reported the design of a high fragment efficiency (HFE) set by extracting, from the 970 million compounds of the GDB-13 virtual library [26], fragment-like compounds that matched target-annotated drug-like compounds at a simple pharmacophore graph level. All the rules mentioned above can now be easily automated in workflows [32] to select a set of chemically diverse fragments whatever the source (screening deck, commercial collections, virtual libraries).

3.2 Backward Approach to Library Design

The backward approach consists of computationally fragmenting existing drugs or drug-like compounds by removing acyclic bonds. It was originally proposed in order to assist combinatorial library design by deconstructing compounds into building blocks. Because building blocks share many interesting properties with fragments (e.g., rule-of-3 compliance), lead fragmentation methods have become popular in recent years, and repurposed to assist the design of fragment libraries. Since fragments may be recombined into novel molecular entities, these methods are also widely used for scaffold hopping.

RECAP (retrosynthetic combinatorial analysis procedure) [33] is the first fragmentation method that has been used widely. It relies on the cleavage of 11 acyclic bond types to generate chemically stable fragments. Interestingly, capping atoms are labeled according to the chemical environment of the parent bond, and thus enable recombination into lead-like structures according to simple chemical reactions. The procedure was refined in the Recore method [34] and recently in BRICS [35] by adding cleavage and fragment filtering rules to remove unwanted fragments, and taking into account 3D structures of the fragments and their possible recombination according to topological (exit vectors) and pharmacophoric constraints. Since such

processing of drug-like libraries gives a large number of possible fragments that usually go beyond FBDD screening capacities, an adequate fragment selection is necessary. After fragmenting drug-like compound collections, Mauser et al. [36] noticed that the maximum coverage of drug-like space after fragment recombination was observed by a mixed selection of fragments based on chemical diversity and frequency of occurrence. Interestingly, splitting the NCI public compound database (>250,000 entries) and computing the resulting fragment co-occurrences gave rise to the observation that some combinations of fragments are observed more frequently than others [37]. Higher than average fragment co-occurrence arises from the synthetic feasibility, versatility and popularity of some chemotypes. Conversely, lower than average co-occurrence of some fragments (e.g., benzene and tetrahydrofuran) may result from the fragment source (e.g., synthetic versus natural origin), synthetic tractability, or bioisosterism reasons. The observations, however, depict holes in the currently accessible chemical space and opportunities for novel molecular frameworks.

4 Binding Site Detection and Structural Druggability

Herein, we will adopt a purely structural view of druggability: the propensity to accommodate high affinity, low molecular weight, drug-like compounds [38]. Obviously, target druggability is not restricted solely to structural aspects, although structural cavity descriptors correlate relatively well with observed NMR-based screening hit rates [39]. We further know that most of the binding free energy of a ligand to a protein cavity is directed towards a few amino acid residues commonly called "hotspots" [40]. Probing a protein surface for regions with favorable interactions with probe atoms or groups is therefore a logical extension of biophysical FBS to very low molecular weight (< 100) fragments or solvent molecules.

4.1 Detection of Hotspots on Protein Surfaces

The experimental method of choice to detect hotspots is the MSCS (multiple solvent crystal structures) method [41]. It consists of soaking protein crystals in a series of solvents (up to six), and determining regions where solvent molecules cluster. Usually, these regions are common to several solvents and correspond to high-affinity ligand-binding pockets. However, it is a lengthy and cumbersome endeavor, and computational surrogates for the MSCS method would significantly reduce the time necessary to propose initial ligands.

A computational analog to MSCS was described in the mid-1980s, far before the first FBDD report, in Goodford's GRID computational approach [42]. GRID can be considered as the precursor of most de novo design and docking algorithms. It maps

probe atoms/groups to regularly spaced grid points and computes the interaction energy with a force-field potential. Preferred positions of small organic probes (atoms or groups) can be identified and the surrounding residues prioritized for ligand design [43, 44]. A variant of GRID was described a few years later in the MCSS (multiple copy simultaneous search) method [45]. Using a combination of Monte Carlo and energy-minimization, MCSS refines the position of many replicas of the same very low molecular weight moiety in the protein cavity. Favorable regions of interaction lead to functionality maps that can be used to bridge and assemble probe moieties into de novo designed ligands [46, 47]. Similar functionality maps could be obtained by knowledge-based approaches (e.g., LUDI [48] and SuperStar [49]), utilizing interaction patterns observed in small-molecule crystal structures.

With the advent of FBDD, binding site detections methods relying on fragment-binding information encountered a revival. The CS-Map (computational solvent mapping) method [50] is the computational approach most analogous to MSCS since it moves a set of 14 organic solvent molecules around the protein surface, refines, clusters them, and identifies consensus regions (hotspots) where various probe clusters overlap. The method was further refined in the FTMap approach using an improved set of probes and an efficient rigid-docking method able to sample billions of probe positions [51]. CS-Map and FTMap were shown to reproduce MSCS experimentally determined solvent consensus binding sites as well as subsites occupied by known inhibitors [51–53]. Both methods have claimed to identify ligand-binding regions with a lower false positive rate than first-generation algorithms [53]. However, translating such functionality maps into newly designed ligands has not yet been reported.

Two main limitations can be noted for the methods described above: (1) the omission of protein flexibility (although multiple structures of the same protein could be read as input), and (2) the approximation of solvation effects. To overcome these limitations, Guvench et al. described a solvent-explicit all-atom molecular dynamics (MD) method (SILCS: site identification by ligand competitive saturation) [54]. SILCS immerses the protein in a high concentration of an aqueous solution of probe groups (propane, benzene), runs multiple MD simulations of the heterogeneous molecular systems, and generates probability maps for the probe groups and water. FBDD knowledge, notably the concept of ligand efficiency (LE) [18], was used in the method for the proper selection of probe group sizes and concentrations. When applied to the computational mapping of the oncoprotein BCL-6, SILCS was able to recapitulate the main interactions observed in the crystal structure of the target protein with two different peptide inhibitors, as well as the known conformational flexibility of one key residue [54]. However, as for any MD-based protocol, the method is computationally demanding, and the value of this extra CPU time has not been established to date. The method does, however, have the potential to be used with other probe groups or fragments to better mirror the experimental conditions observed in biophysical FBS.

