

CHAPTER 1

3D Structure and the Drug Discovery Process

RODERICK E. HUBBARD

Vernalis (R&D) Ltd, Granta Park, Abingdon, Cambridge CB1 6GB, UK and
University of York, Structural Biology Lab, York, YO10 5YW, UK

1 Introduction

The past 30 years has seen an accelerating increase in our understanding of the molecular mechanisms that underlie disease processes. This has had a fundamental impact on the process of drug discovery, and most of modern pharmaceutical research is based on target-focused discovery, where the goal is to affect the biological activity of a particular molecular target to provide a cure or treatment for a disease. As the 3D structures of some of these targets have become available, a range of experimental and computational methods have been developed to exploit that structure in drug discovery. These developments and some of their applications are the subject of this book.

In a target-focused approach, the cycle of discovery is very similar with or without a structure for the target. Initial-hit compounds are found that bind to the target and enter a medicinal chemistry cycle of making compound analogues and testing in suitable biological models. From this, the chemist builds hypotheses of what is important for the activity. Using experience (or inspired guesses) the chemist then makes changes that should improve the properties of the compound and the cycle of synthesis, testing and design begins again. These hypotheses develop a model of the conformations the compounds adopt, the chemical surfaces they project and the interactions made with the active site. For example, the optimisation of sildenafil (Viagra),¹ included consideration of the electronic properties of an initial-hit compound and how it could be improved to more closely mimic the known substrate in the active site of phosphodiesterase, many years before the structure of this enzyme was known.

Nowadays an appreciation of the 3D structure of both the compounds and their target are a part of just about every drug-discovery project. This target structure can

be experimentally determined, a model constructed on the basis of homology or a virtual model of the receptor created on the basis of the chemical structure of the known active compounds. In addition, computational methods such as virtual screening and experimental methods such as fragment screening can generate many new ideas for compound templates and possible interactions with the active site. The major advantage of experimentally determining the structure of these different compounds bound to the target is to increase the confidence in the hypotheses and increase the scope of subsequent design. This encourages the medicinal chemists to embark on novel and often challenging syntheses in the search for novel, distinctive and drug-like lead compounds. Our ability to predict conformational changes in proteins and the binding energy of protein–ligand complexes remains relatively poor, so there is still plenty of scope for experience, inspiration and guess work in the details of design.

This book will provide an overview of the methods currently used in structure-based drug discovery and give some insights into their application. Essentially, all of the examples and methods focus on proteins as the therapeutic target. There has been considerable progress in the structural biology of RNA and DNA molecules and these classes of molecules are the recognised target for some successful drugs. For DNA, our understanding of the binding of compounds that intercalate or bind to the small groove is reasonably well advanced (for an early example, see Henry;² current perspectives are provided in Tse and Boger,³ and Neidle and Thurston,⁴). There have also been spectacular advances in determining the structure of whole ribosome sub-units^{5,6} and of representative portions of the ribosomal RNA⁷ in complex with known natural product antibiotics. These structures have led to some hope that rational structure-based methods may be applied against the ribosome and also other RNA targets where a particular conformation has a role in disease processes (Knowles *et al.*, 2002). Although there has been some progress⁸ and it has been possible to discover compounds with reasonable affinity for RNA, there remain considerable difficulties in designing small, drug-like molecules with the required specificity to discriminate between the very similar sites presented on RNA. For these reasons, the discussions in this book focus on proteins as the therapeutic target.

2 The Drug Discovery Process

As discussed in the Preface, drug discovery is an expensive and time-consuming activity that mostly fails. Retrospective analyses of the pharmaceutical industry during the 1990s estimate that each new drug in the market takes an average 14 years to develop, costing in the region of \$800 million. In addition one in nine compounds that enters clinical trials makes it to the market.^{9,10} The attrition rate in discovery research is similarly high. Depending on the company, therapeutic area and discovery strategy, at best only one in ten research projects that begin with a starting compound will generate an optimised candidate to enter clinical trials. For these reasons, most companies maintain a pipeline with a large number of projects in the early stages, taking a diminishing number forward at each stage. The discovery process gets more expensive as you proceed, hence careful management of the portfolio is

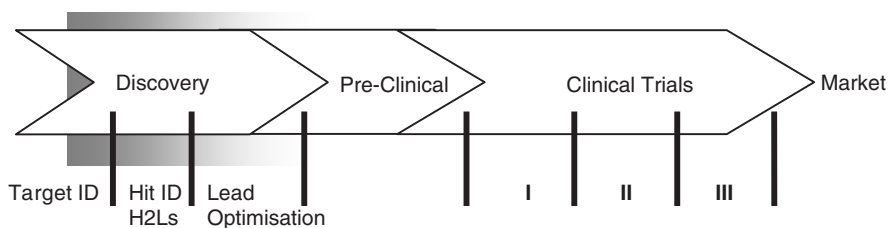


Figure 1 The drug-discovery process. The lightening, shaded box emphasises where structure-based methods can play a significant role. The horizontal axis only approximately scales to time in each stage

essential. The key is to make the right decision at the right time – knowing when to stop a project is often more important than committing to continuing.

Modern, target-oriented drug discovery is usually organised into a series of stages. The definitions of these stages differ from company to company and the details of the boundaries will vary from project to project. The following discussion provides an illustration of the stages, their purpose and duration and the types of resources involved. Clear criteria need to be established for moving from one stage to another as, in general, the stages become progressively more resource and expense intensive (Figure 1).

2.1 Establishing a Target

Clearly, the starting point for a target-oriented drug-discovery project is to identify a relevant target. In the pre-genomic era, targets were discovered through cellular and protein biochemistry methods, where a detailed understanding of the origins of a disease led to isolation and characterisation of key protein molecules. Examples presented in the applications section of this book include neuraminidase described by Colman for anti-influenza therapies and the factor Xa work described by Liebschutz and colleagues to produce anti-thrombotic agents. The nature and significance of these targets were established before much of the modern machinery of molecular biology and genomics methods were available.

The approach to biological research has undergone dramatic changes in the past decade, with successions of omics technologies becoming available. Genomics has recorded the sequence of nucleic acid bases in many genomes, and continuing bioinformatics analyses are identifying the coding regions. Comparing the genomes of both pathogen and host organism can identify potential target genes. Transcriptomics methods monitor the identity and levels of RNA transcribed for each gene, and there have been high hopes that comparison of “normal” and diseased cells will identify targets. There is a vast literature in these areas – Egner *et al.*¹¹ provide an introduction to the methods, and the recent critique by Decherig¹² points out some of the pitfalls. There has been considerable interest (and investment) in applying these methods to find new targets for different diseases and conditions. As the first genomes began to appear, there was intense interest in identifying what all the genes were. An

example of a target discovered in this way is the beta form of the estrogen receptor (see Manas *et al.* in this book).

Whatever the mechanism of identifying a target, there needs to be some level of validation before nominating it for a drug-discovery project. The phrase “target validation” is much misused – a target cannot be said to be truly validated until a drug that uniquely affects that target is on the market. Even then, there can be issues such as the recent challenges facing COX-2 as a target following adverse effects (see 24 February 2005 news item in *Nature*, **433**, 790).

In general, the requirements for a target are to establish a biological rationale for why affecting the target will have the desired therapeutic benefit. This can include assessing the viability of the organisms produced with a particular gene removed, either through knock-out technology or through RNA interference techniques. These are not ideal methods for emulating the actual effect of a drug – with gene knock-outs, there is much redundancy and subtlety in biological pathways and the removal of a gene can often be compensated in other ways as the organism differentiates and grows. An example here is the attempts to discover a function for the beta form of the estrogen receptor. Once the gene had been identified, there were intense efforts to ascribe a function to the gene, with considerable investment in producing and characterising knock-out animals.¹³ There were hints, but in the end, it took the development of isoform-specific compounds to provide chemical tools which could probe the biology and identify which diseases or conditions were associated with the receptor (again, see the chapter from Manas *et al.* in this book).

The best case for a target is to have a compound available that can provide the biological proof of concept. This is a compound that is sufficiently specific for the target of interest that can be studied either in cellular assays or in animal models of disease, to demonstrate that modulating a particular target will have the desired therapeutic benefit, *in vivo*. Such compounds could come from natural products, as in the case of antibiotics that validate the ribosome as a target⁵ and the geldanamycin derivatives that are demonstrating the potential of Hsp90 as an oncology target.¹⁴

In addition to biological validation, targets also need to be considered for what is termed, druggability. That is, does the protein have a binding site which can accommodate a drug-like compound with sufficient affinity and specificity? Although some experimental methods may be used to assess these,¹⁵ analyses of experiences with many targets have generated some general principles discussed in the chapter by Hann *et al.* later in this book. In summary, enzyme active sites tend to be highly druggable consisting of a distinct cleft designed to bind small substrates and with defined shape and directional chemistry. In contrast, most protein–protein interactions are less druggable as they cover quite large areas of protein surface with few shape or chemical features that a small molecule could bind to selectively. Unless particular “hot-spots” of activity can be identified, they are generally regarded as unsuitable drug targets (see Arkin and Wells, 2004 for a discussion).

Finally, for a structure-based project, there is a clear structural gate – that is, the structure of an appropriate form of the target needs to be available. Sometimes (for

example, in a small structure-based company) this is set as a strict gate – that is, unless the structure is available hit identification cannot begin. There can be additional constraints. For example, if the project is relying on fragment screening using crystallography followed by soaking with compound mixtures, then the protein has to crystallise in a suitable crystal form with an open binding site.

2.2 Hit Identification

A hit is a compound that binds to the target and has the desired effect. The conventional method for identifying hits is by screening a compound collection which could consist of natural products or substrate mimetics, legacy compounds in a company's collection, compounds synthesised as potential hits against a particular class of target (focused library) or commercially available compounds. The majority of large pharmaceutical companies have invested considerably in automating this initial phase of hit identification, both in the generation of suitable target libraries and in the initial assay. This High Throughput Screening (HTS) approach places considerable constraints on the robustness of the assay and the availability and properties of the available compound collection (see Davis *et al.*¹⁶ for an up-to-date discussion of the issues).

