CHAPTER

22

Protein Crystallography and Drug Discovery

Jean-Michel Rondeau[1] *and Herman Schreuder*[2]

[1]Novartis Institutes for BioMedical Research, Basel, Switzerland; [2]Sanofi, R&D LGCR/Struct., Design & Informatics FF, Frankfurt am Main, Germany

OUTLINE

I. Introduction 511

II. Historical Background 512
- A. *The Early Days of Crystallography* 512
- B. *The Current State-of-the-Art* 513
- C. *Examples of Structure-Based Drug Discovery* 513

III. Basic Principles and Methods of Protein Crystallography 514
- A. *Crystallization* 514
- B. *Data Collection* 519
- C. *From Diffraction Intensities to a Molecular Structure* 520
- D. *Information Content and Limitations of Protein Crystal Structures* 523

IV. Applications 527
- A. *Target Identification and Selection* 527
- B. *Hit/Lead Generation* 528
- C. *Lead Optimization* 530

V. Two Selected Examples 531
- A. *Imatinib (Gleevec™)* 531
- B. *The β_2-Adrenergic Receptor* 532

VI. Outlook 532

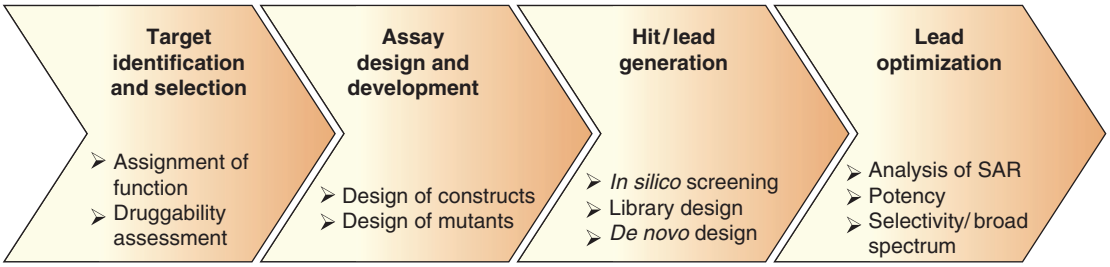
References 533

If you can look into the seeds of time
And say which grain will grow and which will not
Speak then to me ... ***Shakespeare, Macbeth*** *[1]*

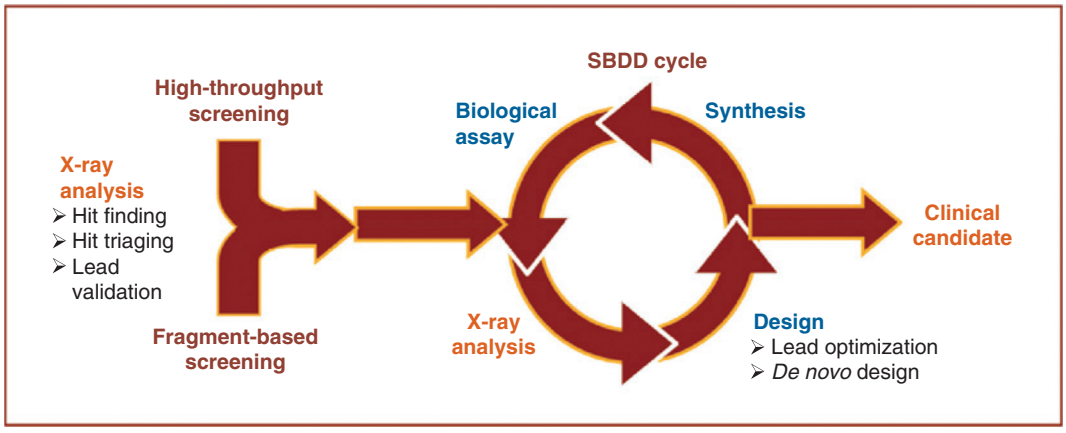
I. INTRODUCTION

Protein crystallography is the only technology to date that allows us to "see" how a ligand (hit, lead) is bound to its target protein. No wonder, therefore, that it has had a profound influence on pharmaceutical research since its inception during the 1970s [2]. Today, the use of structural information pervades all phases of pre-clinical research: target identification and validation, the development of *in vitro* assays, finding the best hit/lead finding strategies, and the entire lead optimization phase [3,4] (Figure 22.1a). The impact of structural biology on the daily work of medicinal chemists by replacing traditional trial and error methods by structure based design has been particularly strong [5,6].

However, exploiting this detailed 3D structural information is not trivial, in part because it must be combined with other constraints such as synthetic accessibility, absorption, distribution, metabolism, and excretion (ADME) properties, toxicology, and intellectual property. Nonetheless, the rewards are immense. Structural information not only clarifies structure–activity relationships, reveals binding modes and bioactive conformations, and



(a)



(b)

FIGURE 22.1 (a) Contributions of X-ray analysis to the drug discovery value chain; (b) Integration of X-ray analysis into hit/lead finding, triaging, validation, and optimization. SBDD: Structure-based drug design.

unveils new binding pockets or allosteric binding sites but also opens new and diverse drug discovery avenues, such as *in silico* screening, design of focused chemical libraries, and *de novo* design of new ligand scaffolds. Having access to such information provides a strong competitive advantage and makes the professional life of medicinal chemists highly stimulating and often very gratifying.

A hundred years after the recording of the first diffraction image, X-ray crystallography is still rapidly evolving with large scale structural genomics efforts focusing on human proteins of medical importance and on potential drug targets from clinically-relevant pathogens [7]. Advances in technologies and methods have led to a breakthrough in the crystallization of membrane proteins and a drastic reduction in the time needed to generate crystal structures. These developments not only produce a wealth of structural data but also enable high-throughput methods in protein crystallography [8] (e.g., hit triaging and *de novo* hit finding by fragment-based screening; Figure 22.1b), which allow the identification of novel, chemically attractive leads and their successful optimization to highly potent drug candidates [9].

In this chapter, we will describe how crystallographic data contribute to the different phases of pharmaceutical research. We will emphasize not only the strengths but also the technical limitations of protein crystallography, so that any medicinal chemist can gauge if and how a project could benefit from this technology. A brief outline of the basic principles and methods of protein crystallography is also provided. To make proper use of structural data, it is essential to be aware of the limitations and potential uncertainties associated with X-ray structures. We also hope that this chapter will contribute to more effective communication between chemists and their fellow crystallographers.

II. HISTORICAL BACKGROUND

A. The Early Days of Crystallography

Crystallography made its first notable contributions to the progress of biology and medicine well before the elucidation by Kendrew and Perutz in 1958–1960 of the first protein structures, myoglobin [10] and hemoglobin [11]. The preparation of "blood crystals"—in fact, hemoglobin crystals—was first reported by Hünefeld in 1840

[12]. During the second half of the nineteenth century, this initial observation sparked considerable interest in the crystallization of hemoglobins and other proteins, mainly from plant seeds [12]. This groundwork set the stage for the first major achievement in this field, which took place 1926–1935: the demonstration of the molecular nature of enzymes and viruses through their isolation in crystalline form by Sumner, Northrop, and Stanley [12]. The second major contribution was made in the early 1950s, when X-ray diffraction photographs of DNA produced by Franklin [13] and Wilkins [14] could be used as a guide by Crick and Watson, ultimately leading to their discovery of the double helical structure of DNA [15]. The next achievement—the determination of the myoglobin and hemoglobin structures by Kendrew and Perutz—revealed for the very first time the intricacies of the architecture of proteins, while also shedding light on the molecular basis of sickle cell anemia [16]. Since then, these milestone studies have been followed by a rich crop of other stunning crystallographic feats. The 3D structures of the human common cold virus [17], the photosynthetic reaction center [18], the F1-ATP synthase [19], the proteasome [20], the nucleosome core particle [21], the 30S and 50S ribosomal particles [22–25], the RNA polymerase II [26], potassium channels [27,28], and the β2 adrenergic G-protein-coupled receptor [29] have all been solved using X-ray diffraction methods, in spite of their daunting size and biochemical complexity.

B. The Current State-of-the-Art

Today, more than 100,000 crystal structures are publicly available from the Protein Data Bank (see Box 22.3), comprising about 40,000 unique sequences, a number that is rapidly increasing due to efforts in academia, structural genomics consortia, and pharmaceutical and biotech companies. This means that for most drug targets, either the structure itself or at least the structure of a homologous protein is available for use in structure-based drug discovery.

On-going developments in miniaturization and robotics allow an extensive screening for crystallization conditions with only a few milligrams of protein, while microfocus beamlines and ever brighter synchrotron sources allow the collection of data from smaller and smaller crystals. Automatic sample mounting and fast hybrid-pixel detectors allow fast data collection on large numbers of crystals, while software pipelines automatically process data [30], calculate electron-density maps, and do the initial ligand fitting [31,32]. Only at this stage does the crystallographer need to examine the structure and manually continue the fitting and refinement process.

X-ray crystallography is now routinely used in drug discovery projects involving soluble targets, and the same may soon be true for membrane proteins. However, a prerequisite for using this method is the availability of suitable crystals. For high-throughput methods such as hit triaging and fragment-based methods, the crystallization process should be robust and routinely produce well-diffracting crystals. This means that in the majority of cases, the main bottleneck does not reside in the X-ray analysis itself but rather in the identification and production of a stable, well-behaved recombinant version of the protein of interest, which is amenable to crystallization.

Today's crystallographers—particularly those working in industry—are often faced with difficult to produce, poorly behaved, poorly characterized targets. Such challenging proteins require a lot of biochemical ingenuity and cannot be conquered without a dedicated, appropriately resourced effort in protein production and characterization. For academic users, large facilities are available to generate and test thousands of clones for expression and solubility [33]. It would be helpful if industrial crystallographers could get access to such facilities as well.

C. Examples of Structure-Based Drug Discovery

1. Captopril

Drugs with anti-sickling properties were the first drugs ever to be studied in complex with their protein target by means of X-ray analysis [16]. These seminal studies spurred the first attempts at designing improved compounds via a structure-based approach [34]. Soon after, the discovery of captopril [35] (Figure 22.2a), an anti-hypertensive agent and the first marketed, orally active inhibitor of the human angiotensin-converting enzyme (ACE), hailed the beginning of a new era for pharmaceutical research. For the first time, a drug had been rationally designed on the basis of structural information, hence providing the first compelling demonstration of the power of the structure-based approach.

Interestingly, the successful design of captopril used simple chemical concepts guided by a hypothetical "paper-and-pencil" model of substrate and inhibitor binding to the enzyme active site that had been inferred from the crystal structure of bovine carboxypeptidase A. The X-ray structure of human ACE became available only in 2003, twenty-five years after the discovery of the captopril class of drugs. While the crystallographic analysis of the ACE complex with captopril [36] confirmed the designed mode of interaction, it revealed little structural similarity overall with carboxypeptidase A.

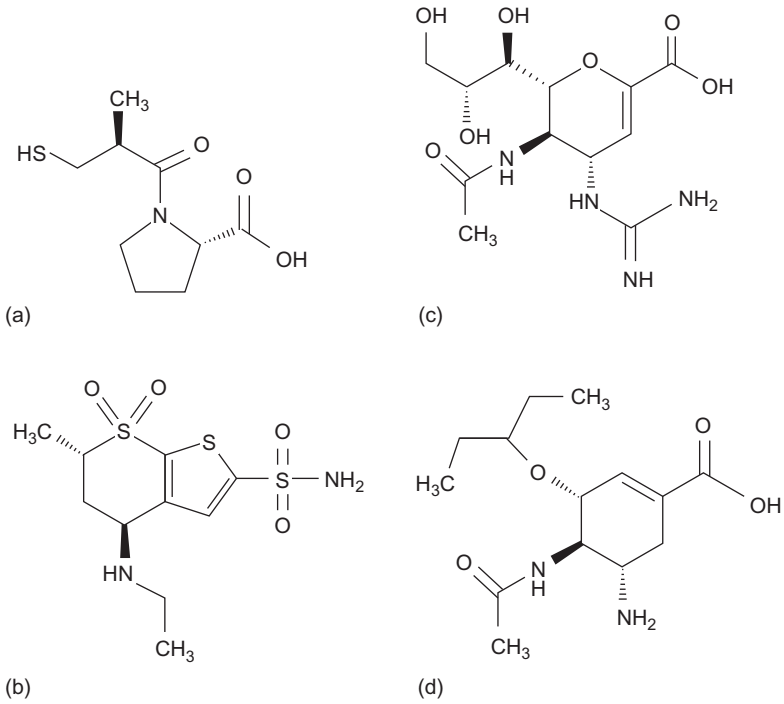


FIGURE 22.2 The first marketed drugs derived from structure-based design: (a) captopril (Capoten™); (b) dorzolamide (Trusopt™); (c) zanamivir (Relenza™); and (d) oseltamivir (Tamiflu™).

2. *Dorzolamide*

Dorzolamide (Figure 22.2b) is the first example of an approved drug that benefited from the complete armament of structure-based design (i.e., multiple X-ray analyses with the human enzyme target [37] combined with sophisticated molecular modeling studies, including in-depth conformational analyses using *ab initio* quantum chemistry calculations) [38]. Dorzolamide is a subnanomolar carbonic anhydrase II inhibitor that was developed in the early 1990s as a topical agent for the treatment of glaucoma.

3. *Relenza and Tamiflu*

The discovery of oseltamivir (Tamiflu™TM; Figure 22.2d), an inhibitor of *influenza* neuraminidase, is another early example of successful structure-based drug design. Interestingly, long standing efforts to identify neuraminidase inhibitors via random screening or the rational design of transition-state analogues had failed to produce any potent compounds [39] until the crystal structure of neuraminidase became available in 1983 [40,41]. A GRID [42] analysis of the sialic acid complex immediately suggested a simple modification of a known sialic acid analogue with low micromolar affinity. Remarkably, only two compounds were synthesized, and both turned out to be extremely potent inhibitors, with Ki values of 50 nM and 0.2 nM, respectively [43]. The most potent compound, zanamivir (Relenza™; Figure 22.2c), became the first marketed neuraminidase inhibitor. Further structure-based design concentrated on the development of an analogue with improved stability and lipophilicity. These efforts very quickly resulted in oseltamivir, a second generation, orally bioavailable drug [44] (Figure 22.2d).