The fragmentation logic has been pushed one step further by breaking up both the protein and the ligand in the BUMBLE method [55]. Known protein–ligand X-ray structures can be represented by interacting fragments, the fragment being either side-chain or main-chain heavy atoms for the protein, and any arrangement of covalently linked three atoms for the ligand. The knowledge-based fragment-oriented interaction database is then used to generate spatial distributions for all possible ligand fragments around all possible protein fragments, and therefore to predict interaction hotspots of any ligand to any protein of known 3D structure. The method was able to recover the correct binding site and ligand-bound conformation in ca. 90% and 60% of test cases, respectively [55]. Interestingly, correct interaction predictions do not necessarily arise from protein structures in the training set, which share the same fold or 3D structure with the query protein.

4.2 Druggability Prediction

The computational approaches mentioned above, which locate important hotspots at the protein surface, can provide valuable information for designing novel ligands. They will output information whatever the target, with the only condition being that a cavity with minimal dimensions exists. However, it does not necessarily mean that this information can be translated into high-affinity drug-like ligands – in other words that the cavity is druggable.

Predicting druggability of a target from its 3D atomic coordinates would significantly improve and speed-up target selection in drug discovery programs. In a seminal study, Hadjuk et al. reported a clear correlation between hit rates obtained by NMR screening of fragment libraries and target druggability, estimated from the availability of high-affinity drug-like noncovalent ligands [39]. Interestingly, the same trend was found by a computational approach combining high-throughput fragment docking and molecular mechanics refinement on the same set of targets [56]. A diverse subset of 11,000 fragment-like compounds was derived from the ZINC library [57] and iteratively docked to 152 binding sites of known druggability and 3D structure, the best-scored pose being later refined with a molecular mechanics/generalized Born surface area (MM/GBSA) approach. Following Hadjuk's classification of druggable and non-druggable binding sites [39], an acceptable correlation was found between computed hit rates (a hit being any pose finally scored above a -40 kcal/mol energy cutoff) and NMR-derived hit rates. Binary classification of cavities according to a druggability score threshold [$-\log($hit rate$)$] could correctly distinguish 37 druggable from 35 non-druggable targets in an external test set [56].

This latter computational approach is computationally more demanding compared with much faster geometry-based and/or energy-based computational methods [58, 59]. Nevertheless, it confirms the very important observation that druggable cavities can accommodate low molecular weight fragments with high hit rates.

5   Fragment Identification and Optimization

Computational surrogates for fragment screening have existed for a while, with the two main technologies being de novo design and docking. For a detailed discussion about methods and practical applications, the reader is referred to recent reviews in the field [60–62]. De novo ligand design aims at reproducing the entire process (fragment positioning, growing, joining or linking) of FBDD but has been rarely successful in delivering high-affinity leads for the main reason that these approaches were mainly computer-driven and did not take into account the necessary experimental knowledge to properly select and score low molecular weight fragments. Likewise, for a long time, docking was only applied to drug-like or lead-like compounds. Starting from the mid-1990s, the development of fragment screening by biophysical methods has put the spotlight on these two computational methods again and fostered novel developments that will be reviewed from here on.

5.1   Docking Approach

Docking ligands to protein active sites was pioneered by USCF Kuntz's group in the early 1980s [63] and numerous methods have been constantly developed since that time, progressively taking into account ligand and protein flexibility [64], water-mediated effects [65], and above-all the development of fast scoring functions to predict binding free energies [62]. At this point, we should distinguish the docking of compounds belonging to fragment chemical space from the impact of fragment-based design on the docking of bigger molecules. Although many of the basic concepts are similar in both scenarios, conformational sampling and scoring issues may be quite different, resulting in different levels of accuracy in docking.

5.1.1   Fragment-Based Approaches to Docking

Many docking programs build up the ligand on-the-fly by first splitting the compound into its basic fragments, then docking a reference fragment (or base fragment) by geometric matching, and incrementally reconstructing the full ligand by adding fragments stepwise until the full molecule has been built [62]. In order to speed up the process, only the most favorable values of a specific dihedral angle are browsed in a look-up table (e.g., MINUMBA conformer library [66]), the selected pose being that scored the best by a scoring function. An advantage of this method with respect to rigid docking of conformers is that the bound conformation of the ligand is generated in the presence of the target. However, if the bound conformer exhibits one or more unexpected torsional angles, incremental build-up techniques are unlikely to find a correct solution. Moreover, since many compounds share identical base fragments (e.g., heterocycles), there is an unnecessary cost in repeating the

same docking operation on simple fragments. It should be mentioned, however, that combinatorial docking algorithms [67] can take advantage of the fact that compounds share the same base fragment. Generating and storing all possible docking poses of all fragments up-front, and then looking whether they have already been docked can significantly speed up the process [68]. Linkers must then be added to connect these fragments together, the full molecule being refined in a last step within the active site. The suitability of the incremental construction approach to docking-based virtual screening of chemical libraries is now well established and many success stories have been reported during the last decade [69].

5.1.2 Docking in Fragment Space

For a long time, all benchmark studies addressing either docking accuracy, scoring accuracy, or virtual screening accuracy have focused on drug-like compounds exclusively. The very first report on fragment docking by Verdonk et al. in 2004 [70] noticed that low molecular weight (< 250) CDK2 inhibitors were much more difficult to retrieve among top scoring hits than higher molecular weight (> 250) CDK2 inhibitors when seeded with decoys for docking-based library screening. Using knowledge-based pharmacophore constraints was helpful in enhancing fragment hit rates [70]. This seminal study was confirmed a couple of years later by the first benchmark on fragment docking applied to a set of 42 diverse fragment-like ligands in the Protein Data Bank (PDB) [71]. Three main explanations have been proposed for this observation [61, 71]: (1) the energy landscape for fragment poses is relatively flat with respect to larger molecules and scoring functions cannot distinguish near-isoenergetic but different solutions to the problem of fragment docking (Fig. 1, left); (2) classic metrics [e.g., root mean square deviations (rmsd) from the X-ray pose] do not always reflect conservation of key molecular interactions (Fig. 1, right); and (3) most scoring functions have been trained on sets of protein X-ray structures co-crystallized with drug-like and not fragment-like ligands.