HTS is also very expensive, consuming large quantities of target and compounds and requiring significant investment in robotic screening devices. Smaller companies that rely on screening usually work with smaller libraries of compounds, and depend on a particular “edge” over the larger companies. That distinctiveness could be either in some detailed knowledge or expertise with the biology of the target class and thus more appropriate configuring of the assay, or through a small library of compounds for that particular class of target. It is in the hit-identification phase that structure-based methods have provided smaller companies an opportunity to establish rapidly effective drug-discovery projects, particularly through the use of virtual screening or fragment-based methods (see later).

In most cases, the hit-identification phase relies on configuring a particular assay to monitor binding or inhibition. Usually, a large number of compounds are being screened, so the first experiment is to measure compounds that exhibit activity (above a certain percentage inhibition) at a set concentration. This is usually followed by confirming the hits, that is where an *in vitro* assay is run at varying concentrations to determine the IC_{50} ^{*} or the K_i or K_d [†] for the compound and the quality of the compound sample checked. Maintaining quality in a compound collection is a major challenge – compounds decompose over time, particularly if held dilute in solution in air. In addition, it is not unusual for 5–10% of compounds purchased

^{*}The IC_{50} represents the concentration of the drug that is required to achieve 50% reduction in activity of the target, usually *in vitro*. A related term is EC_{50} , which represents the plasma concentration required for obtaining 50% of the maximum effect *in vivo*.

[†] K_i is the inhibition constant for a reaction. The precise definition of these constants will depend on the chemical nature of the assay. When comparing values, it is important to know the precise details of the assay – variations in pH, buffer composition, ionic strength, temperature, protein activation state, competitor ligands, *etc.*, can all have a real effect.

from commercial suppliers to either be not what they claim to be, or to contain major contaminants that can give false positive (or false negative) results.

An HTS campaign can require significant resources (compound, target, manpower) and last 6–12 months, depending on how long it takes to configure a robust assay. Where smaller collections of compounds are being used, or structure-based methods applied, the hit-identification phase usually lasts around 6 months and requires a relatively small team of scientists.

The output from a hit-identification campaign is a set of compounds whose chemical structures have been checked and which have reproducibly been shown to have activity.

2.3 Hits to Leads

The hits to leads (H2L) phase is where some of the crucial decisions are made in a project – establishing which chemical series has the potential to be optimised into a drug candidate. This is an important decision as lead optimisation (the next phase) is when significant resources and effort are spent in optimising the properties of compounds. For these reasons, most companies set quite stringent criteria for entering lead optimisation, set for each target and reflecting the projected requirement of the properties of the final drug candidate, often called target-product profile.

The detailed work during the H2L phase varies with the nature of the project and, in particular, the origin of the hit compounds. Wherever the compounds come from, it is usual to re-synthesise the compounds for complete validation of the hit and to either purchase or synthesise close analogues of the compounds. In general, it is during the H2L phase that dramatic changes in chemical template are made and the essential core of the lead series established. The usual aims are to establish preliminary structure–activity relationships (SAR) within one or more series, to explore the indicative physicochemical and ADMET[‡] properties of the compounds, to consider the chemical tractability or synthetic accessibility of the compounds and to understand the IP position on the compound series and target. Depending on the project (and the company policy), entry into lead optimisation can be gated by demonstrating some *in vivo* activity in the series. Setting the right barriers for entry into lead optimisation is one of the most challenging aspects of medicinal chemistry.

[‡]There are a number of phrases and acronyms for these important drug-like properties. *DMPK* is drug metabolism and pharmacokinetics (*PK*). *PK* is the characterisation of what the body does to a drug. Conventionally, this is analysed in terms of four main processes – Absorption, Distribution, Metabolism and Excretion or *ADME*. This is sometimes extended to include Toxicity (*ADMET*). All of these processes are due to complex, interdependent factors within the body and although detailed mechanistic (and increasing structural) information is emerging about individual components, empirically derived models are the only route to prediction. These models usually relate physico-chemical properties of the molecules to activity. The most commonly used measures are lipophilicity, measured as logP – the partition between octanol and water. This can be calculated (clogP) and there are other variants such as logD to reflect the distribution observed when the compound exists in various ionization states. The main challenge for these models is the quantity and consistency of experimental data and the transferability of such models from one compound series to another. As many of the processes are due to interaction with and activities of many different proteins, it is often the case that models are constructed within a compound series, but will not transfer. Although some use is made of these predictive models, in most cases, experimental measurements need to be done. Most can be configured as *in vitro* assays.

This phase usually takes around 6 months, depending on the requirements for biological testing and the degree of synthesis required to establish a lead series with appropriate properties.

2.4 Lead Optimisation

This is the most resource-intensive component in drug discovery, requiring considerable input from synthetic chemistry, modelling, disease biology and assay design. It is not unusual for a lead optimisation (LO) team to consist of over 15 scientists, particularly if more than one lead series of compounds are being progressed. The main challenge is to develop one or more compounds with the desired drug-like properties. As well as having sufficient affinity for the target (nM[§] is the usual goal), the compound needs to have an appropriate selectivity profile, be able to get to the site of action (which for many targets means cell permeability) and have acceptable drug-like properties. In addition, it is important to continue to track that the observed effect on cellular (and later *in vivo* activity) is from interaction with the identified protein target. Although, in the end, the most important feature is that the compound works in the cell, pharmacodynamic markers are important to check if the compound is affecting the biology through the predicted target,[¶] particularly when an understanding of the structure of that target is being used to guide optimisation.

The early stages of the LO process are usually focused on achieving the desired affinity and selectivity. Selectivity requirements vary from target to target and, in particular, between different therapeutic areas. Where a drug is for an acute condition such as cancer, where rapid intervention is required and the course of treatment is likely to be short term, then side-effects can be tolerated. In fact it appears that some oncology drugs achieve efficacy by targeting a number of pathways. Where the drug is for a chronic condition, such as arthritis or diabetes, where the drug will be taken for many years, the selectivity requirements can be much more stringent.

In these early stages, there can still be some modest changes in the central core of the compound. However, as LO progresses, the main changes are on the periphery of the molecule. The main driver is the biology – it is remarkable how quite small changes in the chemistry can have a large effect on the biological activity, particularly *in vivo*.

Lead optimisation typically takes 18–30 months, depending on the complexity of the target biology, the resources deployed and the chemistry of the lead series. The

[§]The phrase, a nanomolar inhibitor, is frequently used in the literature. Usually, this refers to the dissociation constant (K_d) for the *in vitro* equilibrium between target–ligand complex and free target and unbound ligand. Usually (but not always), a higher affinity of a compound for a particular target will increase its selectivity over other proteins in the system.

[¶]Pharmacodynamics (PD) is what the drug does to the body. In many drug discovery programmes, a key part of the early stages of the project is to establish pharmacodynamic markers that can be used to make the link between binding of compound to the target and the effect seen on the cell – *i.e.* being sure that the activity is from interaction with that particular target. As lead optimisation progresses, it is the cellular (and eventually the *in vivo*) activity that guides the medicinal chemistry, so it is essential to ensure that the activity being measured is due to the compound binding to the target that is being used to inform the design.

real challenge in lead optimisation is balancing when certain properties need to be introduced and deciding when to abandon a particular project or lead series.

The output from the LO is a compound (or a set of compounds) that meets the required criteria of *in vivo* efficacy in animal models, with a demonstrable mode of action and with acceptable PK.

2.5 Pre-Clinical Trials

This phase is to prepare for the testing of the compounds in man. This includes scale-up synthesis, formulation, toxicology and design of clinical trials.

The difficulty and cost of synthesising the compounds is considered throughout the discovery process, but becomes particularly important at this stage. A synthetic scheme that works in the laboratory to produce 100 mg of compound may need dramatic modification to produce the many kilograms of compound required for late stage clinical trials. Overall, the difficulty of synthesis or purification of compound will have a marked impact on the cost of goods – *i.e.* how much it will cost to produce the drug – and this can seriously impact the commercial viability of the project. Similarly, formulation – getting the drug into a form that can be administered both for the animal testing and for clinical trials – can have an impact on the project viability.

This phase is to prepare the way for clinical trials where the drug candidate is given to humans. This is covered by a stringent regulatory regime and many of the steps in the pre-clinical stage are covered by regulations and a need to work to certain legal guidelines.

2.6 Clinical Trials

This is usually the most expensive and time consuming of the overall process of discovering a new medicine. It is conventional to think of three separate stages.

Phase 1 studies are primarily concerned with assessing the drug candidate's safety. A small number of healthy volunteers are given the compound to test what happens to the drug in the human body – how it is absorbed, metabolised and excreted. A phase 1 study will investigate side-effects that occur as dosage levels are increased. This initial phase of testing typically takes several months. About 70% of drug candidates pass this initial phase of testing.

In *phase 2*, the drug candidate is tested for efficacy. Usually, this is explored in a randomised trial where the compound or a placebo are given to up to several hundred patients with the condition or disease to be treated. Depending on the condition, the trial can last from several months to a number of years. The output is an increased understanding of the safety of the compound and clear information about effectiveness. Only about one-third of the projects successfully complete both phase 1 and 2 studies, but at the end of this process, the compound can be truly considered as a drug.

In a *phase 3* study, a drug is tested in several hundred to several thousand patients. This provides a more thorough understanding of the drug's effectiveness, benefits and the range of possible adverse reactions. These trials typically last several years

and can include comparison with existing treatments on the market to show increased benefit. These trials provide the necessary data on which to get approval by the regulatory authorities.

As the drug comes towards, and is launched in the market, continued trials and monitoring is required. Sometimes, adverse reactions can only be picked up when a drug is given to a very large population. Problems can sometimes be dealt with by changes in prescribing practice or through defining particular patient populations. However, it is sometimes necessary to remove a drug from the market (*cf.* earlier reference to COX-2 inhibitors).

2.7 Maintaining the Pipeline

As discussed above, the failure rate (or attrition, as it is sometimes termed) in the clinical stages is well documented.⁹ During the 1990s, around one in ten compounds that entered clinical trials was successfully launched as a drug. This drop-out rate can be due to failures in either the target or the compound. There have been significant efforts to reduce problems due to unfavourable bioavailability or ADMET properties. Although our improved understanding of the molecular mechanisms underlying some aspects of toxicology (such as interaction with the hERG channel)¹⁷ allows such features to be screened earlier, there will still be failures due to adverse side-effects when given to man. In addition, it is often not until a suitably selective drug is available to give to man that the hypothesis can be tested that modulating the activity of a particular target will have a therapeutic benefit.