III. BASIC PRINCIPLES AND METHODS OF PROTEIN CRYSTALLOGRAPHY

A. Crystallization

1. *What are Protein Crystals?*

Protein crystals (Figure 22.3), like any crystal of organic or inorganic compounds, are regular 3D arrays of identical molecules or molecular complexes (Figure 22.4). Depending on the symmetry of this arrangement (described by the space group), all molecules in a crystal have a limited number of unique orientations with respect to the crystal lattice. The diffraction of all individual molecules adds up to yield intensities that are sufficiently strong to be measured, the crystal lattice thus acting as an amplifier. An explanation of some common crystallographic terms is given in Box 22.1.

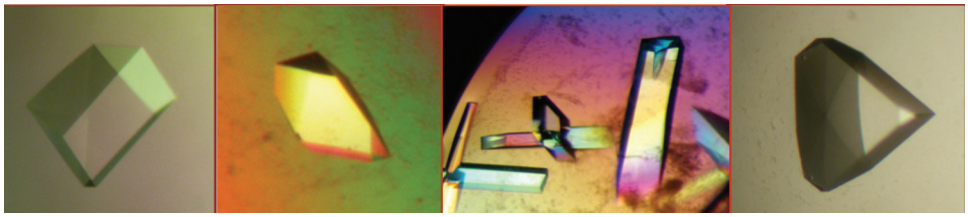


FIGURE 22.3 Examples of protein crystals. From left to right: β-secretase inhibitor complex; human farnesyl pyrophosphatase in complex with zoledronic acid; abl kinase domain in complex with imatinib (courtesy of SW Cowan-Jacob, Novartis); cdk2 inhibitor complex.

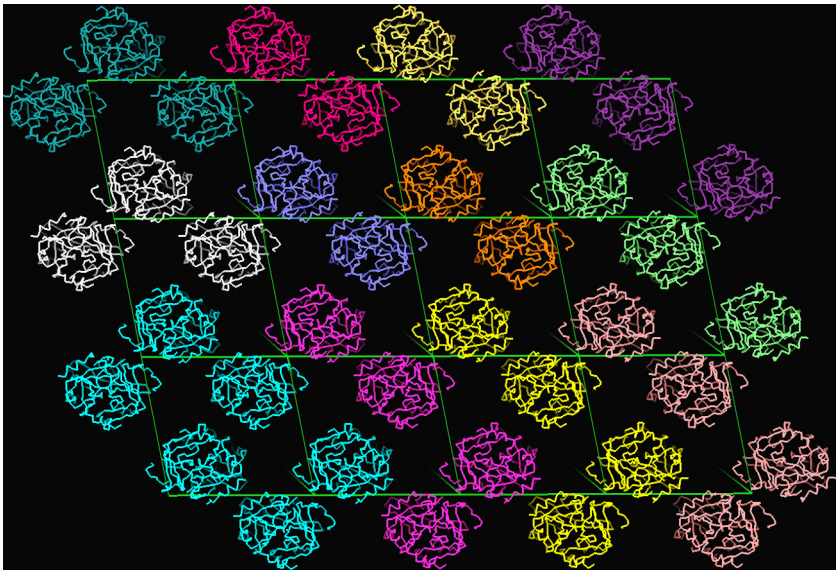


FIGURE 22.4 Crystal packing of a human thrombin complex. Twelve unit cells with one layer of molecules are shown. By looking carefully, one can see that the two molecules in each unit cell are rotated 180° with respect to each other. Protein crystals used for X-ray diffraction extend into three dimensions and consist of many layers of molecules. The next layer of thrombin molecules fits into the holes present in the layer shown.

BOX 22.1

SOME COMMON CRYSTALLOGRAPHIC TERMS

Space group: The group of symmetry operators that describe the symmetry of the crystal. Since biological molecules are optically active, their crystals belong to one of the sixty-five noncentrosymmetric space groups.

Unit cell: The basic building block of a crystal. The whole crystal can be generated by repeated unit translations of the cell in three dimensions. The unit cell is characterized by its axes a, b, c, and the angles (α, β, γ) between them.

Asymmetric unit: The smallest motif from which the whole unit cell can be generated by applying the symmetry operators of the space group. The asymmetric unit may contain one or more copies of the protein or complex under study. In the case of oligomeric particles, the asymmetric unit may contain one or more complete particles or only one or more subunits if some symmetry axes of the particle coincide with some symmetry axes of the crystal.

Reflection: A diffracted beam of X-ray, characterized by its indices h,k,l, and caused by reflection from the lattice planes making intercepts a/h, b/k, and c/l with the unit cell axes. Each reflection contains information on the entire structure. Reflections occurring at high scattering angles have high indices and carry high resolution information (i.e., they correspond to a fine sampling of the structure), while those observed close to the direction of the incident beam have small indices and carry low resolution information (i.e., they correspond to a coarse sampling of the structure).

Resolution: The resolution limit corresponds to the highest scattering angle at which reflections can still be measured (cf. Box 22.2). Individual atoms can be fully resolved when the resolution is better than 1.0Å.

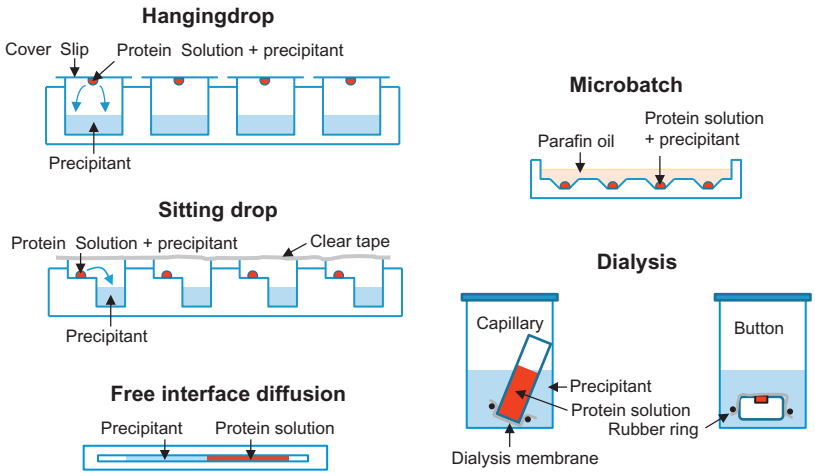


FIGURE 22.5 Methods used to crystallize proteins. For the hanging drop, sitting drop, and microbatch methods, a protein solution is mixed with a precipitant solution (usually in a 1:1 ratio) and set to equilibrate. With the hanging drop and sitting drop methods, water from the less concentrated protein drop will transfer to the more concentrated precipitant solution via the vapor phase, as indicated by blue arrows. This makes the protein drop shrink, thereby increasing the protein and precipitant concentration. With the other two methods, the protein solution is brought into contact with a precipitant solution either directly (free interface diffusion) or via a dialysis membrane (dialysis method). In all cases, adding precipitant to the protein solution creates a supersaturated state, causing to protein to precipitate and—if one is lucky—form crystals.

One notable difference between crystals of small molecules and macromolecular crystals is the very large solvent content of the latter. Protein crystals typically contain 30–80 percent (v/v) solvent (in fact, aqueous crystallization buffer) [45]. Only a fraction of the protein surface is involved in crystal contacts, the rest being fully solvated, pretty much as in solution. As a consequence, protein crystals are very soft and fragile. But on the positive side, low molecular weight ligands, co-factors, and substrates can diffuse from the surrounding mother liquor into the solvent channels in the crystal. If their binding site is not occluded by crystal contacts, the complex can be formed *in situ*. Usually, small conformational changes can take place within the crystal lattice without damaging the crystal, and sometimes very large structural changes can be accommodated as well. Therefore, enzyme crystals are very often active as catalysts.

2. *How do we Get Crystals?*

Crystals are produced by slowly driving a concentrated protein solution into a state of supersaturation [46,47]. Under the right conditions, the protein will not form an amorphous precipitate but will instead settle into a well-ordered crystalline array. Methods for achieving a high degree of supersaturation involve dialyzing away the salt, if the salt has a strong solubilizing (or "salting-in") effect, concentrating a nearly saturated protein solution by evaporation (usually in a hanging or sitting drop set-up; see below) and adding "precipitants" such as poly (ethylene glycol) or high salts such as ammonium sulfate, which has a strong "salting-out" effect on proteins. Other possibilities, which are less often used, are temperature and pH gradients. The methods most often used for crystallization screening are shown in Figure 22.5.

Because of the low protein consumption (100 nl protein solution per drop) and compatibility with crystallization robots and automatic crystal imagers, most crystallization experiments are performed as sitting drop experiments in 96-well microtiter plates [48,49]. Manual crystallizations are usually done in 24-well microtiter plates using the hanging drop set-up with 1 μl protein solution per drop. Other crystallization techniques have been developed, such as crystallization under oil (microbatch) and free interface diffusion using capillaries or microfluidic chips [46,47,49]. Microfluidic methods are also used for determining phase diagrams.

3. *Specific Problems and Solutions*

Crystallization conditions are published. When crystallization conditions are already known, 1 mg of protein may be enough to produce a series of crystals with different inhibitors. One should bear in mind, however, that published crystallization protocols are often difficult to reproduce. It is wise, in a first step, to follow as closely as possible the published expression, purification, and crystallization protocols. Particular attention should be paid to the protein construct, since minor changes to the amino-acid sequence can have a dramatic influence on the solubility, stability, and crystallization behavior.

De novo crystallization. Obtaining X-ray quality crystals is usually the most difficult and time-consuming step of a new structure determination project, notably in the case of a novel, poorly characterized gene product [49]. Modern crystallization robots and miniaturization have considerably simplified the process and shortened the time needed to set up extensive crystallization screening experiments, while dramatically reducing the amount of material needed. By using a protein solution of 10 mg/ml and 100 nl protein solution per drop, 0.1 mg of protein is sufficient to screen 100 conditions, and with a few mg of protein, one can screen several thousand individual

crystallization conditions. Nevertheless, this is often not enough to obtain suitable crystals for new and difficult targets [49] since the protein (construct) may simply not be crystallizable. Several protein constructs or protein variants may have to be generated to increase the chances of finding one that is amenable to crystallization. For this reason, strong and dedicated support in molecular biology, protein expression, and biochemistry are an absolute must for a successful protein crystallography laboratory. Furthermore, in an industrial setting, these activities should be initiated as early as possible, to ensure that crystals are available before chemistry activities are started.

When novel protein targets cannot be produced or crystallized, one should consider the possibility of using a known homolog or anti-target instead. If the binding sites are sufficiently similar, the binding modes of some key compounds or scaffolds can be deciphered, and this information can be fed into the drug-design process.

Conformational heterogeneity. Recombinant proteins designed for assay purposes are often not suitable for crystallization experiments when they contain fusion partners, long tags, floppy ends, disordered or intrinsically unstructured regions, or loosely linked domains [50] that in general prevent crystallization. Intrinsically unstructured regions can usually be identified from the amino-acid sequence as polypeptide segments with low sequence diversity [51]. Domain boundaries can be pinned down by limited proteolysis or with the help of homology modeling, and constructs can be made expressing only a single domain. If tags or fusion partners are needed for enhanced expression and/or ease of purification, then a protease recognition sequence should be engineered to allow their removal before crystallization. Conformational heterogeneity can also be reduced by buffer additives or ligands [52]. Biophysical techniques such as thermal-shift assays and nuclear magnetic resonance (NMR) can be used for the identification of suitable ligands and additives.

Glycosylation. Glycosylated proteins often give poorly diffracting crystals. To circumvent this problem, several strategies can be used: [53] glycosylation can be chopped off enzymatically with PNGase F or other endoglycosidases; glycosylation sites can be mutated away; or a non-glycosylated form can be produced using a prokaryotic expression system. Glycosylation also increases the solubility and stability of proteins, so crystallization of the glycosylated protein should be tried as well.

Proteolytic cleavage. Proteases cleave other proteins, including themselves, and even trace amounts of a contaminating protease may wreak havoc during the time it takes (days to weeks) for crystals to grow. When crystallizing a protease for the first time, it is always a good idea to add the most potent inhibitor available to the crystallization set-up. If a contaminating protease is a problem, one could add a protease inhibitor cocktail or a general broad-spectrum inhibitor like PMSF.

Phosphorylation. Protein kinases and other proteins involved in signaling are often produced from eukaryotic expression hosts as a mixture of inactive (unphosphorylated) and active (with one or more phosphorylations) species, which causes conformational heterogeneity. One possible workaround is mutating the phosphorylation site (s) to glutamate, which mimics the phosphorylation, thus producing a constitutionally active kinase. Other options include mutating the phosphorylation site away or expression in the presence of a ligand or inhibitor [52,54]. Co-expression with a phosphatase is another strategy that may reduce the heterogeneity of the phosphorylation [55].

Membrane proteins The crystallization of membrane proteins is particularly challenging [56] and has long been off-limits for industrial crystallographers. Traditionally, only the soluble domains (catalytic, ligand binding) of multidomain membrane proteins were expressed and crystallized [57]. In recent years, however, the structure of a large number of integral membrane proteins has been published [58], and crystallography of membrane proteins is entering drug discovery laboratories.

Factors that have enabled this breakthrough are improved expression and purification of membrane proteins, optimized detergents, robotics and miniaturization that allow extensive screening of crystallization conditions using very little protein, and microfocus beam lines at synchrotrons that allow the collection of data from tiny crystals. Also new water-detergent phases such as bicelles [59] and lipidic cubic phases [60] have had a large impact on the crystallizing success of membrane proteins. However, the greatest breakthrough came from overcoming the inherent flexibility of many membrane proteins, especially G-protein-coupled receptors (GPCRs).

Most GPCRs such as the β_2-adrenergic receptor (β_2AR) are present as a mixture of conformations ranging from the inactive to the active conformation [61] (Figure 22.6). Adding an agonist merely shifts the equilibrium toward the active conformation, but only binding of a G-protein stabilizes the active state. For crystallization, GPCRs are stabilized by the removal of flexible loops, replacing the flexible intracellular loop three by a stable and soluble protein like T4 lysozyme, the addition of antibodies (Fab fragments or single-chain camelid antibody fragments), and systematic scanning mutagenesis or random evolutionary mutagenesis [58]. In general, these methods stabilize either the active or the inactive conformation, so for studying agonists one would need different antibodies or a different construct than for studying antagonists [62]. At the end of this chapter, the active and inactive conformations of β_2AR will be compared and discussed.