Extending the scope of fragment docking was later reported for the highly polar rRNA A-site [72]. Retrospective virtual screening of fragment libraries seeded with known rRNA A-site inhibitors using GOLD [73] and GLIDE [74] docking tools afforded enrichments in true binders higher than random picking, but only moderate early enrichments. The largest retrospective study on fragment docking has been applied to 190 protein fragment complexes from the PDB [75] with GLIDE5.0 as docking tool. On average, a correct docking pose (rmsd to the X-ray pose <2.0 Å) was found in 80% of the cases, although there was noticeable variation according to ligand and binding site properties. Docking lipophilic fragments to open binding sites was much more challenging than docking polar compounds in closed and polar cavities [75]. Cross-docking experiments (docking into X-ray coordinates of the cognate protein co-crystallized with another fragment) were less accurate than self-docking, suggesting that binding site flexibility also plays an important role for low molecular weight compounds. As expected, predicting absolute binding free
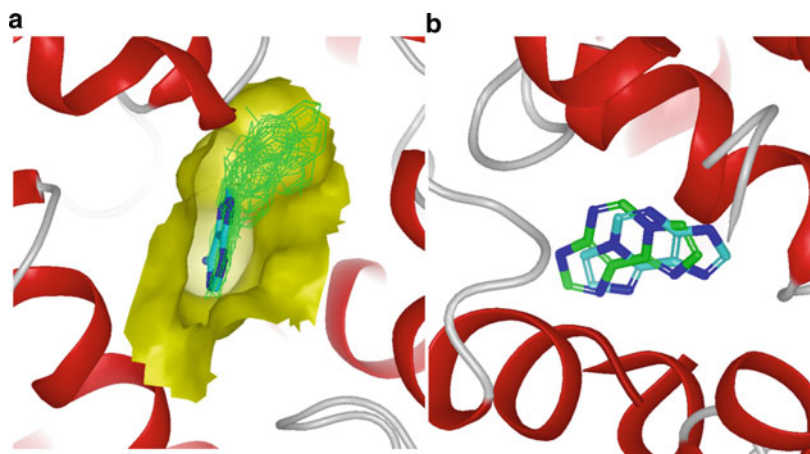
Fig. 1 (**a**) X-ray (*sticks*) and Glide-predicted poses (*green lines*) of a small molecular weight fragment (adenine) to phosphodiesterase 4D. The fragment pocket is delimited by a *yellow surface* and X-ray fragment coordinates extracted from adenosine monophosphate (PDB entry 1tb7). (**b**) FlexX docking of 3H-imidazo[2,1-I]purine to 3-methyladenine DNA glycosylase, (PDB entry 1pu8) produces a pose with high rmsd to the crystal pose (*cyan* carbons, rmsd = 3.90 Å) although 80% of protein–ligand interactions are conserved

Table 2 Fragment hits identified by docking

Target	Library size	Optimization	Hits/tested compounds	Reference
DNA gyrase	350,000	Yes	150/3,000	[76]
Thrombin	7,050	Yes	3/8	[77]
Inosine 5′-monophosphate dihydrogenase	3,425	No	7/74	[78]
L-Xylose reductase	236,000	No	12/39	[79]
Aurora A kinase	70,000	Yes	4/7	[80]
Dipeptidyl peptidase IV	10,000	No	12/14	[81]
AMPC β-lactamase	137,000	No	23/48	[82]
Anthrax edema factor	10,000	No	4/19	[83]
CTX-M β-lactamase	300,000	No	10/69	[84]
Macrophage inhibitory factor	n.a.[a]	No	3/23	[85]
Thermolysin	434	No	2/19	[86]
6-Phosphogluconate dehydrogenase	64,000	No	3/71	[87]
Indoleamine 2,3-dioxygenase	15	Yes	8/15	[88]
NF-κB inducing kinase	67,489	No	2/49	[89]

[a]Not available

energies is out of reach, even when the binding pose is correct [75]. Although the latter study is promising, the real impact of docking on fragment identification can be deduced from reported success stories. We could trace back 14 reports (Table 2) describing the identification of fragment hits upon database docking.

Interestingly, these success stories all have a few characteristics in common:

1. The binding site of interest was small, polar and usually lined with positively charged residues (e.g., S_1 pocket of proteases)
2. Docking was constrained to reproduce the crucial hydrogen bonds observed with substrate analogs or factors
3. The fragment hit was significantly buried in its binding site
4. Validated hits were mostly carboxylic acids or bioisosteres (tetrazole, sulfonamides)

In only a few cases [76, 77, 80, 88] could a structure-guided optimization of the validated fragment hits be undertaken, which led to more potent drug-like compounds. One can therefore see these successes as docking cases where constraints imposed by the topology of the binding site and the strong directionality of key protein–ligand hydrogen bonds favored the identification of fragment hits.

Hopefully, there is plenty of room available for improving canonical docking approaches to fragment identification. First, fragment-specific scoring functions could be developed to better take into account the peculiarity of fragments (low number of H-bond acceptors and donors, few rotatable bonds, high rotational/translational entropy) [75]. More accurate computational methods to estimate binding free energies, at the cost of a lower throughput, could also be used to rescore fragment docking poses [90]. Nevertheless, the benefit of using better free energy prediction methods is a matter of debate. MM-GBSA rescoring with binding site minimization was shown on "toy" cavities [91] (small buried cavities for fragment-like compounds dominated by a single intermolecular interaction) to rescue false negatives (allowing bigger ligands to dock by cavity refinement) while producing additional false positives among top-ranked molecules. It highlights the danger of significantly decreasing hit rates if only a very tiny percentage of top-ranked compounds are tested experimentally. Two reasons were invoked for this intriguing observation: inadequate ligand parameterization, and difficulties in treating electrostatic interactions.