The attrition rates in the early stages of drug discovery are more difficult to quantify as the raw data is not in the public domain. Also, the boundaries between each step vary dramatically between targets, between disease indications and between the varying drug-discovery paradigms of different companies. The definition of success also depends on how high the criteria are set for progression. For example, the problems experienced in clinical trials in the 1990s has led to much more stringent sets of assays and thus higher rates of failure in the research and pre-clinical phase. As a general rule of thumb, the attrition rates in discovery are about the same as in clinical trials – about one in ten. This means that a pharmaceutical enterprise needs to maintain an essentially funnel-shaped pipeline to generate a sustainable business, with larger numbers of projects at the earlier stages. For this to be successful requires some difficult but clear decisions to be made on whether and how to progress the targets from one stage to the next.

3 What is Structure-Based Drug Discovery?

3.1 From Hype to Application

Drug discovery has inspired, suffered and eventually benefited from many waves of new technologies. The drivers are very clear – there is an increasing need and expectation for new medicines and treatments and a patient population that is increasing in both numbers and in affluence. Not surprisingly, this has led to substantial growth in the pharmaceutical industry, which combined with the continuing consolidation

of the sector has provided the financial and scientific resources for huge investments in new technologies and methods. At the same time, there have been waves of new companies established, primarily with venture-capital funding, to develop new methods and either deliver them to the large pharmaceutical companies or to exploit themselves in drug-discovery research or services. As with all new technologies, there has been considerable hype, enthusiasm and ambition for the methods and what they can deliver. Realistically, this is probably needed to ensure sufficient resources are available to develop and assess the methods.

The examples include genome sequencing, transcriptomics and proteomics for target identification and validation, protein engineering for biological therapeutics, combinatorial chemistry, molecular modelling as well as structure-based methods. There have been considerable investments in some of the technologies. For example, combinatorial chemistry was a revolutionary technology for synthesising massive numbers of related compounds. The first paper describing synthesis of a single combinatorial library appeared in 1992¹⁸ and the most recent comprehensive survey of combinatorial library synthesis for 2003 showed 468 new methods.¹⁹ The early years of combinatorial chemistry led to massive investment in parallel synthesis and screening methods in the pharmaceutical industry. Very few compounds from this early investment have entered clinical trials as the early methods were flawed. There was insufficient appreciation that the available synthetic methods suitable for such parallel operation would sample only a relatively small chemical space and produce many compounds without the required drug-like properties. In addition, there were many issues in developing robust, reliable synthesis of individual compounds. However, many lessons were learnt and the design of focused libraries, where particular features of templates are elaborated, are now an integral part of most drug-discovery programmes.

There has been some hype associated with the availability and value of structures of therapeutic targets and the ability to use structure and modelling methods to design compounds. At times, some elements in the pharmaceutical industry and, in particular, some start-up companies have been over-optimistic on what the methods can deliver. However, there has been a steady realisation of the power of the methods for the classes of target for which structures can be determined. The evidence for this is that essentially all pharmaceutical companies have some form of modelling group that constructs models of the structure of targets and uses these in discovery and design of new compounds. And an increasing number of small companies have invested in the ability to determine the structure, particularly with X-ray crystallography.

There are three main contributions that structural methods are making to the drug-discovery process – structural biology, structure-based design and structure-based discovery.

3.2 Structural Biology

The determination of the structure of a protein target, perhaps complexed to partner proteins, lipids, nucleic acid or substrate, can provide a clear insight into the mechanism of action of a protein, which in turn can often be related to its biological or therapeutic role.

Modern structural biology, particularly protein crystallography, is generating the structure for an increasing number of therapeutically important targets (see the chapter by Brown and Flocco). The two main issues limiting the number of structures are the ability to produce sufficient quantities of pure, soluble, functional, homogenous protein for crystallisation trials and the ability of the protein to form regular crystals suitable for diffraction experiments. This combination of limitations often means that a structure is not available for the whole therapeutic target. However, even the structure of individual domains can be sufficient to make a real impact on a discovery project, and provide a context within which to understand the overall function of the protein. The estrogen receptor (see Manas *et al.*'s chapter) provides one example. Although the receptor consists of a number of domains, the structure of just the ligand-binding domain is sufficient against which detailed structure-based design can successfully design selective ligands. However, the subtleties of the function of the receptor in the cell can only be understood in terms of the interplay between the different domains that have an influence on receptor activity.

Another example of where drug discovery against just one domain can be successful is the molecular chaperone, Hsp90. This protein is up-regulated in cells under stress and, in complex with a varying repertoire of co-chaperone proteins, helps to stabilise the folding of a large number of proteins important for cell proliferation, growth and function, such as the estrogen receptor and key cell-signalling kinases. The real breakthrough in identifying this target came with the discovery that Hsp90 is the primary target for natural products such as geldanamycin and radicicol, the derivatives for which a viable therapeutic window has been identified, such that compounds such as 17-AAG are now entering phase 2 clinical trials.¹⁴ Hsp90 contains three domains – a C-terminal domain of unknown function that is thought to be important for the formation of the functional dimer, a central domain with large hydrophobic surfaces that can stabilise nascent, unfolded peptides and an N-terminal domain that harbours the ATP binding site. ATP hydrolysis provides the energy driver for the chaperone function. The natural products, geldanamycin and radicicol, bind to the ATP-binding site on the N-terminal domain, blocking hydrolysis and thereby inhibiting the chaperone action. A number of projects are now embarking on discovery and optimisation of compounds that can selectively inhibit this ATP site.²⁰ However, the detailed mechanism of action has to take into account interactions between the different domains and also the effect of other co-chaperones.²¹

3.3 Structure-Based Design

The crystal structure of a ligand bound to a protein provides a detailed insight into the interactions made between the protein and the ligand. Such understanding can be used to design changes to the ligand to introduce new interactions to modify the affinity and specificity of the ligand for a particular protein. In addition, the structure can be used to identify where the ligand can be changed to modulate the physicochemical and ADME properties of the compound, by showing which parts of the compound are important for affinity and which parts can be altered without affecting binding. There

are numerous examples²² where simple inspection of the protein–ligand complex has identified where solubilising groups can be added. The chapter by Manas *et al.* in this book provides an excellent example of where detailed calculations and design can successfully design changes to affect selectivity between isoforms.

This type of analysis is now well established and has been used in many drug-discovery projects over the past 15 years. Some of the early disappointments in structure-based design arose because of the difficulty of predicting binding affinities between protein and ligand. Although the predictive power of the calculations is beginning to improve,²³ there remain serious challenges in predicting binding affinities. It should be remembered that the equilibrium between target and ligand is governed by the free energy of the complex compared to the free energy of the individual target and ligand. This includes not only the interactions between target and ligand, but also the solvation and entropy of the three different species and the energy of the conformation of the free species. Overall, the equilibrium is a balance between all these different terms and a number of detailed experimental studies have demonstrated that energetically unfavourable changes in the protein, such as conformational strain or disruption of stabilising interactions, can be compensated for by interactions the protein is then able to make with the ligand.^{24,25} These balances are even more difficult to consider in the cellular context, with the many complicating factors of competing ligands, solvent conditions and partner proteins.

3.4 Structure-Based Discovery

As the availability of crystal structures increased in the early 1990s, a number of experimental and computational methods were developed to use the structure of the protein target as a route to discover novel hit compounds. The methods include *de novo* design, virtual screening and fragment-based discovery. These developments are covered in more detail in the later chapters of this book, but their main features can be summarised as follows.

Virtual screening use computational docking methods to assess which of the large database of compounds will fit into the unliganded structure of the target protein. Current protocols and methods can, with up to 80% success, predict the binding position and orientation of ligands that are known to bind to a protein. However, identifying which ligands bind into a particular binding site is much less successful, with many more false positive hits being identified. The major challenges remain the quality of the scoring functions – if these were more accurate, then the challenge of predicting conformational change in the protein on binding of ligand would also be more tractable.

De novo design attempts to use the unliganded structure of the protein to generate novel chemical structures that can bind. There are varying algorithms, most of which depend on identifying initial hot spots of interactions that are then grown into complete ligands. As well as the ubiquitous issue of scoring functions, the major challenge facing these methods is generating chemical structures that are synthetically accessible.

Fragment-based discovery is based on the premise that most ligands that bind strongly to a protein active site can be considered as a number of smaller fragments

or functionalities. Fragments are identified by screening a relatively small library of molecules (400–20,000) by X-ray crystallography, NMR spectroscopy or functional assay. The structures of the fragments binding to the protein can be used to design new ligands by adding functionality to the fragment, by merging together or linking various fragments or by grafting features of the fragments onto existing ligands. The main issues are designing libraries of sufficient diversity and the synthetic challenges of fragment evolution.

The above discussion raises a rather semantic question about the use of the words design and discovery. The word design implies some element of prediction – and some of the methods currently used (such as fragment screening, for example) is clearly not design. In addition, although it is sometimes possible to design modifications to a compound to improve its affinity or selectivity for a target, it is rarely possible to be so predictive in introducing drug-like properties into a molecule. The best you can usually rely on is that the structure of a compound bound to its target will show where a compound should be elaborated (perhaps with a focused library) from which a compound with the desired drug like properties (say, cellular penetration or the desired pharmacokinetics) will be found by assay of the resulting library. For these reasons, this book will use the phrase structure-based drug discovery throughout.

4 The Evolution of the Ideas of Structure-Based Drug Discovery

It is fascinating to look back over the literature of the past 40 years and chart the emergence of the methods and ideas of structure-based drug discovery. The following is a necessarily subjective, idiosyncratic and personal perspective on the key papers and developments, with apologies for any key papers or work that has been overlooked.

The description is chronological and divided into decades. As a starting point for each decade, there is a qualitative summary of the papers in the June issue of the *Journal of Medicinal Chemistry (J. Med. Chem.)* in 1965, 1975, 1985, 1995 and 2005. This is necessarily a snapshot, but it does give some insight into how far structural methods had affected the papers and thinking of drug-discovery scientists at the time.