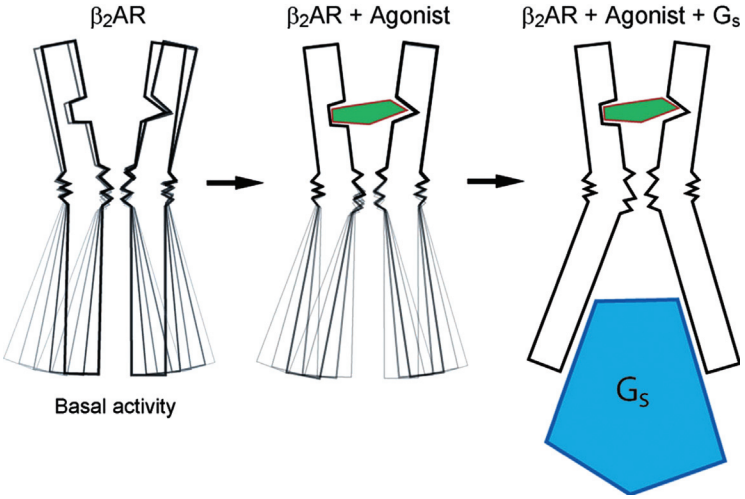


FIGURE 22.6 Cartoon illustrating the dynamic character of the β_2AR by Brian Kobilka [61]. Both the free and the agonist-bound form show a range of conformations. Only binding of the G-protein G_s fully stabilizes the receptor. © *The Nobel Foundation 2012.*

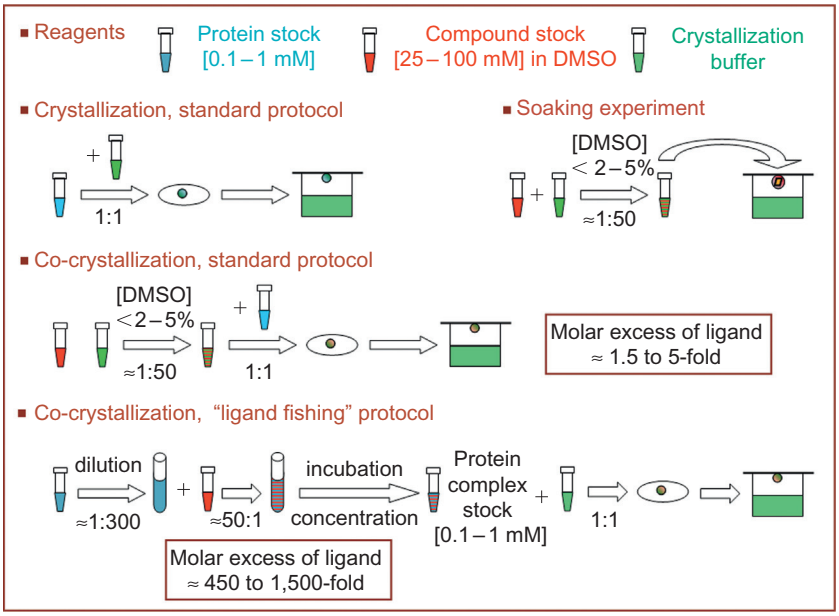


FIGURE 22.7 Protocols for preparing crystals of protein–ligand complexes.

4. *Preparation of Protein–Ligand Complexes*

An important aspect of protein crystallography in the context of drug design concerns the determination of protein–ligand complexes [63]. In many cases, getting crystals of protein–ligand complexes is not trivial, even if the apo protein has been crystallized. The ligand binding site may be blocked by crystal contacts, the ligand may not be soluble in the crystallization buffer, and the ligand may cause conformational changes in the protein that are incompatible with the crystal packing. Several methods exist to obtain crystals of protein–ligand complexes:

Soaking. This is the fastest and easiest way. The ligand is simply added to preformed crystals. If the ligand is a relatively small molecule, the solvent channels in protein crystals are usually large enough to allow the diffusion of the ligand to its binding site. A soaking experiment requires little material (one micromole of compound is usually plenty), but the solubility of the compound under the crystallization conditions is often an issue. The high protein content of the crystallization drop usually requires ligand concentrations in the range of 0.5–5.0 mM. A typical soaking protocol involves the preparation of a concentrated (50–100 mM) stock solution of the ligand, usually in a suitable organic solvent such as dimethylsulfoxide. This solution is then mixed with a crystallization buffer to a final concentration of solvent of up to 5 percent, and a few microliters of this mixture are added to the crystallization drop (Figure 22.7). Compound purity or the use of a diastereomeric mixture may

not be an issue if only one component in the sample binds to the protein. However, the chemical structure of the ligand must be known, since ultra-high resolution would otherwise be required for the unambiguous identification of an unknown binder.

Soaking has some practical advantages, particularly when the crystallization is not robust and good crystals are difficult to prepare. Overnight soaking is usually sufficient, allows a fast feed-back to modeling and chemistry, and opens up the possibility of using X-ray analyses for hit triaging and validation as well as for fragment-based screening. However, the soaking method also has drawbacks. Conformational changes induced by ligand binding may be hindered in the crystal, or access to the binding site may be restricted by protein–protein contacts. As a result, the ligand may not bind at all, or it may adopt an artificial mode of binding. Moreover, the diffraction quality can sometimes suffer from the soaking procedure, or the crystals may even crack or dissolve upon soaking. In such cases, gentle cross-linking of the crystals using the method of Lusty [64] may prove useful, but validation of the soaking approach with a co-crystallization experiment would then be worthwhile.

Co-crystallization. With this method, the ligand is added to the protein in solution, which is subsequently crystallized. When the crystallization is reasonably fast and robust, this is the method of choice and is recommended even in cases where soaking would be possible. The risk of artifacts is minimized, but at the expense of speed, particularly when the crystallization is very slow. A further disadvantage is that, for each and every new complex, crystallization conditions may have to be optimized again, or a full crystallization screening may be required, since crystallization conditions are sometimes very sensitive to changes in the ligand. Seeding is frequently used to accelerate co-crystallization experiments and improve their reproducibility. With modern robotics, this can be done automatically using very little protein [65].

Ligand fishing. Biological assays are usually performed in the presence of a large excess of ligand. This is particularly true for weak ligands, which are assayed at concentrations in the micromolar range while the protein concentration is typically in the nanomolar or sub-nanomolar range. Medicinal chemists should always bear in mind that weak biological activity may sometimes be due to trace amounts of a highly potent compound "contaminating" an otherwise inactive sample. For instance, an IC_{50} of 10 μM could be due to 0.5 percent of an impurity with a potency of 50 nM. This situation is not uncommon in programs where inactive derivatives are sometimes obtained from very potent precursors. Trace impurities of 1.0 percent or less are usually not detected by routine analytical techniques, but may give rise to apparent micromolar activity. These impurities will not be detected by crystallography either, if only a small excess of compound (2 to 5-fold) is added to a concentrated protein aliquot, as is usually the case (Figure 22.7). But if a very large excess of compound (say 500 to 1,500-fold) is added to a diluted sample of protein, there may be enough active impurity to saturate or nearly saturate the protein. The complex can then be concentrated using standard ultrafiltration techniques and crystallization experiments performed. This procedure, which is often referred to as "ligand fishing" (Figure 22.7), is more time-consuming and requires larger amounts of compound (5–10 mgs), but it has the ability to detect very low amounts (down to approximately 0.1 percent) of a potent ligand in a mixture. It may prove useful in cases where a weakly active compound whose structure is at odds with the established structure–activity data could not be observed using the routine crystallization procedure. In addition, this protocol can also be used when ligands are very poorly soluble in the crystallization buffer.

B. Data Collection

Protein crystals contain on average 50 percent solvent and—when exposed to air—they dry out and disintegrate. Moreover, when exposed to high-intensity X-rays at room temperature, they lose their diffraction power very quickly, owing to radiation damage. In order to prolong crystal lifetimes and improve data quality, X-ray measurements are routinely performed at 100 K [66]. Crystals are first mounted on 10–20 μm thin nylon loops and then flash-frozen by immersion into liquid nitrogen. To prevent the formation of ice crystals, it is often necessary to add a cryo-protectant such as glycerol, low molecular weight poly(ethylene glycol), or high salt to the surrounding mother liquor.

For data collection, the crystal is then placed on a goniometer, a device that controls the rotation of the crystal in the X-ray beam, while the temperature is kept at 100 K by blowing dry nitrogen over the crystal (Figure 22.8b). Large, strongly diffracting crystals can be measured in the lab with a rotating anode X-ray generator, but tiny or weakly diffracting crystals must be measured at a synchrotron source, such as the Swiss Light Source (SLS) in Villigen, Switzerland, or the European Synchrotron Radiation Facility (ESRF) in Grenoble, France (Figure 22.8a). For industrial projects, frozen crystals are often sent to the synchrotron by courier service in special cryo-containers. Data collection is remotely controlled by the scientists from their home laboratory or is done by a scientist at the synchrotron for a service fee.

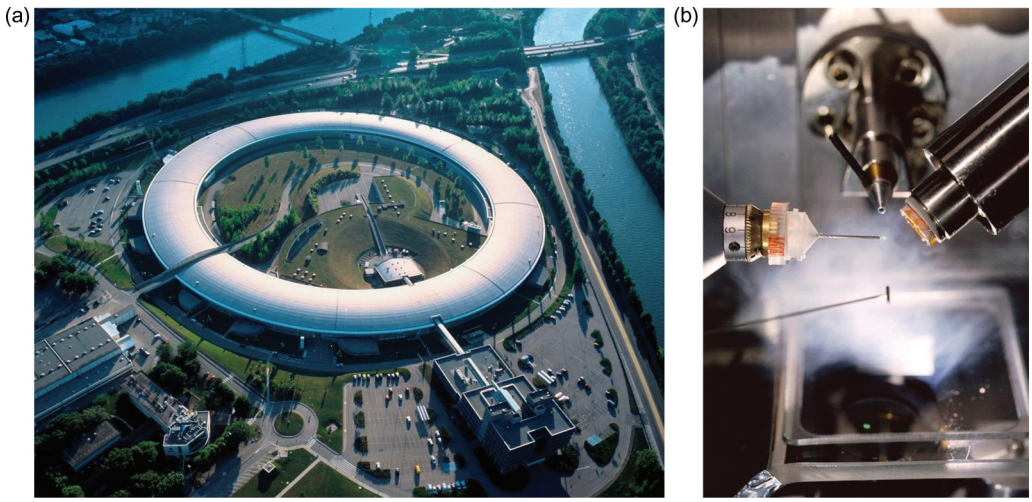


FIGURE 22.8 (a) Aerial view of the ESRF, located between the rivers Isère and Drac in Grenoble, France. Electrons circle around in a large ring inside the circular building, and when they pass bending magnets or assemblies of magnets called undulators they emit powerful X-rays. (b) Crystal being exposed, viewed from the position of the detector. X-rays emanating from the narrow steel tube in the back hit a frozen crystal in the cryo-loop in the center of the picture. The direct beam is stopped by a beam-stop, the small piece of metal just below the center of the picture. On the left is the goniometer, which is used to rotate the crystal, and from the nozzle on the right cold (100 k) nitrogen gas is blown over the crystal. *Pictures courtesy of ESRF/Morel.*

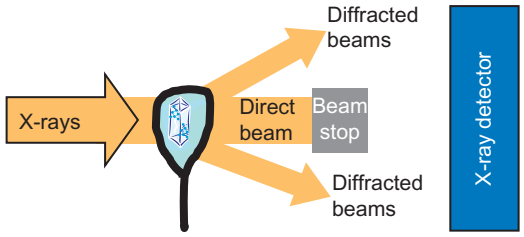


FIGURE 22.9 Schematic picture of an X-ray diffraction experiment. A real-life set-up is shown in Figure 22.8b. Monochromatic X-rays coming from the left hit the crystal—which is usually not longer than 100 μm—and some X-rays are diffracted. Most X-rays pass straight through and are stopped by a small piece of lead, the beam-stop. The diffracted X-rays are detected by a two-dimensional X-ray detector.

During data collection, the crystal is slowly rotated to bring all reflections into diffracting condition (see Box 22.2 about Bragg's law). The diffraction spots are usually recorded by CCD or hybrid-pixel detectors (Figure 22.9). The time to collect complete high quality X-ray data sets from single crystals ranges from a few minutes for decent crystals and high-intensity synchrotron radiation to a few days for weakly diffracting crystals and a conventional X-ray generator. The diffraction images from these detectors (Figure 22.10) are fed directly into a computer, which produces a list of reflection intensities. Ten thousand to several hundred thousand reflections are recorded per crystal, depending on the quality of the crystal and the size of the unit cell.

C. From Diffraction Intensities to a Molecular Structure

Light Microscopy and X-ray Crystallography Share the same Basic principle.

A light microscope allows us to study small objects like insects or cell slices in great detail, but it is physically impossible to resolve any details that are smaller than half the wavelength of the light used. For blue light, this limit is about 200 nm. To resolve atomic details, which are on the order of 1–5Å (0.1–0.5 nm), electromagnetic radiation with a much shorter wavelength than light is required (i.e., X-rays). A light microscope and an X-ray set-up share the same basic principle, although the practical implementation is quite different, owing to the different properties of X-rays and visible light.