When binding modes of true hits are known, knowledge-based rescoring is an easy and powerful scoring method [92]. The basic idea is that true hits generally share key protein–ligand interactions, and that docking poses fulfilling these requirements are more likely to yield experimentally validated hits. This principle was verified by converting protein–ligand coordinates into interaction fingerprints (IFPs) [71, 93]. IFPs are bit strings registering intermolecular interactions (aromatic and hydrophobic contacts, hydrogen bonds, salt bridges, metal coordination) between a ligand and the surrounding binding site (Fig. 2).

Rescoring docking poses by IFP similarity to known references was shown to significantly improve the selection of the top-ranked pose for fragment-like compounds and to enable scaffold docking [71]. In contrast to what might be expected, IFP scoring does not prioritize virtual hits with high 2D similarity to a reference and is suitable for simple hit ranking [93–95], scaffold posing [71], and scaffold hopping [96]. IFP rescoring provides a logical link between FBDD and virtual screening because X-ray poses of fragment hits represent unbiased IFPs for

Fig. 2 Converting a reference protein–ligand complex (*REF*, *bottom left*) into an interaction fingerprint (IFP) registering seven possible interactions with binding site residues. Docked poses (*bottom right*) are fingerprinted and the corresponding IFPs compared to the reference IFP using a standard Tanimoto coefficient. The best pose (*pose 6*, *bottom center*) exhibits the highest IFP similarity to the reference

prioritizing compounds in a later in silico screen for either novel fragment scaffolds or drug-like compounds.

The structural post-processing of docking poses described above does, however, require preliminary knowledge on the binding mode of one true hit. Three promising approaches to fragment docking in the absence of this information have been reported recently. In the first approach, fragment poses produced by docking can be converted into pharmacophoric queries for database screening [97]. An atomic-based energy decomposition of fragment poses was reported upon docking a library of 648 test fragments to seven targets. After mapping interacting atoms to pharmacophoric features, the most energetically favored features were later clustered on the basis of volume overlap to yield reasonably simple and specific queries that proved to be capable of discriminating true actives from decoys in pharmacophore searches [97]. The second application is named fragment screening by replica generation (FSRG) [98]. Instead of docking a library of fragments, this method first generates a set of replicas for each fragment by attaching six side chains to a diversity point and generates higher molecular weight compounds, which are expected to be easier to dock. Analogs are classically docked and initial scaffolds are ordered according to the ranks of their corresponding replicas. Interestingly, polar interactions (Coulombic, hydrogen bonding) were ignored when

ranking replicas and only shape complementarity guided scaffold ranking. The FSRG method was shown to be able to recover true active scaffolds from inhibitors of six different target proteins [98]. The third approach, multiple ligand simultaneous docking (MLSD) [99], aims at simulating the multiple fragment soaking process. Up to four small molecular weight fragments are simultaneously docked using either a Lamarckian genetic algorithm or a particle swarm optimization (PSO) method, considering the input ligands as an additional variable. The advantage of docking multiple ligands simultaneously is that concerted actions of fragments approaching the binding site are better reproduced than sequential docking of single fragments. The method was successfully applied to reproduce the pose of four fragments derived from the ABT-737 inhibitor to the Bcl-xL binding pocket [99]. More applications are required to evaluate the true potential of the MLSD method, which could in theory be applied to docking libraries of pooled fragments.

5.2   De Novo Ligand Design

De novo design algorithms aim to design from scratch small molecular weight ligands that perfectly fulfill steric and electrostatic requirements imposed by the 3D architecture of a binding cavity [60, 97]. They appeared in the early 1990s with the hope that the simple availability of a protein 3D structure and current knowledge of protein–ligand interactions would suffice to produce high-affinity ligands. Although atom-growing approaches have been described, most methods rely on posing and modifying (linking, joining, growing) a set of low molecular weight fragments. Early de novo design programs (1990–1995) were mainly driven by maximizing protein–ligand interactions, with the protein structure as the sole constraint to guide ligand build-up. Despite intense efforts in methodological developments, it appeared quite soon that the outcome of de novo design algorithms was usually disappointing, resulting in highly complex but low-affinity ligands, and always required human expertise for assisting and/or simplifying generated compounds [100, 101]. The main reasons for this poor performance were that the algorithms lacked secondary constraints [e.g., synthetic accessibility, physicochemical and ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties, and knowledge of structure–activity relationships for existing ligands]. Since the designed ligands were mostly low micromolar affinity ligands with often poor synthetic accessibility and pharmacokinetic properties, computational chemists relied on database searching tools (such as similarity measurement, pharmacophore search, and docking) to subsequently identify starting hits that were either purchasable or for which synthetic steps were available. The spectacular success of FBDD in the late 1990s put de novo design methods again under the spotlight since they are basically computational surrogates of biophysical fragment screening. From 2000 on, novel or simplified versions of existing algorithms were reported again in the literature, but with two fundamental changes in their philosophy: (1)