4.1 1960s

Not surprisingly, the papers of the June 1965 issue of *J. Med. Chem.* make no mention of structure. However, this decade saw the first use of two of the central methods of modern structure-based discovery – the determination of protein structure by X-ray diffraction and the development of molecular graphics.

The first structures (myoglobin,²⁶ haemoglobin,²⁷ and lysozyme²⁸) laid the foundation of modern protein crystallography. These established that through structure it was possible to understand the mechanism of action of the proteins and relate this to their biological function. The work on haemoglobin extended to the first attempts to

provide a structural understanding of genetic disease and Perutz and Lehmann²⁹ mapped the known clinically relevant mutations in haemoglobin to the structure.

The first major developments in molecular graphics came in the mid-1960s when Project MAC at MIT produced the first Multiple Access Computer, a prototype for the development of modern computing. The computer included a high performance oscilloscope on which programs could draw vectors very rapidly, and a closely coupled “trackball” through which the user could interact with the representation on the screen. Using this equipment, Levinthal and his team developed the first molecular graphics system and his article in *Scientific American*³⁰ remains a classic in the field. In this paper, he described their achievements, and laid the foundations for many of the features that characterise modern-day molecular graphics systems. It was possible to produce a vector representation of the bonds in a molecule and to rotate it in real time. The representation could be of the whole molecule, or a reduced representation such as an alpha carbon backbone. Because the computer held the atomic coordinates of the molecule, it was possible to interrogate the structure, and to use a computational model to perform crude energy calculations on the molecule and its interaction with other molecules. This work inspired various groups to begin building molecular modelling systems.³¹ Also during this time, scientists such as Hansch laid the foundations for modern predictive cheminformatics methods by establishing that some of the molecular properties of compounds could be computed by considering the individual fragments that make up the molecule (for a fascinating review of the development of ideas on partition coefficients see Leo *et al.*³²).

4.2 1970s

The June 1975 issue of *J. Med. Chem.* includes a few papers that discuss the ideas of common features on small molecules that are indicative of activity.³³ However, these analyses remain focused on the small molecule (little discussion of the protein target) and most of the papers describe very traditional synthesis and testing approaches.

There was a steady increase in the number of available protein structures during the 1970s. The crystallographer was limited to working on naturally abundant proteins and data collection (in general) used rather slow X-ray diffractometers. There were sufficient structures, however, for a data bank to be required and the Protein Data Bank was established in the late 1970s.³⁴ The depository was run for many years at Brookhaven National Labs and moved to the Research Collaboratory in Structural Biology during the 1990s (<http://www.rcsb.org>).³⁵

There are three examples of the use of structure to consider ligand or drug binding that should be highlighted. The first is the studies on dihydrofolate reductase (DHFR) summarised in Matthews *et al.*³⁶ This is a fascinating paper to read. Although the description of the determination of the structure emphasises just how much the experimental methods of protein crystallography have developed, it does illustrate that many of the ideas of modern structure-based design were well established some 30 years ago. The structure of methotrexate bound to bacterial DHFR allowed quite detailed rationalisation of the differences in binding affinity of related ligands and an

understanding of why, although there are sequence variations, the ligand binds tightly to all DHFRs known at that time. This type of structural insight led to structure-based design of new inhibitors.³⁷

The second example is the work of the Wellcome group who explored various aspects of ligand binding to haemoglobin through modelling of the interactions of the ligands with the known structure.^{38,39} The ideas about molecular interactions generated in this work laid the foundation for Goodford's later development of the GRID program (see the 1980s).

The third example is the design of captopril,⁴⁰ an inhibitor of the angiotensin-converting enzyme (ACE) and a major drug for hypertension. Although sometimes quoted as one of the first examples of structure-based design, the structure of ACE was not known in the mid-1970s. However, the design was strongly directed by constructing a crude model of the active site, based on the known structure of carboxypeptidase A.

These papers demonstrate that the central paradigm in structure-based design was well established during the 1970s. This paradigm is that the structure of a ligand bound to its target protein can be used to understand the physicochemical interactions underlying molecular recognition and binding affinity and this insight can then be used to design changes to the ligand to improve its properties.

Alongside the slow emergence of design based on the structure of the target, there were important developments in ligand-based modelling. Computational methods incorporating molecular and quantum mechanical treatments of ligand conformation and properties were being explored. This included conformational analysis to predict the 3D conformations of small molecules and the calculation of molecular properties such as hydrophobicity and electrostatic potential. Brute force methods of quantitative structure activity relationships (QSAR) were developed that considered large sets of active and inactive compounds, computed many properties and then attempted to construct a predictive correlation between some algebraic combination of computed properties and activity. Alongside this, the ideas of "virtual" receptor-based modelling emerged, where the properties of active compounds were analysed to construct a 3D pharmacophore of the features required for activity. Exploring and then applying this range of methods required the development of suites of molecular modelling methods. However, only a few, large laboratories had dedicated computing facilities and these provided the focus for the development of a number of software systems that laid the foundation for modern modelling systems.

It is possible to chart the development of the ideas and methods of molecular graphics and modelling systems in two distinct communities – protein crystallography and molecular modelling in support of ligand design. The first developments in protein crystallography were by Alwyn Jones who developed the program FRODO^{41,42} (re-formulated and extended in the program O⁴³). Protein crystallographers required powerful molecular graphics facilities to help in determining protein structures for visualisation of large electron density maps and fitting of a molecular model of the protein structure into the density. Once the structure had been determined, graphics was again vital in allowing interactive analysis of the structure to not only describe the folding of the protein, but also to understand the mechanism and thus function of the protein. Important examples were the development of the

earliest space-filling representations of molecular structure by Feldman at the NIH⁴⁴ and the developments of the Langridge group at UCSF.⁴⁵

Most of these early developments were in the academic community, but there was also considerable interest in the potential of molecular modelling methods in the pharmaceutical industry and many of the large companies spawned their own software development efforts. The reviews by Gund *et al.*⁴⁶ and Marshall⁴⁷ provide an appreciation of the early developments. The success of these encouraged the development of a whole new industry in the 1980s.

4.3 1980s

Despite all these advances, the June 1985 edition of *J. Med. Chem.* is still very similar in flavour to that of 10 years previously. Most of the papers are ligand oriented, with little evidence that structural models of the target were being used to rationalise and drive synthetic efforts.

However, the 1980s saw many important developments in the scientific disciplines that underpin structure-based drug discovery. Molecular biology and protein chemistry methods were beginning to unravel the biology of many disease processes, identifying new targets and importantly, providing the over-expression methods with which to produce large quantities for structural study. In protein crystallography, synchrotron radiation not only speeded up the data collection process but because of its intensity and focus allowed usable data to be collected from smaller, poorer crystals. This was complemented by developments in methods for refining structures, initially least squares refinement⁴⁸ and later in the 1980s, the simulated annealing approach of X-plor.^{49,50}

There were also important developments in techniques in NMR spectroscopy. Isotopic labelling of protein, instrument and method advances led to multi-dimensional NMR techniques for solving small, soluble protein structures⁵¹ (see the chapter by Davis and Hubbard in this book). The larger pharmaceutical companies invested in these methods alongside the traditional use of NMR in analytical chemistry. However, the size limitations of the technique meant there were few therapeutic targets accessible to NMR.

This decade also provided the core of the methods in computational chemistry that support analysis of protein–ligand complexes. Molecular mechanics techniques such as CHARMm⁵² gained wider application and the computational resources available to most groups increased steadily to allow routine use of energy minimisation and molecular dynamics methods. Of particular note are three papers specifically dealing with protein ligand interactions. Jencks⁵³ provided a simple but powerful analysis of the contributions made by different parts of a molecule to binding. His analysis established that the first part of a molecule overcomes many of the entropic barriers to binding, giving higher affinity for subsequent additions of functionality. This firmly established the ideas that led to fragment-based discovery in the early 2000s. In a similar vein, Andrews *et al.*⁵⁴ analysed the contributions that different functional groups make to binding. Finally, Goodford developed the GRID approach⁵⁵ that used an empirical energy function to generate a very visual analysis of where different types of functional group could interact with a binding site. This approach had a significant

impact on how chemists and molecular modellers viewed protein active sites and the possibility for rational design. An important factor in their application was in the availability of affordable computing. At the beginning of the 1980s, the necessary computing and graphics hardware to support structure analysis and molecular modelling cost many hundreds of thousands of dollars. By the end of the decade, graphics workstations such as the Silicon Graphics IRIS, meant essentially every scientist had access to the technology and software.

A development that had a major impact on the way scientists thought about protein structure was the Connolly surface. The molecular surface is a fundamental aspect of a structure as it is through the complementarity of shape and chemistry of the surface that molecules interact with each other. A variety of different representations of surfaces were developed, the most enduring and informative of which is that developed by Connolly.^{45,56} The molecular surface is defined by the surface in contact with a probe sphere as the sphere “rolls” over the surface of the molecule. Alternatively, the extended solvent accessible surface can be calculated in which the surface is traced out by the centre of the probe sphere as it rolls over the molecule. Although the initial graphics devices could only show this as a continuous envelope of dots, it produced a smooth surface that showed where the protein met the solvent. This approach underlies essentially all the surface representations in use today. In addition, there were developments in the treatment of protein electrostatics, and the program GRASP provided a very visual presentation of the electrostatic surfaces of proteins computed using a Poisson–Boltzman treatment.⁵⁷ These surface images simplified the representation of protein chemistry and provided important insights into function.

A number of structure-based design groups began to emerge in the pharmaceutical companies. One example is the group at Merck. The paper by Boger *et al.*, 1983, describes their work on the design of renin inhibitors which summarises many of the aspects of the discipline at the time. They used homology modelling of the protein structure, and manual docking and inspection of ligands to design peptide mimetics that would find application in many protease inhibitor projects in later years. A second example is also from Merck, where structures of carbonic anhydrase were used to successfully design more potent inhibitors that are now established as treatments for glaucoma.⁵⁸ This work has been cited as one of the earliest examples of structure-based design that has resulted in a drug on the market.

Towards the end of the decade, various scientists within larger companies recognised the power of the structure-based rational approach and established new start-up companies such as Vertex and Agouron, where the resources and organisation could be geared to structure-based discovery.