In a microscope, light from a light source shines on the sample and is scattered in all directions. A set of lenses is used to reconstruct from this scattered light an enlarged image of the original sample. In an X-ray experiment,

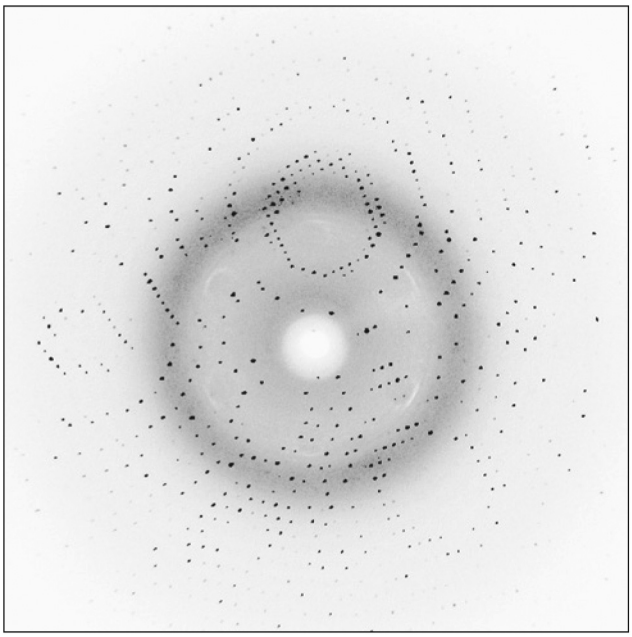
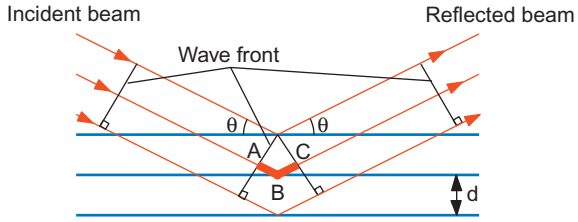


FIGURE 22.10 Example of an X-ray diffraction image.

BOX 22.2

BRAGGS LAW



Observable diffraction is only obtained when waves reflected from adjacent planes reinforce each other, that is, when the path difference ABC in the figure is an integer multiple n of the wavelength λ of the X-ray radiation used. This leads to Bragg's equation:

$$n\lambda = 2d \sin \theta$$

This means that to resolve closely spaced planes (small d), we need to measure high angle (large θ) reflections.

X-rays from an X-ray source hit the crystal and are scattered in all directions, just as with the light microscope. Unfortunately, no lenses can be made which are able to bring the scattered X-rays into focus to reconstruct an enlarged image of the sample. All the crystallographer can do is to record directly the scattered X-rays (the diffraction pattern; see Figure 22.10) and to use computers to reconstruct an enlarged image of the sample.

1. X-rays are Scattered by Electrons

Although X-rays interact only weakly with matter, they are occasionally absorbed by electrons, which start to oscillate. These oscillating electrons serve as X-ray sources that can radiate the X-ray wave in any direction. Waves scattered from different parts of the crystal have to add up constructively in order to produce a measurable intensity. The condition under which the scattered X-rays add up constructively is laid down in Bragg's law, which treats crystals in terms of sets of parallel planes (Box 22.2).

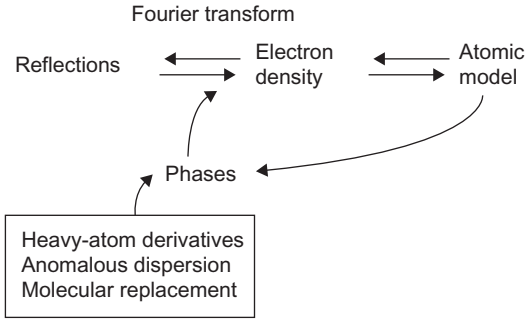


FIGURE 22.11 The phase problem. The experimental data obtained in an X-ray experiment are the intensities of the reflections. By using an inverse Fourier transform, it is possible to calculate electron-density maps from the amplitudes derived from these intensities. However, it is essential for this calculation to know the phase associated with each reflection. Approximate initial phases can be obtained from heavy-atom derivatives, anomalous dispersion, or molecular replacement (see text). More accurate phases can be derived from the refined model, once it has been obtained.

2. *The Diffraction Pattern Corresponds to the Fourier Transform of the Crystal Structure*

Each diffraction spot is caused by reflection of X-rays by a particular set of planes in the crystal. If the crystal contains layers of atoms with the same spacing and orientation as a particular set of planes that would satisfy Bragg's law (if the set of planes is physically present), the corresponding diffraction spot will be strong. On the other hand, if only few atoms in a crystal correspond to a particular set of planes, the corresponding reflection will be weak. The complicated structure present in the crystal is transformed by the diffraction process into a set of diffraction spots that correspond to sets of planes (more precisely, sinusoidal density waves), just as our ear converts a complicated sound signal into a series of (sinusoidal) tones when we listen to music. This conversion of a complicated function into a series of simple sine and cosine functions is called a Fourier transformation.

3. *The Phase Problem*

The original function—in our case the electron-density distribution in the crystal—can be reconstructed by performing the inverse Fourier transformation (i.e., by summing together the corresponding density waves for all reflections; see Figure 22.11). However, in order to make this summation, we need to know not only the amplitude of the density wave but also its relative position with respect to all other density waves (the phase). The amplitude, usually referred to as structure factor or F, can be measured because it is calculated from the intensity of the corresponding diffraction spot, but there is currently no practical way to measure the phases directly. This so-called "phase problem" can be solved by one of the following techniques:

Multiple isomorphous replacement (MIR). Crystals are soaked in solutions with "heavy" atom salts (Hg, Pt, Au, etc.), in the hope that a few heavy atoms will bind to some well-defined sites on the protein molecule. The heavy-atom positions are then found by analyzing the differences between the diffraction pattern of the native and of the soaked crystals. When two or more suitable heavy-atom derivatives are found, phase estimates and an electron-density map can be calculated.

Anomalous scattering (AS). This method makes use of the fact that some inner electrons of the heavier elements have absorption edges in the range of X-ray wavelengths. The method is used to supplement the phase information of a single heavy-atom derivative [67], but also to obtain full phase information from proteins which are labeled with selenomethionine, a selenium-containing amino acid [68]. This method (called "MAD" for multiple wavelength anomalous dispersion) has become the preferred method for the fast structure determination of novel proteins. Other anomalous methods have recently been proposed. For instance, the "halide-soak" approach uses short soaks in solutions containing 0.5−1.0M bromine or iodine, and the anomalous signal of the bound halide ions is then exploited to solve the structure [69]. For well-diffracting crystals, it is also possible to use the sulfur anomalous signal from the cysteines and methionines present in the native protein [70].

The MAD method is performed on a single crystal, but it requires access to tunable radiation (synchrotron source). Moreover, selenomethionine-labeled protein must be produced, purified, and crystallized. This is more easily done for proteins which can be expressed in *E. coli*.

Molecular replacement. When a suitable model of the unknown crystal structure is available, it can be used to solve the phase problem [71]. Examples are the use of the structure of human thrombin to solve the structure of bovine thrombin, the use of a known antibody fragment to solve the structure of an unknown antibody, or the

use of the structure of an enzyme to solve the structure of an inhibitor complex of the same enzyme in a different crystal form. The model is oriented and positioned in the unit cell of the unknown crystal with the use of rotation and translation functions, and the oriented model is subsequently used to calculate phases and an electron-density map.

Molecular replacement is usually straightforward and performed within minutes. However, when only low resolution data and a poor search model are available, model bias can become an issue and experimental phasing may be needed. Nevertheless, if a suitable model is present, which is increasingly likely given the ever increasing number of crystal structures available, molecular replacement is the method of choice to solve the phase problem.

4. Model Building and Refinement

Once a first electron-density map is obtained, it is interpreted by the crystallographer. In the case of a MIR (AS) map, a complete model of the protein has to be fitted to the electron density. The Cα atoms are placed first (chain tracing), and subsequently the complete main-chain and side-chains are built, a process that has become increasingly automated in recent years, particularly when high resolution data are available [72]. In the case of molecular replacement, the search model needs to be updated to reflect the molecule present in the crystal. The model is usually of a similar protein, and the possible changes include the substitution of some amino acids, the introduction of insertions and deletions, and the modification of some loops.

After the (re)building step, the model is refined. Refinement is an iterative procedure that aims at minimizing the differences between the observed diffraction amplitudes (F_o) and the diffraction amplitudes calculated from the model (F_c), while simultaneously optimizing the geometry of the structure. Because of the unfavorable ratio between observations and parameters, a free atom refinement is not possible in protein crystallography, and it is necessary to restrain the bond lengths, valence angles, and dihedral angles toward ideal values. Phases calculated from the refined model at the end of each refinement cycle are then used for the calculation of improved electron-density maps, which are again analyzed by the crystallographer to improve the model further. Cycles of refinement and rebuilding are repeated until convergence is reached. The final set of coordinates is then ready for deposition with the Protein Data Bank (PDB) [73].

5. Most Used Types of Electron-Density Maps

The direct experimental result of a crystallographic analysis is an electron-density map, while the model is derived from a (subjective) interpretation of this map. It is therefore useful to refer to the original data—the electron density—as often as possible. In the following paragraph, we will discuss the different types of electron-density maps most commonly used.

F_o-F_c or difference maps. These maps are obtained after subtracting the calculated structure factors (F_c) from the observed structure factors (F_o), an operation that is—in a first approximation—equivalent to subtracting the calculated electron density from the observed electron density. Features that are present in the "observed" density but not in the calculated density will give peaks, while atoms present in the model (in the F_c), but not in the "observed" electron density will result in holes (Figure 22.12). These maps are frequently used to detect errors in the model and can also be used to obtain an unbiased electron density of a bound inhibitor, for example, by completely removing the inhibitor from the model. In this case, the resulting electron density for the inhibitor is entirely caused by the experimental data and not by any model bias present in the phases. These maps are often referred to as "omit maps."

$2F_o$-F_c maps. These are the standard electron-density maps (Figure 22.12). Because of model bias, maps calculated with F_o and model phases tend to show only electron density associated with the model. As discussed above, F_o-F_c maps show everything that is in F_o but not in the model. By combining a F_o map with a F_o-F_c map, a $2F_o$-F_c electron-density map is obtained, which shows both electron density for the model and electron density for features that are not yet accounted for in the model, such as bound water molecules, carbohydrates, and other molecules associated with the protein. Several weighting schemes exist to minimize model bias. Examples are σ_A [74] and maximum-likelihood [75] weighting.

D. Information Content and Limitations of Protein Crystal Structures

Most chemists are familiar with X-ray analyses of small molecules, which are typically performed at a resolution better than 0.80Å. These subatomic resolution studies deliver highly accurate geometric parameters (bond lengths,

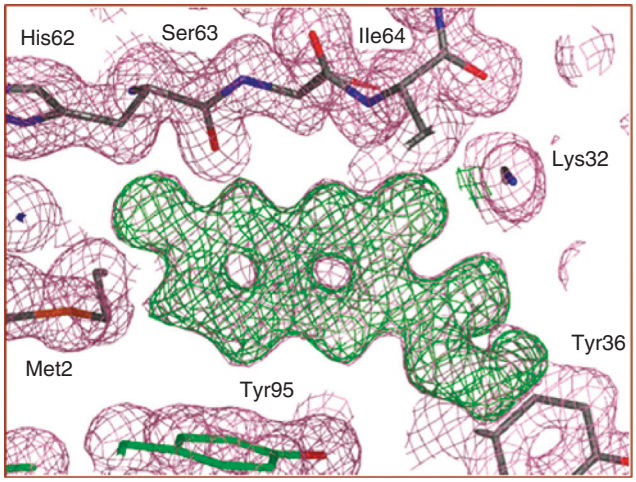


FIGURE 22.12 Close-up view of a protein complex at 1.5Å resolution showing the initial $2F_o$-F_c electron-density map (magenta mesh, 1.0σ contour), as well as the initial F_o-F_c map (green mesh, 3.0σ contour). The ligand has not yet been included in the model. Therefore, it appears as a strong positive difference density in the initial F_o-F_c map.

valence, and dihedral angles), as well as anisotropic displacement parameters ("temperature ellipsoids"). This is made possible by the very favorable observation to parameter ration (typically 50:1) resulting from the ultra-high resolution. Usually, protein crystals do not diffract to atomic or subatomic resolution. The vast majority of protein X-ray studies are performed at much lower resolution (between 3.0Å and 1.50Å), where this level of structural detail is not attainable. In particular, stereochemical parameters such as bond lengths and angles are restrained to standard dictionary values, both for the protein part and any low molecular weight ligand(s), prosthetic group, or post-translational modification. The protonation state and the exact orientation of some amino-acid side-chains (His, Asn, Gln) can only be inferred from potential H-bonded interactions. Substantially fewer solvent molecules and alternate conformations are observed than in the case of ultra-high resolution studies [76,77].

1. *Quality of the Experimental Data*

The quality of a crystal structure cannot be better than the quality of the experimental data upon which it is based. The following criteria are commonly used statistical indicators of the quality of the diffraction data:

Resolution. This corresponds to the shortest spacing of planes (d) whose reflections have been used in map calculation and refinement (see Box 22.2). The smaller this spacing, the sharper and more detailed the electron-density maps will be. The resolution is probably the single most important criterion determining the quality of a crystal structure. At high resolution (better than 2.0Å), the protein and bound water molecules are well defined, and it is unlikely that the structure will contain any serious errors. At low resolution (2.8−3.5Å), it is usually not possible to assign bound waters with certainty, and significant errors can remain unnoticed due to the problem of model bias.

Completeness of the data. One can calculate the total number of reflections to a certain resolution. Ideally, one would like to measure them all. For various reasons, however, it is in practice often not possible to measure all reflections. If only a small fraction of the reflections is missing (~10 percent), and the missing reflections are weak, the electron-density maps will hardly be affected. However, if a significant fraction of the reflections is missing, this may lead to artifacts in the electron-density maps, and the problem of model bias will become more severe.

R_{sym}. This reflects the inconsistency of multiple measurements of the same reflection. The lower the R_{sym}, the better. R_{syms} up to 15 percent are tolerable. Although the PDB still uses R_{sym}, other measures such as R_{meas} are more appropriate [78].

2. *Quality of the Model*

The global quality indicators listed below are commonly reported for refined crystal structures, but many more exist:

R-factor. This is a measure of the disagreement between the observed amplitudes (F_o) and the amplitudes calculated from the model (F_c). Depending on the resolution and quality of the diffraction data, well-refined structures have *R*-factors below 20−25 percent.

Free R-factor. Since refinement programs aim at minimizing the difference between observed and calculated amplitudes (hence the *R*-factor), an unbiased indicator is needed to monitor the progress of refinement. Brünger proposed excluding a subset of reflections from refinement and using these reflections only for the calculation of a "free" *R*-factor [79]. If refinement is progressing correctly, the free *R*-factor will drop as well. But if the model contains serious errors, it will remain stalled above ~35 percent. For correct structures, the free *R*-factor is generally below 30 percent.