Table 3 Novel ligands identified by de novo design methods

Target	Program	Method	Reference
FKBP12	Ludi	Fragment linking	[100]
Thrombin	Ludi	Fragment linking	[102]
Cyclin-dependent kinase 4	Legend	Atom growing	[103]
Kv1.5 potassium channel	Topas	Ligand-based evolutionary algorithm	[104]
Carbonic anhydrase II	SMoG	Fragment growing	[105]
Lanosterol 14-α demethylase	MCSS/Ludi	Probe mapping + fragment linking	[106]
HIV reverse transcriptase	Synopsis	Ligand-based evolutionary algorithm	[107]
Cannabinoid CB1 receptor	Topas	Ligand-based evolutionary algorithm	[108]
HIV-1 protease	Breed	Fragment hybridization	[109]
p38 kinase	Breed	Fragment hybridization	[109]
Phosphodiesterase-4	SLF-Libmaker	Fragment growing	[110]
Dihydroorotate dihydrogenase	Sprout	Fragment linking	[111]
Histamine H3 receptor	Skelgen	Fragment linking	[112]
Protein kinases	In-house	Fragment linking	[113]

that the algorithms should be considered as novel idea generators rather than ligand design methods [97], and (2) that one should accept the statement in Schneider's review [60] that "de novo design will rarely yield novel structures with nanomolar activity in the first instance." With different starting expectations and more secondary constraints (synthetic tractability, drug-likeness, chemical diversity optimization), some significant successes have been noticed (see a non-exhaustive list in Table 3) but still at a lower rate compared with virtual screening methods.

Simpler but more pragmatic ligand-based approaches to fragment-based design have been successfully reported. In the BREED approach [109], protein-bound ligands are overlaid, fragmented, and recombined according to overlapping bonds, to produce high-affinity chimeras for diverse targets [109]. In COREGEN [114], kinase inhibitors were simply reengineered in a combinatorial manner from four rings and eight linkers frequently occurring in protein kinase inhibitors and docked under constraints to propose novel inhibitors. Vieth et al. reported a method to decompose kinase inhibitors into hinge binding cores, hydrophobic groups, and solubilizing moieties [113]. Recombining these building blocks into small-sized libraries of 20–50 compounds produced compound collections highly enriched in true inhibitors for previously unscreened kinases. In SLF-LibMaker [110], an additional layer of building elements is defined. Starting from a known 0.8 μM phosphodiesterase inhibitor (zardaverine), a combinatorial library of 320 compounds was designed by adding at a few positions carefully selected by medicinal chemists, five linkers and 16 groups representing non-redundant pharmacophoric properties. The primary role of the linker was to vary the distance between the core scaffold and the functional group and therefore enable scanning the protein pocket for additional subsites. This method produced a subnanomolar PDE4 inhibitor with a 900-fold increased affinity within a single iteration [110].

5.3 Miscellaneous

Novartis recently reported a very interesting approach (virtual fragment linking, VFL) to couple fragment screening with high-throughput screening (HTS) data in order to prioritize the best possible hits [115]. The underlying idea is to take advantage of existing fragment hits to identify, by Bayesian statistical modeling, fragment substructures (features) frequently found in HTS hits. Hits exhibiting these structural features are further prioritized by a score computed from the Bayesian model. The method is particularly interesting since it trains on low-affinity ligands in fragment space, but predicts high-affinity compounds in drug-like space. Of course, both spaces need to overlap, which means that the training fragment library should recapitulate structural features present in the HTS screening deck. When applied to seven targets for which both fragment and HTS screens were available, the method found between 28% and 67% of existing low micromolar hits among the top 5% of the ranked full library for four of the targets. The VFL approach offers novel opportunities to design focused libraries of lead-like or drug-like compounds from FBS data. Fragment substructures are excellent chemical descriptors and have been implemented in numerous chemical fingerprints [116] to infer pairwise ligand similarity [117], target profiles [118], and even adverse drug reactions [119].

6 Concluding Remarks

As a key technology to provide medicinal and computational chemists with efficient hits, FBS methods have clearly impacted computational approaches to drug discovery during the last 10 years. Firstly, many computational methods (e.g., de novo design, fragment-based docking, detection of binding sites on protein surfaces) that were developed before the advent of FBDD have been revisited and refined while taking advantage of knowledge generated by FBS. Secondly, novel computational algorithms have been designed to address needs coming from FBDD (e.g., fragment space browsers, fragment library design, fragment-to-lead optimization methods). With the increasing availability of FBS data, more improvements should come, including customized scoring functions for docking low molecular weight compounds.

Beside these technological aspects, perhaps the greatest consequence has been the change in the mindset of drug designers. The answer to the basic question "What is a biologically attractive compound?" has significantly evolved in the last 15 years. Focusing on efficiency rather than affinity has revolutionized common practices in computational and medicinal chemistry. Thinking in fragment space has mainly been restricted to the early drug discovery stages (hit identification, hit-to-lead optimization). Some application to later preclinical stages (e.g., side effect predictions, preclinical safety profiling) has begun to appear and will complement the existing tools in the very near future.