4.4 1990s

All the advances in the underlying technologies meant that by the beginning of the 1990s, most large pharmaceutical companies had established structural groups and the results of their early work was beginning to be published. The papers in *J. Med. Chem.* reflect the changes. The striking difference between the June 1990 issue and that of 5 years earlier is that most of the papers in 1990 are more target oriented, with

clear discussions of molecular targets. Of the 45 or so articles, two report on protein–ligand structures and four have explicit discussions of common conformations required for receptor binding. Five years later, the June 23rd 1995 issue has a higher proportion of structural papers. Of the 25 or so articles, 4 contained protein–ligand structures and three used the concepts of pharmacophores and receptor binding. Although most of the reports of protein–ligand structures were post hoc and rationalised the results (rather than guided the design), the increase reflects the growing availability of structural methods.

In addition to the continuing increase in the number of targets for which structures were available, the major change during the 1990s was that much of the equipment for X-ray structure determination and the computing and graphics equipment required for molecular modelling was available in most well-found laboratories in both academia and industry.

At the beginning of the 1990s, there was intense interest in *de novo* design – using the structure of a protein for *ab initio* generation of new ligands. The binding site of the protein was mapped with methods such as GRID⁵⁵ or MCSS⁵⁹ and then a variety of building methods proposed for generating new ligands, such as HOOK.⁶⁰ There were two important developments for computational methods at this time. The first was the work by Bohm to analyse the growing body of experimental structures to develop the LUDI empirical scoring function for prediction of protein–ligand affinity. The second was the development of virtual screening or molecular docking methods. The pioneer in this area was Kuntz⁶¹ and a series of other programs, such as GOLD⁶² and FLEXX⁶³ emerged (for review of virtual screening see Barril *et al.*⁶⁴ and the chapter by Barril in this book).

For X-ray crystallography the major developments were in the speed of structure determination. Synchrotron radiation, coupled to new, faster instrumentation was capable of rapid data collection. A particularly significant development was cryocrystallography,⁶⁵ where flash freezing and maintaining crystals under a stream of dry air at liquid nitrogen temperatures massively reduced the problems of crystal damage. Alongside this, there were continued improvements in methods for structure refinement⁶⁶ and in semi-automated methods for fitting models of structure to the resulting electron density.^{67,68}

The important development in the NMR field was the work of the Abbott group led by Fesik, who developed the SAR by NMR approach⁶⁹ and applied it quite dramatically to develop potent, novel leads against a number of targets.⁷⁰ This approach is described in more detail in the chapter by Davis and Hubbard and exploits the ability of NMR to report selectively on binding events to identify sets of small ligands that bind to the protein and that when linked together produce high affinity ligands. This approach resuscitated interest in protein NMR spectroscopy in drug discovery, but most companies found that there were few targets with appropriate multi-pocket sites and that there were too many challenges in designing appropriate chemistry to link fragments together and maintain binding affinity.

Alongside all this methodology development, there were two high-profile drug-discovery projects that validated the structure-based approach and led to increased investment in the area. The first was work by the groups of von Itztein and Colman who used the structure of the enzyme sialidase to design potent inhibitors against the

influenza virus that became the drug, Relenza⁷¹ (see the chapter by Colman in this book). This is a classic of structure-based drug discovery – the structure of a weak substrate mimic bound to the protein was used to guide lead optimisation to produce a compound with improved affinity and selectivity that also may minimise appearance of drug resistance. The second was the many efforts in developing generations of HIV protease inhibitors. The first generation of drugs²² included the use of structures of protein–ligand complexes to identify where changes could be made on the ligand to improve bioavailability. A paper by Greer *et al.*⁷² summarises how hits were identified by screening of existing aspartyl protease libraries and the structure of these compounds bound to the enzyme was used to guide combining of features of different compounds, adding solubilising groups and making changes to affect PK properties. More recent developments have made wider use of structure-based methods, such as Salituro *et al.*, 1998. Developments in this class of inhibitors are summarised in Randolph and DeGoey⁷³ and Chrusciel and Strohbach.⁷⁴

There are two other major developments of the 1990s that should be summarised – the development of fragment screening methods and the evolution of the ideas of drug and lead-likeness.

The ideas underlying fragment-based discovery can be traced back over many decades. As mentioned above, work by Andrews⁵⁴ and by Jencks (1981) established the idea that the binding affinity of a compound arises from contributions made by different parts of the molecule. This led to the idea of mapping the binding surface of a receptor either computationally (Bohm, 1994) or experimentally.⁵⁹ The NMR methods have been mentioned above, but crystallographers also saw the potential. Work by Ringe⁷⁵ and others⁷⁶ characterised how different solvent fragments bound to protein active sites. Nienaber *et al.*⁷⁷ took the approach a step further, soaking crystals with mixtures of small molecular fragments as a starting point for drug design. These ideas have been taken forward by many other groups to provide a basis for structure-based discovery^{78,79} described in more detail in the chapter by Hann *et al.* in this book.

Analysis of the successes and failures of drug discovery in the 1990s has led to some important concepts for modern and future rational drug discovery. The analysis by Lipinski *et al.*⁸⁰ has had a profound effect on rational approaches to drug discovery by identifying some relatively simple guidelines on the properties of compounds that are orally bioavailable. This idea has been further refined⁸¹ and extended to identify the properties needed for lead compounds to be successfully optimised into leads – lead-likeness and ligand complexity.^{82,83}

4.5 2000s

The “methods” chapters in this book provide a detailed survey of the current techniques and strategies in structure-based drug discovery. The June 30th 2005 issue of *J. Med. Chem.* reflects how widespread these ideas have become. Of the 33 articles, 10 either contain a crystal structure of a protein–ligand complex or use structures to dock compounds or guide compound optimisation. In addition, 6 articles use concepts of target structure for guiding design through use of pharmacophore descriptions or similar approaches. A quick survey of other issues in 2005 suggests this is not un-representative.

Over the past five years, the increased ubiquity of structure-based methods has been built on the ideas discussed above and the increased evidence of how structural insights can not only speed up, but improve the success of drug discovery efforts. Articles such as Williams *et al.* (2005)¹¹⁵ provide a summary of the current range of therapeutic targets and discovery projects.

As well as continuing refinements and improvements in these methods, the principle advance in the past five years has been the availability of an increasing number of structures of therapeutic targets. The major barrier to structure-based discovery is the production of a pure, homogenous preparation of soluble, functional protein that forms crystals suitable for X-ray structure determination. A recent review by Hunt (Hunt, 2005)¹¹⁶ summarises the current best practise in producing protein for the range of post-genomic experimental studies. For structural work, the primary workhorses are over-expression in bacterial systems (see review by Goulding and Perry, 2003)¹¹⁷ or, where folding or post-translational modification can be a challenge, the use of insect cells driven by baculovirus infection (see review by Kost *et al.*, 2005).¹¹⁸ Although there remain considerable challenges, the massive investments in structural genomics are slowly providing improved methods and protocols for generating protein structures for an increased number of proteins (Yokoyama, 2003,¹¹⁹ Braun and LaBaer, 2003¹²⁰) aided considerably by the development of a variety of affinity tags (Waugh, 2005)¹²¹ to facilitate isolation and purification of expressed protein. A recent volume of the review journal, *Progress in Biophysics and Molecular Biology* (vol 88) contains a series of articles describing current methods for protein crystallisation.

The complete genome sequences are available for human and for many major pathogens, and many new targets are being identified and validated. A potentially valuable development for drug discovery is the recently established Structural Genomics Initiative, which aims to generate structures for many hundreds of therapeutically relevant human proteins and place them in the public domain (see <http://www.sgc.utoronto.ca>).

Where there is not structure available, there has been considerable interest in using homology models to provide a starting point for structure-based discovery. The review by Hillisch *et al.* (2004)⁸⁴ summarises the current state of the field.

5 What isn't in this Book

This book aims to provide an introduction and overview to the methods of structure-based drug discovery. The methods chapters provide a reasonably comprehensive coverage of the ideas and tools available. However, the chapters illustrating how these methods have been successfully applied cover only a few of the applications areas. The following provides a brief summary of the major areas of omission.

5.1 Drug Discovery Against GPCR Targets

The G-Protein-Coupled Receptors (GPCRs) represent a major class of target for therapeutic intervention.⁸⁵ Over 50% of the current marketed drugs target this class of receptor in many therapeutic areas. Over the past 20 years, conventional SAR methods in medicinal chemistry have been highly successful in generating new

generations of drugs. However, to date, the only crystal structure available is for bovine rhodopsin, which shares variable sequence and functional similarity to the many hundreds of possible target proteins. It is thus very difficult to construct accurate models of the active site of a GPCR target.⁸⁶ Some hints about which amino acids are important in the active site can be derived from binding data for ligands with mutated receptors, sufficient in some cases to generate models of the target against which to guide virtual screening and compound design.⁸⁷ However, this type of modelling is fraught with many challenges. What is awaited is a breakthrough in the determination of the structures of this class of proteins. Current progress is reviewed in Lundstrom.⁸⁸

5.2 Protein–Protein Interactions

Many aspects of biological control and function operate through specific interactions between protein partners, bringing functionalities together or inducing conformational change for modulation of biological activity. However, disruption of protein–protein interactions by small molecules remains a considerable challenge. Molecular recognition in most protein–protein interactions relies on bringing together of large, often chemically and structurally featureless surfaces. There have been notable successes with targeting proteases where there are discrete active sites, but these experiences have emphasised the difficulties of creating drug-like ligands able to reach the spatially separated recognition sites that are important for specificity. One approach has been to identify the so-called “hot spots” important for binding and targeting these for small molecule intervention. One interesting idea is to exploit nearby cysteine residues for covalent localisation of ligands. The review by Arkin and Wells (2004) provides an overview of the current prospects and achievements in the area.