Deviations from ideality of bond lengths and bond angles. A correctly fitted model is generally not strained. Significant deviations from ideal values for bond lengths and bond angles usually point to problems with the structure. Root-mean-square (r.m.s.) deviations from ideality should not be much larger than 0.02Å for bond lengths, and 3° for bond angles. The bond lengths and angles are biased toward the target values that are used during refinement. Accurate, unbiased values for these parameters can only be derived when ultra-high resolution (0.85Å or better) is available.

φ,ψ plot. Because of steric hindrance, only certain combinations of the main-chain dihedral angles φ and ψ are "allowed." The protein fold may force some residues to assume unallowed φ,ψ values, and this may have functional significance for some active site residues [80,81]. However, if more than a few percent of all the residues have φ,ψ values completely outside allowed regions, one should suspect errors.

3. Errors in Crystal Structures

Serious errors in crystal structures are rare and are usually associated with the first structure determination of a novel target, particularly when only low resolution data are available (3.0−5.0Å). Small errors and inaccuracies, however, are very common and virtually unavoidable. These errors are often underestimated, and small details of crystal structures are frequently overinterpreted by noncrystallographers. Medicinal chemists making use of crystal structures should be well aware of their limitations [82].

A major source of errors in macromolecular crystallography results from our inability to detect and model "disorder" appropriately [83], owing to the limited resolution and unfavorable parameter-to-observation ratio. Crystallographic refinement often attempts to fit a single model to some blurred electron density originating from several distinct but overlapping conformational states. This may lead to distorted geometry or to several distinct but equally valid interpretations.

A second important source of errors results from the fact that hydrogen atoms cannot be detected and atom types cannot be assigned at the resolution that is typically attainable with most protein crystals (1.5−3.0Å). This leads to ambiguities in the exact orientation of some groups, such as the side-chain amide of Asn and Gln residues or the imidazole ring of histidine side-chains.

Errors affecting the ligand. The exact orientation of one or more ligand groups can sometimes be uncertain. The choice of sensible geometric restraints for the refinement of nonstandard groups—in particular the ligands—is not always trivial and constitutes a potential source of errors [84]. For instance, the nitrogen atom of a tertiary amine bearing one aromatic substituent is usually planar, but it can also be pyramidal. At high resolution (better than 2.0Å), it may be possible to select the appropriate geometric restraints on the basis of the electron density. At lower resolutions, the refined model may mainly reflect the arbitrary choice of geometric restraints.

Errors affecting the solvent model. Water molecules are usually identified on the basis of residual electron-density peaks that meet certain criteria, such as the peak height, the distance, and the angle with respect to H-bond donor or acceptor groups. Since atom types and protonation states cannot be determined, a "water" may as well be a hydroxide, a hydroxonium, an ammonium, a sodium, or a magnesium ion. The assignment of metal ions becomes more reliable when the resolution of the data is good enough to reveal the coordination sphere or when the anomalous signal of the metal can be used.

4. Flexibility and Temperature Factors

Proteins are flexible molecules [85], and they usually retain a substantial degree of flexibility in the crystalline state. The mobility of the atoms in a crystal is expressed in terms of "temperature factors" or "B-factors," which are optimized during refinement. The relationship between mean total displacement and B-factors is given in Figure 22.13. The mean displacement of atoms with B-factors in excess of 60Å^2 is larger than 1.5Å, which is the length of a carbon−carbon bond. These atoms are generally poorly defined in the electron-density maps (Figure 22.14). For functional analysis, one should bear in mind that these flexible surface residues are either put in an arbitrary, low-energy conformation or deleted from the coordinate file. Not taking this into account could lead to serious artifacts, especially with electrostatic calculations.

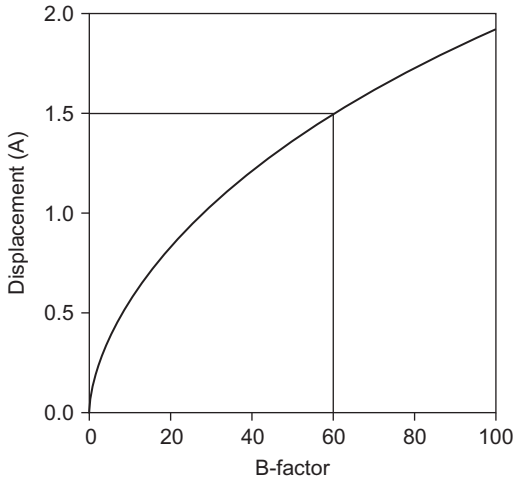


FIGURE 22.13 Relationship between mean total displacement and temperature factor B. At temperature factors of 60Å^2 and higher, the displacement becomes larger than 1.5Å and the electron density becomes very poor (see Figure 22.14). The formula used in the figure is derived from the relationship $B = 8\pi^2 < u^2 >$ where $< u^2 >$ represents the displacement perpendicular to the diffracting planes. The total mean square displacement $< u^2_{tot} > = 3 < u^2 >$, hence $< u_{tot} > = \sqrt{(3B/8\pi^2)}$.

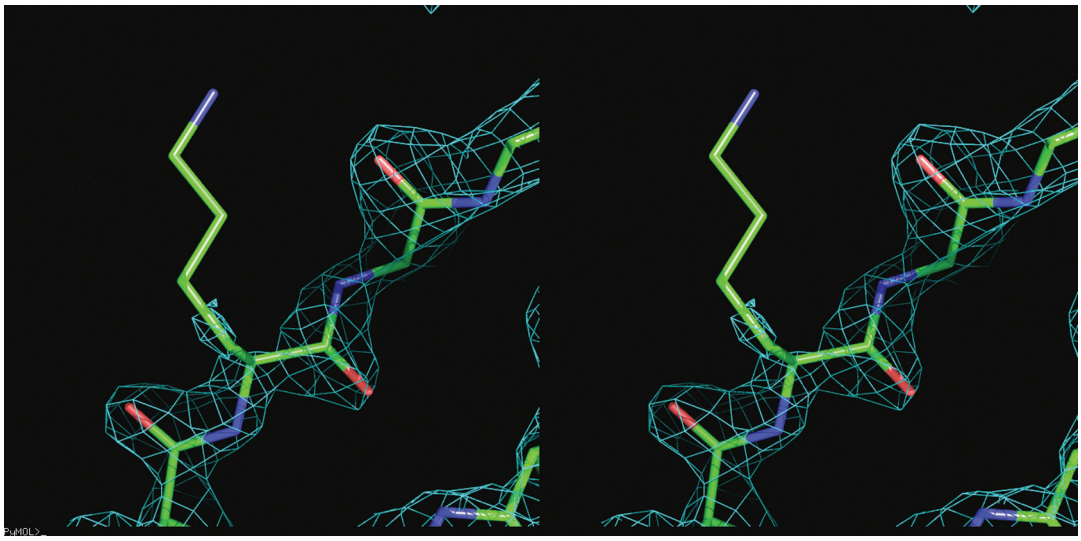


FIGURE 22.14 Long and flexible side-chains (such as Arg, Lys, Glu, and Gln) that are exposed to the solvent often move around freely. As a result, these side-chains have very high temperature factors, are very poorly or not at all defined in the electron-density maps, and are often fitted in an arbitrary, low-energy conformation. Lys87, located at the surface of human thrombin, is shown as an example. If one uses protein crystal structures for drug design, one should bear in mind that many exposed surface residues do not have a well-defined conformation.

5. *Misinterpretations of Electron-Density Maps*

Protein crystal structures are the result of a human interpretation of electron-density maps that are biased by the very model one is building. It is therefore no surprise that misinterpretations occur. Reasons for these errors include insufficient resolution or data quality, the presence of multiple overlapping binding modes, binding of several buffer components or fragments to the same site, and the lack of experience of the crystallographer.

False positives. False positives occur when a ligand is fitted to electron density belonging to bound solvent atoms, buffer components, or PEG molecules. Deleting the inhibitor and running a few rounds of refinement usually reveals this type of error clearly. Alternatively, one could run the Twilight script [86].

False negatives. There are two types of false negatives. With the first type, no bound ligand is found, although the ligand should bind according to biochemical and other binding assays [87]. In this case, crystal packing or the crystallization conditions (pH, high salt, high PEG) may prevent binding, or compound solubility may have

BOX 22.3

USING PDB FILES—TIPS AND TRICKS

Crystal structures are stored in formatted text files called "PDB" files. These files can be freely downloaded from the RCSB Protein Data Bank at http://www.pdb.org/. Information on how to search and navigate the PDB is available on the PDB home page.

Tip1: Always download a complete biological assembly. PDB files usually contain only the portion of the structure forming the asymmetric unit of the crystal. The asymmetric unit may include only a fraction of the functional biological molecule, for instance a single subunit of a homodimer. In such a case, information derived from viewing a single subunit may be very misleading, since binding sites or active sites are sometimes located at the interface between two or more subunits. Although a complete biological assembly can be generated with help of crystallographic software, it is also possible to download the corresponding file directly from the PDB (Download Files → Biological Assembly).

Tip2: Check all molecules of the asymmetric unit. Sometimes the asymmetric unit of the crystal contains several copies of the molecule or complex of interest. In such cases, individual copies of the biological assembly can be downloaded individually as separate files or all together in the original PDB file. It is very important to inspect them all, since significant differences can exist between these molecules due to different crystal contacts, disorder, partial occupancy of a ligand or co-factor, or as a consequence of different conformational states.

Tip3: Do not look only at the 3D model. Check the actual experimental information as well: the electron-density map. Electron-density maps contain more information than can possibly and accurately be included in an atomic model, even after careful refinement by an experienced crystallographer. For instance, some alternate conformations may not have been modeled (also note that many graphic programs ignore alternate conformations and do not display them). Some portions of a ligand molecule may be disordered, but coordinates for the complete molecule have been included. These and other important details can be revealed by displaying the electron-density map together with the atomic model. For structures which have been deposited with the PDB together with the corresponding diffraction data, electron-density map files can be downloaded from the Uppsala electron-density server [89] at http://eds.bmc.uu.se/eds/.

Tip4: Browse through your PDB file to find out more about its content. While the 3D structures encoded in the PDB files are best visualized using a graphic program (some interactive viewers are directly accessible from the PDB web pages), bear in mind that PDB files are simply text files that can also be displayed using a text editor. Browsing through PDB files can reveal some important information, notably on the method used to derive the structure (NMR, X-ray, or modeling), some data statistics, the amino-acid sequence with comments about engineered residues, the numbering of the protein residues and associated co-factors, ligands and solvent molecules, and more.

been too low to form enough complex. Hence, the failure to observe binding in an X-ray experiment does not necessarily disqualify a compound from being a genuine ligand. Before abandoning dubious but interesting hits, one should first verify them using other methods, such as protein NMR [88], mass spectroscopy, surface plasmon resonance, microcalorimetry, or thermophoresis.

In the second type of false negatives, compounds do bind but are not recognized as such. For example, at resolutions normally used in protein crystallography, the electron densities of ammonia, water, and sodium ions are virtually indistinguishable. Also bound buffer components and side products of the synthesis of the compound might not be recognized as such due to disorder or just because the exact chemical nature of the molecule is not known. In most cases, water molecules get fitted to these unknown densities.

IV. APPLICATIONS

A. Target Identification and Selection

1. Assignment of Function

Current low molecular-weight medicines exploit a fraction of all potential drug targets [90,91]. However, the large scale sequencing of whole genomes, including the human genome, has uncovered thousands of previously

unknown genes or "open reading frames." Functional annotation of these novel gene products is mainly based on sequence homologies to previously known proteins. For distant relatives, these homologies are often limited to a few short—but usually characteristic—sequence motifs. Tentative assignments are reinforced by sophisticated and powerful approaches, such as threading techniques, which verify the compatibility of a given amino-acid sequence with a 3D fold [92].

However, a substantial fraction of the novel genes code for proteins with no apparent relationship to any of the currently known ones. Structural genomic centers are solving the crystal structures of many of these novel proteins. In many cases the protein family and sometimes also the function can be deduced from the 3D structure, for example, from the presence of certain structural motifs with known catalytic functions like the Ser-His-Asp catalytic triad of serine proteases [93].

2. *Druggability (ligandability) Assessment*

While biology plays a key role in the selection of new targets, chemistry must have a strong say too, for pursuing a non-druggable target is a waste of time and resources. To get potent, selective, and orally active drugs, binding pockets with suitable properties must be present on the protein target. Potency and selectivity are usually achieved by optimizing the fit of the ligand to its receptor site, while oral bioavailability requires certain criteria to be met, such as Lipinski's "rule of five" [94]. Hence, a druggable target may be defined as a protein with a binding site of suitable size (that can accommodate compounds of MW $<$ 500Da), appropriate lipophilicity, and sufficient H-bonding potential [91], which can be deduced from crystal structures. Also, the presence of allosteric binding sites [95] and the existence of distinct structural conformations [85] can greatly increase the odds of finding a drug.

Most receptors and enzymes possess beautiful binding sites and are druggable [90,91]. In contrast, many protein−protein interaction sites are large and flat, and are therefore hopeless drug targets. However, because of their relevance to many diseases, protein−protein interactions are still attracting considerable interest, and a few may ultimately turn out to be druggable [96].

As a final remark, the methods mentioned above only estimate whether a ligand with a particular size and particular physiochemical properties is likely to bind to the target. They do not say whether such a compound will be a good drug that can be used to treat patients. For that reason, it might be more appropriate to speak of ligandability assessment instead of druggability assessment [97].

B. Hit/Lead Generation

1. *Structure-Based De Novo Drug Design*

The *de novo* design of novel scaffolds usually starts with careful scrutiny of multiple X-ray structures of the target in complex with a variety of ligands or tool compounds that in themselves are not attractive for chemical optimization due to issues with—for example—ease of synthesis and derivatization, intellectual property, and drug-likeness.