References

1. Shuker SB, Hajduk PJ, Meadows RP et al (1996) Discovering high-affinity ligands for proteins: SAR by NMR. Science 274:1531–1534
2. Kola I, Landis J (2004) Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov 3:711–715
3. Venter JC, Adams MD, Myers EW et al (2001) The sequence of the human genome. Science 291:1304–1351
4. Schreiber SL (2000) Target-oriented and diversity-oriented organic synthesis in drug discovery. Science 287:1964–1969
5. Pereira DA, Williams JA (2007) Origin and evolution of high throughput screening. Br J Pharmacol 152:53–61
6. Lipinski CA, Lombardo F, Dominy BW et al (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46:3–26
7. Hann MM, Leach AR, Harper G (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. J Chem Inf Comput Sci 41:856–864
8. Congreve M, Marshall F (2010) The impact of GPCR structures on pharmacology and structure-based drug design. Br J Pharmacol 159:986–996
9. Congreve M, Chessari G, Tisi D et al (2008) Recent developments in fragment-based drug discovery. J Med Chem 51:3661–3680
10. Murray CW, Rees DC (2009) The rise of fragment-based drug discovery. Nat Chem 1:187–192
11. Orita M, Warizaya M, Amano Y et al (2009) Advances in fragment-based drug discovery platforms. Expert Opin Drug Discov 4:1125–1144
12. Warr WA (2009) Fragment-based drug discovery. J Comput Aided Mol Des 23:453–458
13. Erlanson DA (2006) Fragment-based lead discovery: a chemical update. Curr Opin Biotechnol 17:643–652
14. Hajduk PJ, Greer J (2007) A decade of fragment-based drug design: strategic advances and lessons learned. Nat Rev Drug Discov 6:211–219
15. Law R, Barker O, Barker JJ et al (2009) The multiple roles of computational chemistry in fragment-based drug design. J Comput Aided Mol Des 23:459–473
16. Jencks WP (1981) On the attribution and additivity of binding energies. Proc Natl Acad Sci USA 78:4046–4050
17. Kuntz ID, Chen K, Sharp KA et al (1999) The maximal affinity of ligands. Proc Natl Acad Sci USA 96:9997–10002
18. Hopkins AL, Groom CR, Alex A (2004) Ligand efficiency: a useful metric for lead selection. Drug Discov Today 9:430–431
19. Reynolds CH, Bembenek SD, Tounge BA (2007) The role of molecular size in ligand efficiency. Bioorg Med Chem Lett 17:4258–4261
20. Orita M, Ohno K, Niimi T (2009) Two 'Golden Ratio' indices in fragment-based drug discovery. Drug Discov Today 14:321–328
21. Abad-Zapatero C, Perisic O, Wass J et al (2010) Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an atlas-like representation. Drug Discov Today 15:804–811
22. Murray CW, Verdonk ML (2002) The consequences of translational and rotational entropy lost by small molecules on binding to proteins. J Comput Aided Mol Des 16:741–753
23. Borsi V, Calderone V, Fragai M et al (2010) Entropic contribution to the linking coefficient in fragment based drug design: a case study. J Med Chem 53:4285–4289
24. Babaoglu K, Shoichet BK (2006) Deconstructing fragment-based inhibitor discovery. Nat Chem Biol 2:720–723
25. Chung S, Parker JB, Bianchet M et al (2009) Impact of linker strain and flexibility in the design of a fragment-based inhibitor. Nat Chem Biol 5:407–413

26. Blum LC, Reymond JL (2009) 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. J Am Chem Soc 131:8732–8733
27. Siegal G, Ab E, Schultz J (2007) Integration of fragment screening and library design. Drug Discov Today 12:1032–1039
28. Congreve M, Carr R, Murray C et al (2003) A 'rule of three' for fragment-based lead discovery? Drug Discov Today 8:876–877
29. Baurin N, Aboul-Ela F, Barril X et al (2004) Design and characterization of libraries of molecular fragments for use in NMR screening against protein targets. J Chem Inf Comput Sci 44:2157–2166
30. Schuffenhauer A, Ruedisser S, Marzinzik AL et al (2005) Library design for fragment based screening. Curr Top Med Chem 5:751–762
31. Venhorst J, Núñez S, Kruse CG (2010) Design of a high fragment efficiency library by molecular graph theory. ACS Med Chem Lett 1:499–503
32. Gianti E, Sartori L (2008) Identification and selection of "privileged fragments" suitable for primary screening. J Chem Inf Model 48:2129–2139
33. Lewell XQ, Judd DB, Watson SP et al (1998) RECAP–retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. J Chem Inf Comput Sci 38:511–522
34. Maass P, Schulz-Gasch T, Stahl M et al (2007) Recore: a fast and versatile method for scaffold hopping based on small molecule crystal structure conformations. J Chem Inf Model 47:390–399
35. Degen J, Wegscheid-Gerlach C, Zaliani A et al (2008) On the art of compiling and using 'drug-like' chemical fragment spaces. ChemMedChem 3:1503–1507
36. Mauser H, Stahl M (2007) Chemical fragment spaces for de novo design. J Chem Inf Model 47:318–324
37. Lameijer EW, Kok JN, Back T et al (2006) Mining a chemical database for fragment co-occurrence: discovery of "chemical cliches". J Chem Inf Model 46:553–562
38. Hopkins AL, Groom CR (2002) The druggable genome. Nat Rev Drug Discov 1:727–730
39. Hajduk PJ, Huth JR, Fesik SW (2005) Druggability indices for protein targets derived from NMR-based screening data. J Med Chem 48:2518–2525
40. Ciulli A, Williams G, Smith AG et al (2006) Probing hot spots at protein-ligand binding sites: a fragment-based approach using biophysical methods. J Med Chem 49:4992–5000
41. Mattos C, Ringe D (1996) Locating and characterizing binding sites on proteins. Nat Biotechnol 14:595–599
42. Goodford PJ (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J Med Chem 28:849–857
43. Rognan D, Scapozza L, Folkers G et al (1995) Rational design of nonnatural peptides as high-affinity ligands for the HLA-B*2705 human leukocyte antigen. Proc Natl Acad Sci USA 92:753–757
44. von Itzstein M, Dyason JC, Oliver SW et al (1996) A study of the active site of influenza virus sialidase: an approach to the rational design of novel anti-influenza drugs. J Med Chem 39:388–391
45. Miranker A, Karplus M (1991) Functionality maps of binding sites: a multiple copy simultaneous search method. Proteins 11:29–34
46. Eisen MB, Wiley DC, Karplus M et al (1994) HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. Proteins 19:199–221
47. Schubert CR, Stultz CM (2009) The multi-copy simultaneous search methodology: a fundamental tool for structure-based drug design. J Comput Aided Mol Des 23:475–489
48. Bohm HJ (1992) The computer program LUDI: a new method for the de novo design of enzyme inhibitors. J Comput Aided Mol Des 6:61–78
49. Verdonk ML, Cole JC, Taylor R (1999) SuperStar: a knowledge-based approach for identifying interaction sites in proteins. J Mol Biol 289:1093–1108