An alternative strategy is to generate or identify peptide fragments that can disrupt a protein–protein interaction. Structures of the protein–peptide complex can then be used to derive peptidomimetic compounds. Recent successes include the discovery of compounds against MDM2⁸⁹ and XIAP.⁹⁰

5.3 Using Structural Models of ADMET Mechanisms

The past 10 years has seen an increased molecular (and at times structural) understanding for some of the mechanisms in human biology responsible for drug pharmacokinetics. Structures for some of the cytochrome P450 enzymes responsible for oxidative metabolism have been determined.⁹¹ As more structures of ligand complexes are determined, it may be possible for *in silico* screening to highlight metabolic liability in a compound. Similarly, models have been developed for binding to the hERG channel, responsible for the cardiac side effects of some compounds.⁹² An interesting use for NMR screening is to characterise binding to albumin as a model for the plasma protein binding that can affect drug bioavailability.⁹³

These developments all contribute additional methods that can be used as early filters or structural alerts to guide the design of new compounds. However, the mechanisms contributing to ADMET are clearly very complex and multi-factorial, so it will be a long time before they can replace *in vivo* experiments.

5.4 Protein Therapeutics

This book focuses entirely on small molecule therapeutics. However, there have also been some important applications of structure-based methods in the design of protein therapeutics. The developments in molecular biology methodology of the 1980s provided the ability to introduce specific mutations into proteins to modify functional or physical properties. This found particular application in the development of industrial enzymes but also in engineering proteins for therapy. An early example is the work at Novo (and York), summarised in Brange *et al.*⁹⁴ where a structural understanding of the oligomerisation of insulin directed the design of specific mutations to create monomeric insulins with improved absorption characteristics. Another example is the extensive work on engineering of antibodies for therapy.^{95,96}

5.5 Other Targets for Structure-Based Drug Discovery

The three applications chapters in this book describe structure-based drug discovery against just three targets – neuraminidase, Factor Xa and the estrogen receptor. The methods have now been applied to many different classes of target, many of which are referred in the other chapters (see the table in the chapter by Brown and Flocco).

The following is a summary of the efforts for some of these other target classes.

Kinases. Noble *et al.*⁹⁷ have reviewed the detailed insights for drug design provided by kinase structures and Vieth *et al.*⁹⁸ have provided an up-to-date review of the kinome with valuable annotation for which kinases there are structures and those that are homologous to them. Kinases have seen intense activity over the past 10 years and have been a major focus for many innovations and investments in structure-based drug discovery. The next 5 years or so will see how many of these projects can deliver drug candidates into the clinic.

HIV proteins. The unraveling of the life cycle and molecular biology of the HIV virus prompted intense efforts to determine the structures of the distinctive proteins that support replication and infectivity of the virus. The earliest studies on HIV protease are described in Blundell *et al.*⁹⁹ and Greer *et al.*,⁷² and the article by Vacca and Condra²² describes the development of the first generation of inhibitors. The article by De Clercq¹⁰⁰ provides an update on more recent studies. In addition to the protease, there has also been intense efforts to develop drugs against HIV reverse transcriptase, some of which are summarised in the article by Ren and Stammers.¹⁰¹

The ribosome as an antibacterial target. The determination of the structures of the 30S and 50S subunits of the bacterial ribosome represent one of the triumphs of structural biology of the late 1990s Ramakrishnan, 2002,⁶ and continuing structural work is beginning to elucidate the detailed mechanisms of transcription and control. The ribosome is the target for many natural product antibiotics and the determination of the structures led to many efforts to use the structures for rational design (for example, the companies RiboTargets and Rib-X). These efforts have generated some excellent science, but as discussed in the introduction, there are many challenges in drug discovery against this class of target. It is possible to discover many compounds that bind with high affinity to the ribosome (and RNA in general). The difficulty is in achieving specificity – it is no accident that natural product antibiotics are large,

complex molecules. Such complexity and size may well be what is required to achieve selectivity. In addition, the compounds that bind RNA tend to be quite polar and not drug-like. Again, the natural product antibiotics have evolved to find some mechanism to gain cellular access that is extremely difficult to design in a small molecule.

Nuclear receptors. The chapter by Manas *et al.* in this book describes just one example of drug discovery against this class of protein that play a key regulatory role in cells. The article by Folkertsma *et al.*¹⁰² summarises the structural information available.

Phosphatases. There has been considerable activity in structure-based discovery against this class of proteins whose primary role is removing phosphate groups attached by kinases, thus providing balancing regulation for many cellular processes.¹⁰³ Examples of structure-based discovery efforts include Black *et al.*,¹⁰⁴ Lund *et al.*¹⁰⁵ and Zhao *et al.*¹⁰⁶. A challenge with this target is that the active site recognises a phosphate group and design of cell penetrating phosphate mimics is difficult. In addition, individual phosphatases are usually active against a number of kinases. As yet, no phosphatase inhibitor has progressed far in clinical trials necessary to demonstrate that the class of proteins is a true therapeutic target.

Phosphodiesterases. The discovery and development of Viagra¹ stimulated great interest in this protein class. Phosphodiesterases are implicated in a range of therapeutic areas as reviewed in Manallack *et al.*¹⁰⁷

6 Concluding Remarks

This introduction has hopefully provided an overall perspective on the field of structure-based drug discovery. The major challenges for the methods are the ability to determine the structure of the target, the ability to predict binding of ligands and the ability to design novel chemistry that is synthetically accessible.

The determination of the structure of the therapeutic target, if possible in complex with as many different ligand starting points as possible is clearly central to structure-based discovery. Many major classes of therapeutic target are still inaccessible to routine structure determination – such as the GPCRs and ion channels. In addition, many aspects of mammalian biology is governed by the transient assembly of large, multi-protein, multi-domain complexes and these remain a formidable challenge for structural study. These prizes remain available for the ambitious structural biologist.

Our ability to predict the conformational and energetic changes that accompany binding of a ligand to a protein target remains relatively weak. The methods that can be practically applied have remained essentially on a plateau since the development of empirical scoring methods in the early 1990s. Recent advances in techniques such as MM-PBSA¹⁰⁸ may offer the next level of improvement (see Barril and Soliva chapter). This ability to accurately determine interaction energy is the key for the next step of being able to model protein conformational change on ligand binding – a phenomenon which currently limits success (and confidence) in detailed structure-based design.

Finally, a major challenge is how to bring this wealth of structural, computational and assay data together to design new, improved compounds that can be readily

synthesised. There are few computational/informatics tools available to guide this process currently, and successful design crucially relies on effective inter-working and understanding between the different disciplines. It is hoped that the descriptions of the methods and selected applications provided in this book will give some insight into how this integration of the various methods is important, and emphasise how structure can provide insight and confidence to inspire and enable successful design.

References

1. S.F. Campbell, Science, art and drug discovery: a personal perspective, *Clin. Sci.*, 2000, **99**, 255–260.
2. D. Henry, Intercalation mechanisms: antitumor drug design based upon helical DNA as a receptor site, *Cancer Chemoth. Rep.*, 1972, **3**, 50.
3. W.C. Tse and D.L. Boger, Sequence-selective DNA recognition: natural products and nature's lessons, *Chem. Biol.*, 2004, **11**, 1607–1617.
4. S. Neidle and D.E. Thurston, Chemical approaches to the discovery and development of cancer therapies, *Nat. Rev. Cancer*, 2005, **5**, 285–296.
5. V. Ramakrishnan, Ribosome structure and the mechanism of translation, *Cell*, 2002, **108**, 557–572.
6. P.B. Moore and T.A. Steitz, The ribosome revealed, *Trends in Biochem. Sci.*, 2005, **30**, 281–283.
7. Q. Vicens and E. Westhof, Crystal structure of geneticin bound to a bacterial 16S ribosomal RNA A site oligonucleotide, *J. Mol. Biol.*, 2003, **326**, 1175–1188.
8. N. Foloppe, I.J. Chen, B. Davis, A. Hold, D. Morley and R. Howes, R., A structure-based strategy to identify new molecular scaffolds targeting the bacterial ribosomal A-site, *Bioorg. Med. Chem.*, 2004, **12**, 935–947.
9. I. Kola and J. Landis, Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.*, 2004, **3**, 711–715.
10. M. Dickson and J.P. Gagnon, Key factors in the rising cost of new drug discovery and development, *Nat. Rev. Drug Discov.*, 2004, **3**, 417–429.
11. U. Egner, J. Kratzschmar, B. Kreft, H.D. Poblenz and M. Schneider, The target discovery process, *Chembiochem.*, 2005, **6**, 468–479.
12. K.J. Dechering, The transcriptome's druggable frequenters, *Drug Discov. Today*, 2005, **15**, 857–864.
13. K.F. Koehler, L.A. Helgner, L.A. Haldosen, M. Warner and J.A. Gustafsson, Reflections on the discovery and significance of estrogen receptor beta, *Endocr. Rev.*, 2005, **26**, 465–478.
14. U. Banerji *et al.*, A phase I pharmacokinetic (PK) and pharmacodynamic (PD) study of 17-allylamino, 17 demethoxygeldanamycin (17-AAG) in patients with advanced malignancies, *J. Clin. Oncology*, 2005, **23**, 4152–4161.
15. P.J. Hajduk, J.R. Huth and S.W. Fesik, Druggability indices for protein targets derived from NMR-based screening data, *J. Med. Chem.*, 2005, **48**, 2518–2525.