However, even the most promising templates designed with the most sophisticated computational tools are likely to have very weak potencies during the early stages of the *de novo* design process, so that standard biochemical assays may not be appropriate to evaluate these prototypic compounds. Protein crystallography may be of great help here, since in favorable cases it can detect high micromolar or even low millimolar binders. Once a first co-crystal structure with the designed template is obtained, subsequent optimization is usually straightforward.

2. *In Silico Screening*

Structure-based virtual screening, also called "high-throughput docking," involves the automatic docking and scoring of thousands of compounds to binding sites on protein targets [98−100]. Although the method has some shortcomings, like imperfect handling of receptor plasticity and reliability of scoring functions, its high-throughput and relatively low cost combined with its versatility outweigh these deficiencies. Most importantly, several recent success stories demonstrate that these methods do indeed deliver useful hits [101,102]. Obviously the effectiveness of high-throughput docking critically depends on the amount and quality of the structural information that is available for the drug target [102]. The outcome may further be improved by using target-based scoring and an expert system [103]. Particularly important is the understanding of the relevant conformational states and possible induced-fit mechanisms of the receptor binding site. Multiple co-crystal structures of the target of interest with different chemotypes, as well as any X-ray structures of related targets, contribute to this

understanding. Furthermore, it is essential that the most critical interaction sites or binding-site "hot spots" are identified [99]. Example of such key interaction sites include the hinge region of protein kinases and the flap and catalytic aspartates of aspartic proteinases. Sometimes, one or more conserved water molecules have been found to play an important role in ligand recognition and binding. Since the incorporation of such waters can strongly influence both the docking and the scoring steps, it is wise to search the available crystallographic data for the presence of conserved waters at critical locations within the receptor binding site [99,104]. Last but not least, it is also important to be aware of the limitations and uncertainties of crystal structures that can affect the virtual screening experiment. These uncertainties include the protonation state of protein residues and the exact orientation of some donor/acceptor groups, such as imidazole side-chains and the side-chain amide groups of asparagine and glutamine residues. Moreover, some important protein loops lining a binding site may not have well-defined electron density due to partial disorder (multiple conformations are present in the crystal), or the observed conformation may be influenced by the crystallization conditions or protein−protein contacts.

3. *Fragment-Based Screening*

While the two previous methods were computational (virtual), fragment-based screening (FBS) is an experimental method. Here, small fragments the size of a decorated benzene ring are screened for binding to the target protein. FBS emerged out of a need to overcome the current shortcomings of existing experimental or computational hit-finding approaches. The rationale behind the FBS strategy is well known: because the likelihood of a compound fitting a binding site decreases exponentially as the size increases [105], high-throughput screening approaches often fail to deliver hits, or they provide hits that are difficult to optimize owing to their low ligand efficiency [106,107] and "drug-like" rather than "lead-like" properties [108].

In contrast, the aim of FBS is to find hits that are easy to optimize by using a carefully selected fragment library [109]. Because of their small size and the fact that the entropic penalty associated with the loss of rigid-body translational and rotational freedom upon complex formation is independent of molecular weight [110,111], small fragments bind weakly, even when their ligand efficiency is high. Consequently, highly sensitive robust experimental techniques are needed to detect these weak binders. Historically, NMR has played a pioneering role in the development of FBS [9], but other technologies are applied as well, such as mass spectroscopy, surface plasmon resonance, and protein crystallography [100]. We hope that the reader will not take it amiss if we concentrate below on protein crystallographic applications to FBS. More general information on FBS can be found in [112]) or chapter 8 of this book.

To optimize FBS hits, it is essential to know their binding mode for a couple of reasons. First, the biophysical techniques used in FBS detect binding and not biochemical or biological activity and binding may be anywhere. Second, even if we know the binding site from competition experiments, modeling or docking of small fragments is usually ambiguous.

When one has access to a large compound store, it is often possible to dig out analogues with improved potency by substructure or similarity searches, and in this way to generate structure−activity data easily. For this reason, such an approach is often called "SAR by inventory." Nevertheless, in the absence of more detailed structural information, the optimization of weak FBS hits into potent leads can be a lengthy and cumbersome process.

A variety of NMR techniques exist to infer structural details on protein−ligand interactions [88]. However, protein crystallography remains the preferred approach for elucidating binding modes with certainty and guiding the hit-to-lead phase. Unfortunately, experience shows that only a fraction of FBS hits discovered by NMR or other biophysical techniques can be observed by protein crystallography.

Fragment-based screening by X-ray crystallography. When a suitable crystallization platform is available, one may consider using protein crystallography as the main FBS screening technique. Since crystallographic information is usually essential for the subsequent hit optimization, the use of X-ray analysis from the start can save time and certainly avoids the frustration of finding hits that cannot be reproduced later by crystallography.

Before an FBS by X-ray campaign can be launched, an initial investment in the preparation of suitable crystals may be needed [113], since crystals originally used for the first structure determination of a new drug target may not be suited. They may be difficult to grow or not diffract well enough, or the binding site of interest may be occupied by a strong ligand. For FBS by X-ray, it is essential that the crystals diffract to high resolution (better than 2.5Å, preferably 2.0Å or better) and are amenable to soaking, which implies that the targeted binding site is free and accessible. In cases where the crystallization is particularly robust, co-crystallization with the fragment cocktails can be attempted, but this strategy is usually less effective than the soaking approach. High crystal symmetry is not a must but makes data collection faster. When suitable crystals are not available, it may be necessary to engineer and produce new protein variants.

Fragments should be highly soluble under crystallization conditions. With typical protein concentrations in a crystallization experiments in the 0.1−1.0 mM range, fragments should be soluble up to concentrations of 1−10 mM. This is particularly critical when crystals are grown under high salt conditions. Apolar and aromatic scaffolds should feature one or more solubilizing group, such as a carboxylic or ammonium group. The risk that electrostatic interactions dominate binding is largely alleviated under high salt conditions, which strengthen hydrophobic interactions at the expense of the electrostatic ones.

With current technology, a library of 500 to 1,000 fragments split up into cocktails of five to ten compounds can be screened by X-ray crystallography within reasonable timelines. The cocktails should be designed in such a way that each component of a mixture has a distinct shape to allow unambiguous identification of any bound fragment on the basis of the shape of the electron density.

Over the past decade, FBS by X-ray has made notable contributions to the overall success of the fragment-based screening approach. It has provided novel, chemically attractive leads for some notoriously difficult targets, such as β-secretase [114,115], and these hits could be successfully optimized to highly potent drug candidates, hence fulfilling the initial promise of this approach.

4. *Triaging and Validation of HTS Hits*

Protein crystallography plays an important role in hit validation and selection for further optimization. Whereas before, only a few selected HTS hits could be analyzed crystallographically, high-throughput crystallography allows a more systematic approach. It is now possible to analyze many hits, including some of those that in the past would have been discarded. This analysis can provide highly valuable information regarding novel binding sites or subsites, alternative binding modes, privileged interaction patterns, and protein conformational substates. This information can then be fed into the structure-based design process, even when some of these hits are not pursued any further. Moreover, among the weak hits many compounds are often "fragment-like," with molecular weights in the 150−250 Da range. It may be of particular interest to investigate this region of chemical space where HTS meets FBS.

Finally, having a co-crystal structure of a hit bound to its target protein provides definitive experimental proof that the compound was not a false positive of some sort [116] and reassures the chemists that fast, structure-based optimization of the compound will be feasible.

C. Lead Optimization

Crystallographic information greatly enhances the speed and efficiency of lead optimization. However, crystal structures only show part of the picture and do not provide information on factors such as physicochemical properties, toxicity issues, metabolic weak points, thermodynamic parameters such as entropy and enthalpy, or protonation states of active site residues. For successful lead optimization, data obtained from many different sources need to be brought together.

1. *Optimizing Potency*

Large weakly binding compounds have very poor ligand efficiencies and are generally difficult to optimize. In most cases, fragments with a high ligand efficiency are much better starting points. The identification of key interaction sites (or "hot spots") within the binding pocket is a first and essential step when an enhancement in potency is sought [117]. To this end, an experimental fragment-based approach can be used [118] or computational methods can be utilized [42,119,120]. Interactions with the binding site hot spots should be maximized through the introduction of new substituents or the replacement of functionalities making sub-optimum contacts.

Protein ligands rarely bind in their lowest energy conformation [121]. When present, unfavorable strain energy should be detected and minimized. An analysis of the conformation of related compounds in the Cambridge data bank can guide this process [122], and *ab initio* calculations are often useful [122,123]. Compounds requiring minimal conformational reorganization on enzyme binding should be favored. Small-molecule ligands frequently adopt an extended conformation in the bound state [121]. Hence, hydrophobic ligands exhibiting a folded conformation in solution may incur a high reorganization energy cost on binding. Introduction of conformational restraints through (macro)cyclization [124] or the introduction of rigid linkers [125] is another strategy that has been successfully used in many cases to minimize entropic penalties.

2. *Optimizing Selectivity*

Exquisite selectivity can often be achieved by exploiting binding subsites or pockets adjacent to the main binding site that are not involved in the normal biological function of the drug target and are thus poorly conserved in other family members [126,127]. Likewise, taking advantage of the flexibility of the protein by targeting an unusual conformational state is an excellent means to achieving high selectivity. Protein kinase inhibitors provide numerous examples of this kind [128,129].

One disadvantage of the above approaches is that resistance mutations are more likely to emerge when nonfunctional states or cavities are used by the drug [127]. Designing more flexible compounds has also been proposed as a possible strategy for achieving broad-spectrum activity [130], since rigid molecules are less likely to adapt to structural changes. The entropic cost of the built-in flexibility needs to be compensated by a larger enthalpic contribution to binding through an optimization of all available polar interactions. Hence, enthalpic optimization of the binding affinity has been proposed as a better alternative to potency enhancement through hydrophobic binding and rigid fit [131]. In doing so, however, it is important to ensure that the strongest interactions involve residues with a low probability to mutate [132].

3. *Optimizing ADME Properties*

A general recipe for turning a potent lead into a real drug candidate does not exist, but some guidelines are available, such as the well-known "rule of five" [94]. Structure-guided design can aid in achieving the right balance between lipophilicity and polar surface area by guiding the introduction or replacement of heteroatoms, polar groups, and other solubilizing groups. Essential hydrogen-bonded interactions can be identified along with dispensable acceptor/donor groups. Minimizing molecular weight is frequently an effective strategy to achieving good oral bioavailability [122,133] and X-ray structures can identify groups and substituents that do not contribute much to binding and can therefore be replaced or removed. Suitable replacements for functionalities that are detrimental to ADME properties can be sought using focused chemical libraries or a FBS approach. Ultimately, it may be necessary to switch to a different chemotype. The availability of multiple X-ray structures of diverse molecular scaffolds may allow the combination of two different scaffolds into a new one [134], or the grafting of one particular motif from one inhibitor onto another [135,136].

V. TWO SELECTED EXAMPLES

A. Imatinib (Gleevec™)

The development of protein kinase inhibitors targeting the ATP binding site was initially received with great skepticism, on the grounds that it would not be possible to achieve a sufficient level of selectivity to turn them into useful therapeutic agents. In view of the large size of the human kinome [137] (518 genes) and the high conservation of the ATP binding site, this criticism was well founded. However, the discovery of imatinib [138,139] (Glivec®, Gleevec™), an inhibitor of the tyrosine kinase activity of the Bcr-Abl oncogene and an effective, frontline therapy for chronic myelogenous leukemia, provided compelling evidence for the viability of this approach.

The X-ray structure of the abl kinase domain in complex with des-methylpiperazinyl imatinib became available in 2000 [140], soon followed by the imatinib complex [141,142]. The N-methylpiperazine moiety of imatinib had been introduced during the lead-optimization phase to improve solubility, at a point in time where the exact binding mode of the drug was not known. Unexpectedly, the X-ray analyses revealed that the drug was binding to an inactive conformation of the kinase, with the benzamide and piperazinyl groups accessing a channel at the back of the ATP site (Figure 22.15). A conformational switch of the DFG motif of the kinase was responsible for the formation of this channel, which is therefore referred to as the "DFG-out" pocket. In this mode of binding, the N-methylpiperazine moiety was only partially exposed to solvent and strongly interacted with the kinase [127]. More importantly, several structural features of the inactive state of the abl kinase were important for imatinib binding, and detailed structural comparisons indicated that these features were poorly conserved in other protein kinases, thus explaining the high selectivity of this compound [127]. In addition, these and follow-up structures provided a platform for the analysis of resistance mutants [127,128]. The concept of DFG-in and DFG-out conformations has become a central theme in the search for kinase inhibitors.

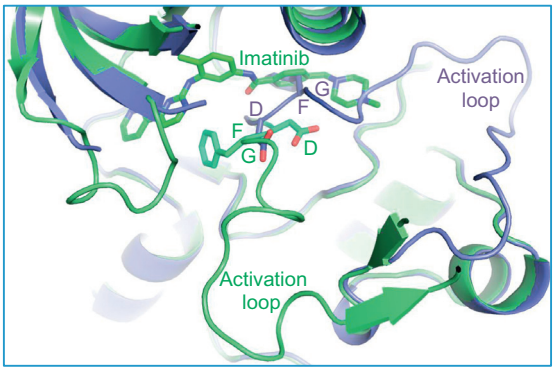


FIGURE 22.15 Close-up of a superposition of unligated abl kinase [127] (blue; pdb code 2hz4, chain A; DFG-in) and abl kinase complexed with imatinib [127] (green, pdb code 2hyy, chain A; DFG-out). In the DFG-in conformation, the position of the Phe of the DFG motif completely overlaps with the imatinib molecule bound to the DFG-out conformation. Also, the activations loops, which contain the DFG motif, assume completely different conformations.

B. The β_2-Adrenergic Receptor

The human genome contains at least 800 GPCRs that respond to a broad range of molecules and other entities, including photons, protons, odorants, neurotransmitters, hormones, and glycoproteins [61]. Since GPCRs are involved in most physiological processes, they represent the largest class of drug targets [58]. However, due to extreme difficulties in crystallizing them, very little structural information was available on GPCRs until recently. Structure-based drug design on GPCRs was mainly done using homology models, constructed on the basis of the rhodopsin crystal structure [143].