50. Dennis S, Kortvelyesi T, Vajda S (2002) Computational mapping identifies the binding sites of organic solvents on proteins. Proc Natl Acad Sci USA 99:4290–4295
51. Brenke R, Kozakov D, Chuang GY et al (2009) Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. Bioinformatics 25:621–627
52. Chuang GY, Kozakov D, Brenke R et al (2009) Binding hot spots and amantadine orientation in the influenza a virus M2 proton channel. Biophys J 97:2846–2853
53. Landon MR, Lieberman RL, Hoang QQ et al (2009) Detection of ligand binding hot spots on protein surfaces via fragment-based methods: application to DJ-1 and glucocerebrosidase. J Comput Aided Mol Des 23:491–500
54. Guvench O, MacKerell AD Jr (2009) Computational fragment-based binding site identification by ligand competitive saturation. PLoS Comput Biol 5:e1000435
55. Kasahara K, Kinoshita K, Takagi T (2010) Ligand-binding site prediction of proteins based on known fragment-fragment interactions. Bioinformatics 26:1493–1499
56. Huang N, Jacobson MP (2010) Binding-site assessment by virtual fragment screening. PLoS One 5:e10109
57. Irwin JJ, Shoichet BK (2005) ZINC–a free database of commercially available compounds for virtual screening. J Chem Inf Model 45:177–182
58. Halgren TA (2009) Identifying and characterizing binding sites and assessing druggability. J Chem Inf Model 49:377–389
59. Schmidtke P, Barril X (2010) Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. J Med Chem 53:5858–5867
60. Schneider G, Fechner U (2005) Computer-based de novo design of drug-like molecules. Nat Rev Drug Discov 4:649–663
61. Loving K, Alberts I, Sherman W (2010) Computational approaches for fragment-based and de novo design. Curr Top Med Chem 10:14–32
62. Moitessier N, Englebienne P, Lee D et al (2008) Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. Br J Pharmacol 153(Suppl 1):S7–S26
63. Kuntz ID, Blaney JM, Oatley SJ et al (1982) A geometric approach to macromolecule-ligand interactions. J Mol Biol 161:269–288
64. B-Rao C, Subramanian J, Sharma SD (2009) Managing protein flexibility in docking and its applications. Drug Discov Today 14:394–400
65. Rarey M, Kramer B, Lengauer T (1999) The particle concept: placing discrete water molecules during protein-ligand docking predictions. Proteins 34:17–28
66. Klebe G, Mietzner T (1994) A fast and efficient method to generate biologically relevant conformations. J Comput Aided Mol Des 8:583–606
67. Sun Y, Ewing TJ, Skillman AG et al (1998) CombiDOCK: structure-based combinatorial docking and library design. J Comput Aided Mol Des 12:597–604
68. Zsoldos Z, Reid D, Simon A et al (2007) eHiTS: a new fast, exhaustive flexible ligand docking system. J Mol Graph Model 26:198–212
69. Huang D, Caflisch A (2010) Library screening by fragment-based docking. J Mol Recognit 23:183–193
70. Verdonk ML, Berdini V, Hartshorn MJ et al (2004) Virtual screening using protein-ligand docking: avoiding artificial enrichment. J Chem Inf Comput Sci 44:793–806
71. Marcou G, Rognan D (2007) Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. J Chem Inf Model 47:195–207
72. Li Y, Shen J, Sun X et al (2010) Accuracy assessment of protein-based docking programs against RNA targets. J Chem Inf Model 50:1134–1146
73. Verdonk ML, Cole JC, Hartshorn MJ et al (2003) Improved protein-ligand docking using GOLD. Proteins 52:609–623
74. Friesner RA, Banks JL, Murphy RB et al (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 47:1739–1749

75. Sandor M, Kiss R, Keseru GM (2010) Virtual fragment docking by Glide: a validation study on 190 protein-fragment complexes. J Chem Inf Model 50:1165–1172

76. Boehm HJ, Boehringer M, Bur D et al (2000) Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. J Med Chem 43:2664–2674

77. Makino S, Kayahara T, Tashiro K et al (2001) Discovery of a novel serine protease inhibitor utilizing a structure-based and experimental selection of fragments technique. J Comput Aided Mol Des 15:553–559

78. Pickett SD, Sherborne BS, Wilkinson T et al (2003) Discovery of novel low molecular weight inhibitors of IMPDH via virtual needle screening. Bioorg Med Chem Lett 13:1691–1694

79. Carbone V, Ishikura S, Hara A et al (2005) Structure-based discovery of human L-xylulose reductase inhibitors from database screening and molecular docking. Bioorg Med Chem 13:301–312

80. Warner SL, Bashyam S, Vankayalapati H et al (2006) Identification of a lead small-molecule inhibitor of the Aurora kinases using a structure-assisted, fragment-based approach. Mol Cancer Ther 5:1764–1773

81. Rummey C, Nordhoff S, Thiemann M et al (2006) In silico fragment-based discovery of DPP-IV S1 pocket binders. Bioorg Med Chem Lett 16:1405–1409

82. Teotico DG, Babaoglu K, Rocklin GJ et al (2009) Docking for fragment inhibitors of AmpC beta-lactamase. Proc Natl Acad Sci USA 106:7455–7460

83. Chen D, Misra M, Sower L et al (2008) Novel inhibitors of anthrax edema factor. Bioorg Med Chem 16:7225–7233

84. Chen Y, Shoichet BK (2009) Molecular docking and ligand specificity in fragment-based inhibitor discovery. Nat Chem Biol 5:358–364

85. McLean LR, Zhang Y, Li H et al (2010) Fragment screening of inhibitors for MIF tautomerase reveals a cryptic surface binding site. Bioorg Med Chem Lett 20:1821–1824

86. Englert L, Silber K, Steuber H et al (2010) Fragment-based lead discovery: screening and optimizing fragments for thermolysin inhibition. ChemMedChem 5:930–940

87. Ruda GF, Campbell G, Alibu VP et al (2010) Virtual fragment screening for novel inhibitors of 6-phosphogluconate dehydrogenase. Bioorg Med Chem 18:5056–5062