16. A.M. Davis, D.J. Keeling, J. Steele, N.P. Tomkinson and A.C. Tinker, Components of successful lead generation, *Curr. Top. Med. Chem.*, 2005, **5**, 421–439.
17. A. Cavali, E. Poluzzi, F. de Ponti and M. Recanatini, Toward a pharmacophore for drugs inducing the long QT syndrome: insights from a COMFA study of HERG K⁺ channel blockers, *J. Med. Chem.*, 2002, **45**, 3844–3853.
18. B.A. Bunin and J.A. Ellman, A general and expedient method for the solid-phase synthesis of 1,4-benzodiazepine derivatives, *JACS*, 1992, **114**, 10997–10999.
19. R.E. Dolle, Comprehensive survey of combinatorial library synthesis: 2003, *J. Comb. Chem.*, 2004, **6**, 623–679.
20. B.W. Dymock, X. Barril, P.A. Brough, J.E. Cansfield, A. Massey, E. McDonald, R.E. Hubbard, A. Surgenor, S.D. Roughly, P. Webb, P. Workman, L. Wright and M.J. Drysdale, Novel, potent small molecule inhibitors of the molecular chaperone Hsp90 discovered through structure-based design, *J. Med. Chem.*, 2005, **48**, 4212–4215.
21. S.M. Roe, M.M.U. Ali, P. Meyer, C.K. Vaughan, B. Panaretou, P.W. Piper, C. Prodromou and L.H. Pearl, The mechanism of Hsp90 regulation by the protein kinase-specific cochaperone p50 (cdc37), *Cell*, 2004, **116**, 87–98.
22. J.P. Vacca and J.H. Condra, Clinically effective HIV-1 protease inhibitors, *Drug Discov. Today*, 1997, **2**, 261–272.
23. N. Foloppe, L.M. Fisher, R. Howes, P. Kierstan, A. Potter, A.G. Robertson and A.E. Surgenor, Structure-based design of novel chk1 inhibitors: insights into hydrogen bonding and protein–ligand affinity, *J. Med. Chem.*, 2005, **48**, 4332–4445.
24. S.H. Done, J.A. Brannigan, P.C.E. Moody and R.E. Hubbard, Ligand-induced conformational change in penicillin acylase, *J. Mol. Biol.*, 1998, **284**, 463–475.
25. T.G. Davies, R.E. Hubbard and J.R.H. Tame, Relating structure to thermodynamics: the crystal structures and binding affinity of eight OppA-peptide complexes, *Protein Sci.*, 1999, **8**, 1432–1444.
26. J.C. Kendrew, G. Bodo, H.M. Dintzis, R.G. Parrish, H. Wyckoff and D.C. Phillips, A three-dimensional model of the myoglobin molecular obtained by x-ray analysis, *Nature*, 1958, **181**, 662–666.
27. M.F. Perutz and H. Mazzarella, A preliminary x-ray analysis of haemoglobin H, *Nature*, 1963, **199**, 633–638.
28. C.C. Blake, D.F. Koenig, G.A. Mair, A.C. North, D.C. Phillips and V.R. Sarma, Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Å resolution, *Nature*, 1965, **206**, 757–761.
29. M.F. Perutz and J. Lehmann, Molecular pathology of human haemoglobin, *Nature*, 1968, **219**, 902–909.
30. C. Levinthal, Molecular model building by computer, *Sci. Am.*, 1966, **214**, 42.
31. C.D. Barry, H.E. Bosshard, R.A. Ellis and G.R. Marshall, Evolving macromolecular molecular modeling system, *Fed. Proc.*, 1974, **33**, 2368–2372.
32. A. Leo, C. Hansch and D. Elkins, Partition coefficients and their uses, *Chem. Rev.*, 1971, **71**, 525.

33. K.C. Chu, R.J. Feldmann, M.B. Shapiro, G.F. Hazard and R.I. Geran, Pattern recognition and structure-activity relation studies. Computer-assisted prediction of antitumor activity in structurally diverse drugs in an experimental mouse brain tumor system, *J. Med. Chem.*, 1975, **18**, 539–545.
34. F.C. Bernstein, T.F. Koetzle, G.J. Williams, E.E. Meyer, M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, Protein Data Bank – computer-based archival file for macromolecular structures, *J. Mol. Biol.*, 1977, **112**, 535–542.
35. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov and P.R. Bourne, The protein data bank, *Nucl. Acid Res.*, 2000, **28**, 235–242.
36. D.A. Matthews, R.A. Alden, J.T. Bolin, S.T. Freer, R. Hamlin, N. Xuong, J. Kraut, M. Poe, M. Williams and K. Hoogsteen, Dihydrofolate reductase: X-ray structure of the binary complex with methotrexate, *Science*, 1977, **197**, 452–455.
37. L.F. Kuyper, B. Roth, D.P. Baccanari, R. Ferone, C.R. Beddell, J.N. Champness, D.K. Stammers, J.G. Dann, F.E. Norrington, D.J. Baker and P.J. Goodford, Receptor-based design of dihydrofolate reductase inhibitors: comparison of crystallographically determined enzyme binding with enzyme affinity in a series of carboxy-substituted trimethoprim analogues, *J. Med. Chem.*, 1982, **25**, 1120–1122.
38. C.R. Beddell, P.J. Goodford, D.K. Stammers and R. Wootton, Species differences in the binding of compounds designed to fit a site of known structure in adult human haemoglobin, *Br. J. Pharmacol.*, 1979, **65**, 535–543.
39. F.F. Brown and P.J. Goodford, The interaction of some bis-arylhydroxysulphonic acids with a site of known structure in human haemoglobin, *Br. J. Pharmacol.*, 1977, **60**, 337–341.
40. D.W. Cushman, H.S. Cheung, E.F. Sabo and M.A. Ondetti, Design of potent competitive inhibitors of angiotensin-converting enzyme. Carboxyalkonyl and Mercaptoalkanoyl amino acids, *Biochemistry*, 1977, **16**, 5484–5491.
41. T.A. Jones, A graphics model building and refinement system for macromolecules, *J. Appl. Cryst.*, 1978, **11**, 268.
42. T.A. Jones, Diffraction methods for biological macromolecules. Interactive computer graphics: FRODO, *Methods Enzymol.*, 1985, **115**, 157–171.
43. T.A. Jones, J.Y. Zou, S.W. Cowan and M. Kjeldgaard, Improved methods for building protein models in electron-density maps and the location of errors in these models, *Acta Cryst.*, 1991, **A47**, 110–119.
44. R.J. Feldmann, D.H. Bing, B.C. Furie and B. Furie, Interactive computer surface graphics approach to study of the active site of bovine trypsin, *PNAS*, 1978, **75**, 5409–5412.
45. R. Langridge, T.E. Ferrin, I.D. Kuntz and M.L. Connolly, Real time color graphics in studies of molecular interactions, *Science*, 1981, **211** 661.
46. P. Gund, J.D. Andose, J.B. Rhodes and G.M. Smith, Three-dimensional molecular modeling and drug design, *Science*, 1980, **208**, 1425–1431.
47. C. Humblet and G.R. Marshall, Three dimensional modelling as an aid to drug design, *Drug Develop. Res.*, 1981, **1**, 409.

48. W.A. Hendrickson, Stereochemically restrained refinement of macromolecular structures, *Methods Enzymol.*, 1985, **115**, 252–270.
49. A.T. Brunger, G.M. Clore, A.M. Gronenborn and M. Karplus, 3-dimensional structure of proteins determined by molecular dynamics with interproton distance restraints – application to crambin, *PNAS*, 1986, **83**, 3801–3805.
50. A.T. Brunger, J. Kuriyan and M. Karplus, Crystallographic r-factor refinement by molecular dynamics, *Science*, 1987, **235**, 458–460.
51. G.M. Clore and A.M. Gronenborn, Multidimensional heteronuclear nuclear magnetic resonance of proteins, *Methods Enzymol.*, 1994, **239**, 349–363.
52. B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan and M. Karplus, CHARMM – a program for macromolecular energy minimization and dynamics calculations, *J. Comp. Chem.*, 1983, **4**, 187–217.
53. W.P. Jencks, On the attribution and additivity of binding energies, *PNAS*, 1981, **78**, 4046–4050.
54. P.R. Andrews, D.J. Craik and J.L. Martin, Functional group contributions to drug receptor interactions, *J. Med. Chem.*, 1984, **27**, 1648–1657.
55. P.J. Goodford, A computational procedure for determining energetically favourable binding sites on biologically important macromolecules, *J. Med. Chem.*, 1985, **28**, 849–857.
56. M.L. Connolly, Solvent-accessible surfaces of proteins and nucleic acids, *Science*, 1983, **221**, 709–713.
57. A. Nicholls, K. Sharp and B. Honig, Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons, *Proteins*, 1991, **11**, 281–296.
58. J.J. Baldwin *et al.*, Thienothiopyran-2-sulfonamides: novel topically active carbonic anhydrase inhibitors for the treatment of glaucoma, *J. Med. Chem.*, 1989, **32**, 2510–2513.
59. A. Miranker and M. Karplus, Functionality maps of binding sites: a multiple copy simultaneous search method, *Proteins*, 1991, **11**, 29–34.
60. M.B. Eisen, D.C. Wiley, M. Karplus and R.E. Hubbard, HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site, *Proteins, Struct., Func. and Gen.*, 1994, **19**, 199–221.
61. E.C. Meng, B.K. Shoichet and I.D. Kuntz, Automated docking with grid-based energy evaluation, *J. Comput. Chem.*, 1992, **13**, 505–524.
62. G. Jones, P. Willett, R.C. Glen, A.R. Leach and R. Taylor, Development and validation of a genetic algorithm for flexible docking, *J. Mol. Biol.*, 1997, **267**, 727–748.
63. M. Rarey, S. Wefing and T. Lengauer, Placement of medium-sized molecular fragments into active sites of proteins, *J. Comput. Aid. Mol. Des.*, 1996, **10**, 41–54.
64. X. Barril, R.E. Hubbard and S.D. Morley, Virtual screening in structure-based drug discovery, *Mini Rev. Med. Chem.*, 2004, **4**, 779–791.
65. D.W. Rodgers, Cryocrystallography, *Structure*, 1994, **2**, 1135–1140.
66. G.N. Murshudov, A.A. Vagin and E.J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method, *Acta Cryst. D*, 1997, **53**, 240–255.