The β_2AR is not a drug target, but the closely related β_1AR is the target of beta blockers, a class of drugs widely used to treat heart patients. For that reason, a rich diversity of commercial ligands (full, partial, and inverse agonists, and neutral antagonists) was available. Using such compounds, Brian Kobilka and colleagues solved the mechanism of GPCRs by determining crystal structures of active and inactive states of β_2AR. For this work, Brian Kobilka—together with Robert Lefkowitz—was awarded the 2012 Nobel Prize in chemistry.

Crucial for this success was the stabilization of the inherently flexible GPCR (see Figure 22.6). The inactive state was crystallized using the potent inverse agonist carazolol and an antibody fragment binding the flexible loop between helices M5 and M6 [29], or, alternatively, by replacing this loop with the stable protein T4 lysozyme [144]. To crystallize the active state [145], the ultra-high affinity agonist BI-167107 from Boehringer-Ingelheim was used with either the G-protein mimicking nanobody [146] (single-chain camelid antibody fragment) Nb80 or a combination of inserted T4 lysozyme, Gs protein, and the nanobody Nb35.

Comparison of the inactive and active states (Figure 22.16) reveals that small differences in the ligand binding pocket due to antagonist or agonist binding are amplified via a repacking of Ile121, Pro211, Phe282, and Asn318 in the core of the β_2AR molecule. A rotation of helix TM6 results in a 14Å outward movement of the tip of this helix and causes conformational changes in the associated G-protein, ultimately resulting in the exchange of GDP by GTP and activation of the G-protein.

VI. OUTLOOK

A sequence-based search in the January 2014 release of the PDB with the sequence of the β_1-adrenergic receptor yielded 101 GPCR structures corresponding to thirty-one unique sequences. This number is rapidly increasing, paving the way for true structure-based GPCR drug discovery.

Finally, the construction of free electron lasers all over the world may again radically change the way protein crystallography is done. By using extremely short (10–200 fs) and extremely bright ($>10^{12}$ photons) X-ray pulses, it is possible to record useful X-ray diffraction before the atoms in the crystal have had time to move (i.e., before any radiation damage could occur). In a test experiment, a full 8.5Å data set was recently collected from nanocrystals of Photosystem I [147], a large membrane protein complex with a molecular mass of 1 MDa, 36 proteins, and 381 co-factors. Diffraction was observed from crystals smaller than ten unit cells on a side. Being able to use nanocrystals would remove one of the biggest hurdles in protein crystallography, namely the preparation of large, well-diffracting crystals.

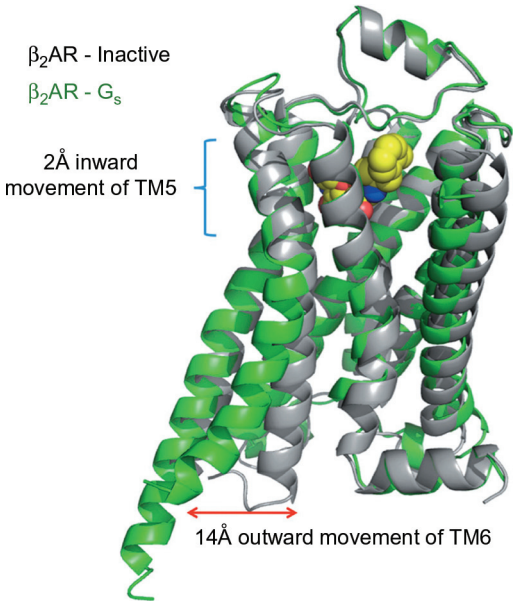


FIGURE 22.16 A comparison of the carazolol-bound, inactive-state structure of the β_2AR [144] (grey) and the active-state structure of the β_2AR (green) from the β_2AR–Gs complex [145]. *Figure by Brian Kobilka [61].* © *The Nobel Foundation 2012.*

References

[1] Shakespeare W. Macbeth. Paris: Editions Aubier Montaign; 1977.
[2] Congreve M, Murray CW, Blundell TL. Keynote review: structural biology and drug discovery. Drug Discov Today 2005;10(13):895–907.
[3] Scapin G. Structural biology and drug discovery. Curr Pharm Des 2006;12(17):2087–97.
[4] Tintelnot-Blomley M, Lewis RA. A critical appraisal of structure-based drug design. IDrugs 2006;9(2):114–18.
[5] Lombardino JG, Lowe JA. The role of the medicinal chemist in drug discovery—then and now. Nat Rev Drug Discov 2004;3(10):853–62.
[6] Williams SP, Kuyper LF, Pearce KH. Recent applications of protein crystallography and structure-guided drug design. Curr Opin Chem Biol 2005;9(4):371–80.
[7] Weigelt J. Structural genomics: impact on biomedicine and drug discovery. Exp Cell Res 2010;316(8):1332–8.
[8] Blundell TL, Jhoti H, Abell C. High-throughput crystallography for lead discovery in drug design. Nat Rev Drug Discov 2002;1(1):45–54.
[9] Hajduk PJ, Greer J. A decade of fragment-based drug design: strategic advances and lessons learned. Nat Rev Drug Discov 2007;6 (3):211–19.
[10] Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by X-Ray analysis. Nature 1958;181(4610):662–6.
[11] Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North ACT. Structure of haemoglobin: a three-dimensional fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis. Nature 1960;185(4711):416–22.
[12] McPherson A. A brief history of protein crystal growth. J Cryst Growth 1991;110(1–2):1–10.
[13] Franklin RE, Gosling RG. Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. Nature 1953;172(4369):156–7.
[14] Wilkins MHF, Seeds WE, Stokes AR, Wilson HR. Helical structure of crystalline deoxypentose nucleic acid. Nature 1953;172(4382): 759–62.
[15] Watson JD, Crick FHC. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. Nature 1953;171(4356):737–8.
[16] Perutz M. Protein structure: new approaches to disease and therapy. New York: WH Freeman and Company; 1992.
[17] Rossmann MG, Arnold E, Erickson JW, Frankenberger EA, Griffith JP, Hecht H-J, et al. Structure of a human common cold virus and functional relationship to other picornaviruses. Nature 1985;317(6033):145–53.
[18] Deisenhofer J, Epp O, Miki K, Huber R, Michel H. Structure of the protein subunits in the photosynthetic reaction centre of Rhodopseudomonas viridis at 3Å resolution. Nature 1985;318(6047):618–24.
[19] Abrahams JP, Leslie AGW, Lutter R, Walker JE. Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria. Nature 1994;370(6491):621–8.
[20] Lowe J, Stock D, Jap B, Zwickl P, Baumeister W, Huber R. Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution. Science 1995;268(5210):533–9.
[21] Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8A resolution. Nature 1997;389(6648):251–60.
[22] Brodersen DE, Clemons Jr WM, Carter AP, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V. The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. Cell 2000;103(7):1143–54.
[23] Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Science 2000;289(5481):905–20.
[24] Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. The structural basis of ribosome activity in peptide bond synthesis. Science 2000;289 (5481):920–30.

[25] Pioletti M, Schlünzen F, Harms J, Zarivach R, Glühmann M, Avila H, et al. Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine, and IF3. EMBO J 2001;20(8):1829–39.
[26] Cramer P, Bushnell DA, Kornberg RD. Structural basis of transcription: RNA polymerase II at 2.8 Ångstrom resolution. Science 2001;292(5523):1863–76.
[27] Doyle DA, Cabral JM, Pfuetzner RA, Kuo A, Gulbis JM, Cohen SL, et al. The structure of the potassium channel: molecular Basis of K+ conduction and selectivity. Science 1998;280(5360):69–77.
[28] Jiang Y, Lee A, Chen J, Cadene M, Chait BT, MacKinnon R. Crystal structure and mechanism of a calcium-gated potassium channel. Nature 2002;417(6888):515–22.
[29] Rasmussen SGF, Choi H-J, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, et al. Crystal structure of the human [bgr]2 adrenergic G-protein-coupled receptor. Nature 2007;450(7168):383–7.
[30] Kroemer M, Dreyer MK, Wendt KU. APRV: a program for automated data processing, refinement, and visualization. Acta Crystallogr D 2004;60(9):1679–82.
[31] Mooij WTM, Hartshorn MJ, Tickle IJ, Sharff AJ, Verdonk ML, Jhoti H. Automated protein–ligand crystallography for structure-based drug design. Chem Med Chem 2006;1(8):827–38.
[32] Evrard GX, Langer GG, Perrakis A, Lamzin VS. Assessment of automatic ligand building in ARP/wARP. Acta Crystallogr D 2007;63(1):108–17.
[33] Yumerefendi H, Tarendeau F, Mas PJ, Hart DJ. ESPRIT: an automated, library-based method for mapping and soluble expression of protein domains from challenging targets. J Struct Biol 2010;172(1):66–74.
[34] Beddell CR, Goodford PJ, Norrington FE, Wilkinson S, Wootton R. Compounds designed to fit a site of known structure in human haemoglobin. Br J Pharmacol 1976;57(2):201–9.
[35] Ondetti M, Rubin B, Cushman D. Design of specific inhibitors of angiotensin-converting enzyme: new class of orally active antihypertensive agents. Science 1977;196(4288):441–4.
[36] Natesh R, Schwager SLU, Evans HR, Sturrock ED, Acharya KR. Structural details on the binding of antihypertensive drugs captopril and enalaprilat to human testicular angiotensin I-converting enzyme. Biochemistry 2004;43(27):8718–24.
[37] Eriksson AE, Jones TA, Liljas A. Refined structure of human carbonic anhydrase II at 2.0 Å resolution. Proteins: structure. Funct Bioinform 1988;4(4):274–82.
[38] Baldwin JJ, Ponticello GS, Anderson PS, Christy ME, Murcko MA, Randall WC, et al. Thienothiopyran-2-sulfonamides: novel topically active carbonic anhydrase inhibitors for the treatment of glaucoma. J Med Chem 1989;32(12):2510–13.
[39] Wade RC. 'Flu' and structure-based drug design. Structure 1997;5(9):1139–45.
[40] Varghese JN, Laver WG, Colman PM. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. Nature 1983;303(5912):35–40.
[41] Colman PM, Varghese JN, Laver WG. Structure of the catalytic and antigenic sites in influenza virus neuraminidase. Nature 1983;303(5912):41–4.
[42] Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J Med Chem 1985;28(7):849–57.
[43] von Itzstein M, Wu W-Y, Kok GB, Pegg MS, Dyason JC, Jin B, et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. Nature 1993;363(6428):418–23.
[44] Kim CU, Lew W, Williams MA, Liu H, Zhang L, Swaminathan S, et al. Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: Design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. J Am Chem Soc 1997;119(4):681–90.
[45] Matthews BW. Solvent content of protein crystals. J Mol Biol 1968;33(2):491–7.
[46] Ducruix A, Giegé R. Crystallization of nucleic acids and proteins: a practical approach. 2nd ed. Oxford: Oxford University Press; 1993.
[47] McPherson A. Crystallization of biological macromolecules. Cold Spring Harbor: Laboratory Press; 1999.
[48] Kundrot CE. Which strategy for a protein crystallization project? Cell Mol Life Sci 2004;61(5):525–36.
[49] Pusey ML, Liu Z-J, Tempel W, Praissman J, Lin D, Wang B-C, et al. Life in the fast lane for protein crystallization and X-ray crystallography. Prog Biophys Mol Biol 2005;88(3):359–86.
[50] Derewenda ZS. The use of recombinant methods and molecular engineering in protein crystallization. Methods 2004;34(3):354–63.
[51] Wright PE, Dyson HJ. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. J Mol Biol 1999;293(2):321–31.
[52] McPherson A, Cudney B. Searching for silver bullets: an alternative strategy for crystallizing macromolecules. J Struct Biol 2006;156(3):387–406.
[53] Chang VT, Crispin M, Aricescu AR, Harvey DJ, Nettleship JE, Fennelly JA, et al. Glycoprotein structural genomics: solving the glycosylation problem. Structure 2007;15(3):267–73.
[54] Strauss A, Fendrich G, Horisberger MA, Liebetanz J, Meyhack B, Schlaeppi J-M, et al. Improved expression of kinases in Baculovirus-infected insect cells upon addition of specific kinase inhibitors to the culture helpful for structural studies. Protein Expr Purif 2007;56(2):167–76.
[55] Brading RL, Abbott WM, Green I, Davies A, McCall EJ. Co-expression of protein phosphatases in insect cells affects phosphorylation status and expression levels of proteins. Protein Expr Purif 2012;83(2):217–25.
[56] Wiener MC. A pedestrian guide to membrane protein crystallization. Methods 2004;34(3):364–72.
[57] Minor Jr DL. The neurobiologist's guide to structural biology: a primer on why macromolecular structure matters and how to evaluate structural data. Neuron 2007;54(4):511–33.
[58] Yong K, Scott D. Engineering stabilised G protein-coupled receptors for biochemical and structural studies. Aust Biochem 2013;44(2):7–12.
[59] Faham S, Bowie JU. Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. J Mol Biol 2002;316(1):1–6.
[60] Caffrey M. Membrane protein crystallization. J Struct Biol 2003;142(1):108–32.