88. Rohrig UF, Awad L, Grosdidier A et al (2010) Rational design of indoleamine 2,3-dioxygenase inhibitors. J Med Chem 53:1172–1189

89. Mortier J, Masereel B, Remouchamps C et al (2010) NF-kappaB inducing kinase (NIK) inhibitors: identification of new scaffolds using virtual screening. Bioorg Med Chem Lett 20:4515–4520

90. Gleeson MP, Gleeson D (2009) QM/MM as a tool in fragment based drug discovery. A cross-docking, rescoring study of kinase inhibitors. J Chem Inf Model 49:1437–1448

91. Graves AP, Shivakumar DM, Boyce SE et al (2008) Rescoring docking hit lists for model cavity sites: predictions and experimental testing. J Mol Biol 377:914–934

92. Novikov FN, Stroylov VS, Stroganov OV et al (2010) Improving performance of docking-based virtual screening by structural filtration. J Mol Model 16:1223–1230

93. Deng Z, Chuaqui C, Singh J (2004) Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions. J Med Chem 47:337–344

94. Kelly MD, Mancera RL (2004) Expanded interaction fingerprint method for analyzing ligand binding modes in docking and structure-based drug design. J Chem Inf Comput Sci 44:1942–1951

95. Mpamhanga CP, Chen B, McLay IM et al (2006) Knowledge-based interaction fingerprint scoring: a simple method for improving the effectiveness of fast scoring functions. J Chem Inf Model 46:686–698

96. Venhorst J, Nunez S, Terpstra JW et al (2008) Assessment of scaffold hopping efficiency by use of molecular interaction fingerprints. J Med Chem 51:3222–3229

97. Loving K, Salam NK, Sherman W (2009) Energetic analysis of fragment docking and application to structure-based pharmacophore hypothesis generation. J Comput Aided Mol Des 23:541–554

98. Fukunishi Y, Mashimo T, Orita M et al (2009) In silico fragment screening by replica generation (FSRG) method for fragment-based drug design. J Chem Inf Model 49:925–933
99. Li H, Li C (2010) Multiple ligand simultaneous docking: orchestrated dancing of ligands in binding sites of protein. J Comput Chem 31:2014–2022
100. Babine RE, Bleckman TM, Kissinger CR et al (1995) Design, synthesis and X-ray crystallographic studies of novel FKBP-12 ligands. Bioorg Med Chem Lett 5:1719–1724
101. Rich DH, Bohacek RS, Dales NA et al (1997) Transformation of peptides into non-peptides. Synthesis of computer-generated enzyme inhibitors. Chimia 51:45–47
102. Bohm HJ, Banner DW, Weber L (1999) Combinatorial docking and combinatorial chemistry: design of potent non-peptide thrombin inhibitors. J Comput Aided Mol Des 13:51–56
103. Honma T, Hayashi K, Aoyama T et al (2001) Structure-based generation of a new class of potent Cdk4 inhibitors: new de novo design strategy and library design. J Med Chem 44:4615–4627
104. Schneider G, Lee ML, Stahl M et al (2000) De novo design of molecular architectures by evolutionary assembly of drug-derived building blocks. J Comput Aided Mol Des 14:487–494
105. Grzybowski BA, Ishchenko AV, Kim CY et al (2002) Combinatorial computational method gives new picomolar ligands for a known enzyme. Proc Natl Acad Sci USA 99:1270–1273
106. Ji H, Zhang W, Zhang M et al (2003) Structure-based de novo design, synthesis, and biological evaluation of non-azole inhibitors specific for lanosterol 14alpha-demethylase of fungi. J Med Chem 46:474–485
107. Vinkers HM, de Jonge MR, Daeyaert FF et al (2003) SYNOPSIS: SYNthesize and OPtimize System in Silico. J Med Chem 46:2765–2773
108. Rogers-Evans M, Alanine AI, Bleicher KH et al (2004) Identification of novel cannabinoid receptor ligands via evolutionary de novo design and rapid parallel synthesis. QSAR & Comb Sci 23:426–430
109. Pierce AC, Rao G, Bemis GW (2004) BREED: Generating novel inhibitors through hybridization of known ligands. Application to CDK2, p38, and HIV protease. J Med Chem 47:2768–2775
110. Krier M, Araujo-Junior JX, Schmitt M et al (2005) Design of small-sized libraries by combinatorial assembly of linkers and functional groups to a given scaffold: application to the structure-based optimization of a phosphodiesterase 4 inhibitor. J Med Chem 48:3816–3822
111. Heikkila T, Thirumalairajan S, Davies M et al (2006) The first de novo designed inhibitors of *Plasmodium falciparum* dihydroorotate dehydrogenase. Bioorg Med Chem Lett 16:88–92
112. Roche O, Rodriguez Sarmiento RM (2007) A new class of histamine H3 receptor antagonists derived from ligand based design. Bioorg Med Chem Lett 17:3670–3675
113. Vieth M, Erickson J, Wang J et al (2009) Kinase inhibitor data modeling and de novo inhibitor design with fragment approaches. J Med Chem 52:6456–6466
114. Aronov AM, Bemis GW (2004) A minimalist approach to fragment-based ligand design using common rings and linkers: application to kinase inhibitors. Proteins 57:36–50
115. Crisman TJ, Bender A, Milik M et al (2008) "Virtual fragment linking": an approach to identify potent binders from low affinity fragment hits. J Med Chem 51:2481–2491
116. Willett P (2006) Similarity-based virtual screening using 2D fingerprints. Drug Discov Today 11:1046–1053
117. Bender A, Mussa HY, Gill GS et al (2004) Molecular surface point environments for virtual screening and the elucidation of binding patterns (MOLPRINT 3D). J Med Chem 47:6569–6583
118. Nidhi GM, Davies JW et al (2006) Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. J Chem Inf Model 46:1124–1133
119. Clark M, Wiseman JS (2009) Fragment-based prediction of the clinical occurrence of long QT syndrome and torsade de pointes. J Chem Inf Model 49:2617–2626

Index