67. R.J. Morris, A. Perrakis and V.S. Lamzin, ARP/wARP and automatic interpretation of protein electron density maps, *Method. Enzymol.*, 2003, **374**, 229–244.
68. T.J. Oldfield, Automated tracing of electron-density maps of proteins, *Acta Cryst. D*, 2003, **59**, 483–491.
69. S.B. Shuker, P.J. Hajduk, R.P. Meadows and S.W. Fezick, Discovering high-affinity ligands for proteins: SAR by NMR, *Science*, 1996, **274**, 1531–1534.
70. T. Oltersdorf *et al.*, An inhibitor of Bcl-2 family proteins induces regression of solid tumours, *Nature*, 2005, **435**, 677–681.
71. M. Von Itzstein *et al.* Rational design of potent sialidase-based inhibitors of influenza-virus replication, *Nature*, 1993, **363**, 418–423.
72. J. Greer, J.W. Erickson, J.J. Baldwin and M.D. Varney, Application of three-dimensional structures of protein target molecules in structure-based drug design, *J. Med. Chem.*, 1994, **37**, 1035–1054.
73. J.T. Randolph and D.A. DeGoey, Peptidomimetic inhibitors of HIV protease, *Curr. Top. Med. Chem.*, 2004, **10**, 1079–1095.
74. R.A. Chrusciel and J.W. Strohbach, Non-peptidic HIV protease inhibitors, *Curr. Top. Med. Chem.*, 2004, **4**, 1097–1114.
75. C. Mattos and D. Ringe, Locating and characterizing binding sites on proteins. *Nat. Biotechnol.*, 1996, **14**, 595–599.
76. A.C. English, C.R. Groom and R.E. Hubbard, Experimental and computational mapping of the binding surface of a crystalline protein, *Protein Eng.*, 2001, **14**, 47–59.
77. V.L. Nienaber, P.L. Richardson, V. Klighofer, J.J. Bouska, V.L. Giranda and J. Greer, Discovering novel ligands for macromolecules using X-ray crystallographic screening, *Nat. Biotechnol.*, 2000, **18**, 1105–1108.
78. M.J. Hartshorn, C.W. Murray, A. Cleasby, M. Frederickson, I.J. Tickle and H. Jhoti, Fragment based lead discovery using X-ray crystallography, *J. Med. Chem.*, 2005, **48**, 403–413.
79. E.R. Zartier and M.J. Shapiro, Fragonomics: fragment-based drug discovery, *Curr. Opin. Chem. Biol.*, 2005, **9**, 366–370.
80. C.A. Lipinski, F. Lombardo, B.W. Dominy and P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv. Drug Deliver. Rev.*, 1997, **23**, 3–25.
81. C. Lipinski and A. Hopkins, Navigating chemical space for biology and medicine, *Nature*, 2004, **432**, 855–861.
82. T.I. Oprea, A.M. Davis, S.J. Teague and P.D. Leeson, Is there a difference between leads and drugs? A historical perspective, *J. Chem. Inf. Comp. Sci.*, 2001, **41**, 1308–1315.
83. M.M. Hann, A.R. Leach and G. Harper, Molecular complexity and its impact on the probability of finding leads for drug discovery, *J. Chem. Inf. Comp. Sci.*, 2001, **41**, 856–864.
84. A. Hillisch, L.F. Pineda and R. Hilgenfeld, Utility of homology models in the drug discovery process, *Drug Discov. Today*, 2004, **9**, 659–669.
85. T. Klabunde and G. Hessler, Drug design strategies for targeting G-protein-coupled receptors, *Chem. Bio. Chem.*, 2002, **3**, 928–944.

86. E. Archer, B. Maignet, C. Escricout, L. Pradayrol and D. Fourmy, Rhodopsin crystal; new template yielding realistic models of G-protein-coupled receptors? *Trends Pharmacol. Sci.*, 2003, **24**, 36–40.
87. A. Evers and G. Klebe, Ligand-supported homology modeling of G-protein-coupled receptor sites: models sufficient for successful virtual screening, *Angew Chem. Int Edit.*, 2004, **43**, 248–251.
88. K. Lundstrom, Structural biology of G protein-coupled receptors, *Biorg. Med. Chem. Lett.*, 2005, **15**, 3654–3657.
89. N. Fotouhi and B. Graves, Small molecule inhibitors of p53/MDM2 interactions, *Curr. Top. Med. Chem.*, 2005, **5**, 150–165.
90. T.K. Oost *et al.*, Discovery of potent antagonists of the antiapoptotic protein XIAP for the treatment of cancer, *J. Med. Chem.*, 2004, **47**, 4417–4426.
91. P.A. Williams, J. Cosme, D.M. Vinkovic, A. Ward, H.C. Angove, P.J. Day, C. Vornrhein, I.J. Tickle and H. Jhoti, Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone, *Science*, 2004, **305**, 683–686.
92. F. Osterberg and J. Aqvist, Exploring blocker binding to a homology model of the open hERG K⁺ channel using docking and molecular dynamics methods, *FEBS Lett.*, 2005, **579**, 2939–2944.
93. P.J. Hajduk, R. Mendoza, A.M. Petros, J.R. Huth, M. Bures, S.W. Fesik and Y.C. Martin, Ligand binding to domain-3 of human serum albumin: a chemometric analysis, *J. Comp. Aid. Mol. Des.*, 2003, **17**, 93–102.
94. J. Brange, U. Ribel, J.F. Hansen, G.G. Dodson, M.T. Hansen, S. Havelund, S.G. Melberg, F. Norris, K. Norris, L. Snel, A.R. Sorensen and H.O. Voigt, Monomeric insulins obtained by protein engineering and their medical implications, *Nature*, 1988, **333**, 679–682.
95. J.R. Adair, Engineering antibodies for therapy, *Immunol. Rev.*, 1992, **130**, 5–40.
96. R.L. Brady, D.J. Edwards, R.E. Hubbard, J.S. Jiang, G. Lange, S.M. Roberts, R.J. Todd, J.R. Adair, J.S. Emtage, D.S. King and D.C. Low, Crystal structure of a chimeric Fab' fragment of an antibody-binding tumor cells, *J. Mol. Biol.*, 1992, **227**, 253–264.
97. M.E.M. Noble, J.A. Endicott and L.N. Johnson, Protein kinase inhibitors: insights into drug design from structure, *Science*, 2004, **303**, 1800–1805.
98. M. Vieth, J.J. Sutherland, D.H. Robertson and R.M. Campbell, Kinomics: characterizing the therapeutically validated kinase space, *Drug Discov. Today*, 2005, **10**, 839–846.
99. T.L. Blundell, R. Lapatto, A.F. Wilderspin, A.L. Hemmings, P.M. Hobart, D.E. Danley and P.J. Whittle, The 3-D structure of HIV-1 proteinase and the design of anti-viral agents for the treatment of AIDS, *TIBS*, 1990, **15**, 425–430.
100. E. De Clercq, Emerging anti-HIV drugs, *Expert Opin. Emerg. Drugs*, 2005, **10**, 241–273.
101. J. Ren and D.K. Stammers, HIV reverse transcriptase structures: designing new inhibitors and understanding mechanisms of drug resistance, *Trends Pharmacol. Sci.*, 2005, **26**, 4–7.

102. S. Folkertsma, P. van Noort, J. van Durme, H.-J. Joosten, E. Bettler, W. Fleuren, L. Oliveira, F. Horn, J. de Vlieg and G. Vriend, A family-based approach reveals the function of residues in the nuclear receptor ligand-binding domain, *J. Mol. Biol.*, 2004, **341**, 321–335.
103. A. Alonso, J. Sasin, N. Bottini, I. Friedberg, A. Osterman, A. Godzik, T. Hunter, J. Dixon and T. Mustelin, Protein tyrosine phosphatases in the human genome, *Cell*, 2004, **117**, 699–711.
104. E. Black *et al.*, Structure-based design of protein tyrosine phosphatase-1B inhibitors, *Bioorg. Med. Chem. Letts.*, 2005, **15**, 2503–2507.
105. I.K. Lund *et al.*, Structure-based design of selective and potent inhibitors of protein-tyrosine phosphatase ?, *J. Biol. Chem.*, 2004, **279**, 24226–24235.
106. H. Zhao *et al.*, Isoxazole carboxylic acids as protein tyrosine phosphatase 1B (PTP1B) inhibitors, *Bioorg. Med. Chem. Lett.*, 2004, **15**, 5543–5546.
107. D.T. Manallack, R.A. Hughes and P.E. Thompson, The next generation of phosphodiesterase inhibitors: structural clues to ligand and substrate selectivity of phosphodiesterases, *J. Med. Chem.*, 2005, **48**, 3449–3462.
108. P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case and T.E. Cheatham III, Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts Chem. Res.*, 2000, **33**, 889–897.
109. M.R. Arkin and J.A. Wells, Small-molecule inhibitors of protein–protein interactions: progressing towards the dream, *Nat. Rev. Drug Discov.*, 2004, **3**, 301–317.
110. C.T. Baker, F.G. Salituro, J.J. Court, D.D. Deiningner, E.E. Kim, B. Li, P.M. Novak, B.G. Rao, S. Pazhanisamy, W.C. Schairer and R.D. Tung, Design, synthesis, and conformational analysis of a novel series of HIV protease inhibitors, *Bioorg. Med. Chem. Lett.*, 1998, **15**, 3631–3636.
111. J. Boger *et al.*, Novel renin inhibitors containing the amino acid statine, *Nature*, 1983, **303**, 81–84.
112. H.J. Bohm, The development of a simple empirical scoring function to estimate the binding constant for a protein ligand complex of known 3-dimensional structure, *JCAMD*, 1994, **8**, 243–256.
113. D.J. Knowles, N. Foloppe, N.B. Matassove and A.I. Muchie, The bacterial ribosome, a promising focus for structure-based drug design, *Curr. Opin. Pharmacol.*, 2002, **2**, 501–506.
114. F.G. Salituro, C.T. Baker, J.J. Court, D.D. Deiningner, E.E. Kim, B. Li, P.M. Novak, B.G. Rao, S. Pazhanisamy, M.D. Porter, W.C. Schairer and R.D. Tung, Design and synthesis of novel conformationally restricted HIV protease inhibitors, *Bioorg. Med. Chem. Lett.*, 1990, **8**, 3637–3642.
115. S.P. Williams, L.F. Kuyper and K.H. Pearce, Recent applications of protein crystallography and structure-guided drug design, *Curr. Opin. Chem. Biol.*, 2005, **9**, 371–380
116. I. Hunt, From gene to protein: a review of new and enabling technologies for multi-parallel protein expression, *Prot. Expr and Purif.*, 2005, **40**, 1–22.

117. C. W. Goulding and L. J. Perry, Protein production in *Escherichia coli* for structural studies by X-ray crystallography, *J. of Struct. Biol.*, 2003, **142**, 133–143.
118. T.A. Kost, J.P. Condreay and D.L.Jarvis, Baculovirus as versatile vectors for protein expression in insect and mammalian cells, *Nature Biotech.*, 2005, **23**, 567–575.
119. S. Yokoyama, Protein expression systems for structural genomics and proteomics, *Curr. Opin. Chem. Biol.*, 2003, **7**, 39–43.
120. P. Braun and J. LaBaer, High throughput protein production for functional proteomics, *Trends in Biotech.*, 2003, **21**, 383–388.
121. D.S. Waugh, Making the most of affinity tags, *Trends in Biotech.*, 2005, **23**, 316–320.