[61] Kobilka B. The structural basis of G-protein-coupled receptor signaling (Nobel Lecture). Angew Chem Int Ed 2013;52(25):6380–8.
[62] Congreve M, Rich RL, Myszka DG, Figaroa F, Siegal G, Marshall FH. Fragment screening of stabilized G-protein-coupled receptors using biophysical methods N.p. In: Lawrence CK, editor. Methods in enzymology. Academic Press; 2011. p. 115–36
[63] Hassell AM, An G, Bledsoe RK, Bynum JM, Carter HL, Deng S-JJ, et al. Crystallization of protein–ligand complexes. Acta Crystallogr D 2007;63(1):72–9.
[64] Lusty CJ. A gentle vapor-diffusion technique for cross-linking of protein crystals for cryocrystallography. J Appl Crystallogr 1999;32 (1):106–12.
[65] D'Arcy A, Villard F, Marsh M. An automated microseed matrix-screening method for protein crystallization. Acta Crystallogr D 2007;63 (4):550–4.
[66] Garman EF, Owen RL. Cryocooling and radiation damage in macromolecular crystallography. Acta Crystallogr D 2006;62(1):32–47.
[67] Dauter Z, Dauter M, Dodson E. Jolly SAD. Acta Crystallogr D 2002;58(3):494–506.
[68] Ealick SE. Advances in multiple wavelength anomalous diffraction crystallography. Curr Opin Chem Biol 2000;4(5):495–9.
[69] Dauter Z, Dauter M. Entering a new phase: using solvent halide ions in protein structure determination. Structure 2001;9(2):R21–6.
[70] Debreczeni JÉ, Bunkóczi G, Ma Q, Blaser H, Sheldrick GM. In-house measurement of the sulfur anomalous signal and its use for phasing. Acta Crystallogr D 2003;59(4):688–96.
[71] Vagin A, Teplyakov A. MOLREP: an automated program for molecular replacement. J Appl Crystallogr 1997;30(6):1022–5.
[72] Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D 2006;62(9):1002–11.
[73] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. Nucleic Acids Res 2000;28(1):235–42.
[74] Read RJ. Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Crystallogr A 1986;42(3):140–9.
[75] McCoy AJ. Liking likelihood. Acta Crystallogr D 2004;60(12–1):2169–83.
[76] Petrova T, Podjarny A. Protein crystallography at subatomic resolution. Rep Prog Phys 2004;67(9):1565.
[77] Howard EI, Sanishvili R, Cachau RE, Mitschler A, Chevrier B, Barth P, et al. Ultrahigh resolution drug design. I. Details of interactions in human aldose reductase-inhibitor complex at 0.66 Å. Proteins: structure. Funct Bioinform 2004;55(4):792–804.
[78] Diederichs K, Karplus PA. Improved R-factors for diffraction data analysis in macromolecular crystallography. Nat Struct Biol 1997;4 (4):269–75.
[79] Brunger AT. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature 1992;355(6359):472–5.
[80] Jia Z, Vandonselaar M, Quail JW, Delbaere LTJ. Active-centre torsion-angle strain revealed in 1.6 A-resolution structure of histidine-containing phosphocarrier protein. Nature 1993;361(6407):94–7.
[81] Chevrier B, Schalk C, D'Orchymont H, Rondeau J-M, Moras D, Tarnus C. Crystal structure of Aeromonas proteolytica aminopeptidase: a prototypical member of the co-catalytic zinc enzyme family. Structure 1994;2(4):283–91.
[82] Davis AM, Teague SJ, Kleywegt GJ. Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. Angew Chem Int Ed 2003;42(24):2718–36.
[83] DePristo MA, de Bakker PIW, Blundell TL. Heterogeneity and inaccuracy in protein structures solved by X-Ray crystallography. Structure 2004;12(5):831–8.
[84] Kleywegt GJ. Crystallographic refinement of ligand complexes. Acta Crystallogr D 2007;63(1):94–100.
[85] Teague SJ. Implications of protein flexibility for drug discovery. Nat Rev Drug Discov 2003;2(7):527–41. Available from: http://dx.doi.org/10.1038/nrd1129.
[86] Weichenberger CX, Pozharski E, Rupp B. Visualizing ligand molecules in twilight electron density. Acta Crystallogr F 2013;69 (2):195–200.
[87] Chung C-W. The use of biophysical methods increases success in obtaining liganded crystal structures. Acta Crystallogr D 2007;63(1): 62–71.
[88] Jahnke W, Widmer H. Protein NMR in biomedical research. Cell Mol Life Sci 2004;61(5):580–99.
[89] Kleywegt GJ, Harris MR, Zou J-Y, Taylor TC, Wählby A, Jones TA. The uppsala electron-density server. Acta Crystallogr D 2004;60 (12–1):2240–9.
[90] Overington JP, Al-Lazikani B, Hopkins AL. How many drug targets are there? Nat Rev Drug Discov 2006;5(12):993–6.
[91] Hopkins AL, Groom CR. The druggable genome. Nat Rev Drug Discov 2002;1(9):727–30.
[92] Blundell TL, Sibanda BL, Montalvão RW, Brewerton S, Chelliah V, Worth CL, et al. Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. Philos Trans R Soc B Biol Sci 2006;361(1467):413–23.
[93] Laskowski RA, Watson JD, Thornton JM. Protein function prediction using local 3D templates. J Mol Biol 2005;351(3):614–26.
[94] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 1997;23(1–3):3–25.
[95] Hardy JA, Wells JA. Searching for new allosteric sites in enzymes. Curr Opin Struct Biol 2004;14(6):706–15.
[96] Chène P. Drugs targeting protein–protein interactions. Chem Med Chem 2006;1(4):400–11.
[97] Edfeldt FNB, Folmer RHA, Breeze AL. Fragment screening to predict druggability (ligandability) and lead discovery success. Drug Discov Today 2011;16(7–8):284–7.
[98] Shoichet BK. Virtual screening of chemical libraries. Nature 2004;432(7019):862–5.
[99] Klebe G. Virtual ligand screening: strategies, perspectives, and limitations. Drug Discov Today 2006;11(13–14):580–94.
[100] Rees DC, Congreve M, Murray CW, Carr R. Fragment-based lead discovery. Nat Rev Drug Discov 2004;3(8):660–72.
[101] Cavasotto CN, Orry AJ. Ligand docking and structure-based virtual screening in drug discovery. Curr Top Med Chem 2007;7 (10):1006–14.
[102] Alvarez JC. High-throughput docking as a source of novel drug leads. Curr Opin Chem Biol 2004;8(4):365–70.
[103] Jansen JM, Martin EJ. Target-biased scoring approaches and expert systems in structure-based virtual screening. Curr Opin Chem Biol 2004;8(4):359–64.
[104] Zentgraf M, Steuber H, Koch C, La Motta C, Sartini S, Sotriffer CA, et al. How reliable are current docking approaches for structure-based drug design? Lessons from aldose reductase. Angew Chem Int Ed 2007;46(19):3575–8.

[105] Hann MM, Leach AR, Harper G. Molecular complexity and its impact on the probability of finding leads for drug discovery. J Chem Inf Comput Sci 2001;41(3):856–64.
[106] Hopkins AL, Groom CR, Alex A. Ligand efficiency: a useful metric for lead selection. Drug Discov Today 2004;9(10):430–1.
[107] Abad-Zapatero C. Ligand efficiency indices for effective drug discovery. Expert Opin Drug Discov 2007;2(4):469–88.
[108] Teague SJ, Davis AM, Leeson PD, Oprea T. The design of leadlike combinatorial libraries. Angew Chem Int Ed 1999;38(24):3743–8.
[109] Congreve M, Carr R, Murray C, Jhoti HA. "Rule of Three" for fragment-based lead discovery?. Drug Discov Today 2003;8(19):876–7.
[110] Finkelstein AV, Janin J. The price of lost freedom: entropy of bimolecular complex formation. Protein Eng 1989;3(1):1–3.
[111] Murray C, Verdonk M. The consequences of translational and rotational entropy lost by small molecules on binding to proteins. J Comput Aided Mol Des 2002;16(10):741–53.
[112] Jahnke W, Erlanson DA, editors. Fragment-based approaches in drug discovery. Weinheim: Wiley-VCH; 2006.
[113] Nienaber V, Wang J, Davidson D, Henkin J. Re-engineering of human urokinase provides a system for structure-based drug design at high resolution and reveals a novel structural subsite. J Biol Chem 2000;275(10):7239–48.
[114] Murray CW, Callaghan O, Chessari G, Cleasby A, Congreve M, Frederickson M, et al. Application of fragment screening by X-ray crystallography to β-secretase†. J Med Chem 2007;50(6):1116–23.
[115] Congreve M, Aharony D, Albert J, Callaghan O, Campbell J, Carr RAE, et al. Application of fragment screening by X-ray crystallography to the discovery of aminopyridines as inhibitors of β-secretase§. J Med Chem 2007;50(6):1124–32.
[116] Böcker A, Bonneau PR, Edwards PJ. HTS Promiscuity analyses for accelerating decision making. J Biomol Screening 2011;16(7):765–74.
[117] Brenk R, Klebe G. "Hot spot" analysis of protein-binding sites as a prerequisite for structure-based virtual screening and lead optimization. In: Langer T, Hoffmann RD, editors. Pharmacophores and pharmacophore searches, vol. 32. Weinheim: Wiley-VCH; 2006. p. 171–92.
[118] Ciulli A, Williams G, Smith AG, Blundell TL, Abell C. Probing hot spots at protein − ligand binding sites: A fragment-based approach using biophysical methods. J Med Chem 2006;49(16):4992–5000.
[119] Gohlke H, Hendlich M, Klebe G. Predicting binding modes, binding affinities and "hot spots" for protein–ligand complexes using a knowledge-based scoring function. Perspect Drug Discov Des 2000;20(1):115–44.
[120] Verdonk ML, Cole JC, Watson P, Gillet V, Willett P. Superstar: improved knowledge-based interaction fields for protein binding sites. J Mol Biol 2001;307(3):841–59.
[121] Perola E, Charifson PS. Conformational analysis of drug-like molecules bound to proteins: An extensive study of ligand reorganization upon binding. J Med Chem 2004;47(10):2499–510.
[122] Kim EE, Baker CT, Dwyer MD, Murcko MA, Rao BG, Tung RD, et al. Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. J Am Chem Soc 1995;117(3):1181–2.
[123] Greer J, Erickson JW, Baldwin JJ, Varney MD. Application of the three-dimensional structures of protein target molecules in structure-based drug design. J Med Chem 1994;37(8):1035–54.
[124] MacPherson LJ, Bayburt EK, Capparelli MP, Bohacek RS, Clarke FH, Ghai RD, et al. Design and synthesis of an orally active macrocyclic neutral endopeptidase 24.11 inhibitor. J Med Chem 1993;36(24):3821–8.
[125] Kim S-H, Pudzianowski AT, Leavitt KJ, Barbosa J, McDonnell PA, Metzler WJ, et al. Structure-based design of potent and selective inhibitors of collagenase-3 (MMP-13). Bioorg Med Chem Lett 2005;15(4):1101–6.
[126] Rahuel J, Rasetti V, Maibaum J, Rüeger H, Göschke R, Cohen NC, et al. Structure-based drug design: the discovery of novel nonpeptide orally active inhibitors of human renin. Chem Biol 2000;7(7):493–504.
[127] Cowan-Jacob SW, Fendrich G, Floersheimer A, Furet P, Liebetanz J, Rummel G, et al. Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukaemia. Acta Crystallogr D 2007;63(1):80–93.
[128] Noble MEM, Endicott JA, Johnson LN. Protein kinase inhibitors: insights into drug design from structure. Science 2004;303 (5665):1800–5.
[129] Cowan-Jacob SW. Structural biology of protein tyrosine kinases. Cell Mol Life Sci 2006;63(22):2608–25.
[130] Das K, Clark AD, Lewi PJ, Heeres J, de Jonge MR, Koymans LMH, et al. Roles of conformational and positional adaptability in structure-based design of TMC125-R165335 (Etravirine) and related non-nucleoside reverse transcriptase inhibitors that are highly potent and effective against wild-type and drug-resistant HIV-1 variants. J Med Chem 2004;47(10):2550–60.
[131] Velazquez-Campoy A, Todd MJ, Freire E. HIV-1 Protease inhibitors: Enthalpic versus entropic optimization of the binding affinity. Biochemistry 2000;39(9):2201–7.
[132] Ohtaka H, Velázquez-Campoy A, Xie D, Freire E. Overcoming drug resistance in HIV-1 chemotherapy: the binding thermodynamics of Amprenavir and TMC-126 to wild-type and drug-resistant mutants of the HIV-1 protease. Protein Sci 2002;11(8):1908–16.
[133] Brown FJ, Andisik DW, Bernstein PR, Bryant CB, Ceccarelli C, Damewood Jr JR, et al. Design of orally active, non-peptidic inhibitors of human leukocyte elastase. J Med Chem 1994;37(9):1259–61.
[134] Pierce AC, Rao G, Bemis GW. BREED: Generating novel inhibitors through hybridization of known ligands. Application to CDK2, P38, and HIV protease. J Med Chem 2004;47(11):2768–75.
[135] Anderson M, Beattie JF, Breault GA, Breed J, Byth KF, Culshaw JD, et al. Imidazo[1,2-a]pyridines: a potent and selective class of cyclin-dependent kinase inhibitors identified through structure-based hybridisation. Bioorg Med Chem Lett 2003;13(18):3021–6.
[136] Terasaka T, Kinoshita T, Kuno M, Nakanishi I. A highly potent non-nucleoside adenosine deaminase inhibitor: efficient drug discovery by intentional lead hybridization. J Am Chem Soc 2003;126(1):34–5.
[137] Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science 2002;298 (5600):1912–34.
[138] Zimmermann J, Buchdunger E, Mett H, Meyer T, Lydon NB. Potent and selective inhibitors of the Abl-kinase: phenylamino-pyrimidine (PAP) derivatives. Bioorg Med Chem Lett 1997;7(2):187–92.
[139] Capdeville R, Buchdunger E, Zimmermann J, Matter A. Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. Nat Rev Drug Discov 2002;1(7):493–502.

[140] Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, Kuriyan J. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. Science 2000;289(5486):1938–42.
[141] Nagar B, Bornmann WG, Pellicena P, Schindler T, Veach DR, Miller WT, et al. Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and Imatinib (STI-571). Cancer Res 2002;62(15):4236–43.
[142] Manley PW, Cowan-Jacob SW, Buchdunger E, Fabbro D, Fendrich G, Furet P, et al. Imatinib: a selective tyrosine kinase inhibitor. Eur J Cancer 2002;38:S19–27.
[143] Bourdon H, Trumpp-Kallmeyer S, Schreuder H, Hoflack J, Hibert M, Wermuth C-G. Modelling of the binding site of the human m1 muscarinic receptor: experimental validation and refinement. J Comput Aided Mol Des 1997;11(4):317–32.
[144] Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, et al. High-resolution crystal structure of an engineered human β2-adrenergic G protein–coupled receptor. Science 2007;318(5854):1258–65.
[145] Rasmussen SGF, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, et al. Crystal structure of the [bgr]2 adrenergic receptor-Gs protein complex. Nature 2011;477(7366):549–55.
[146] Muyldermans S. Nanobodies: natural single-domain antibodies. Annu Rev Biochem 2013;82(1):775–97.
[147] Chapman HN, Fromme P, Barty A, White TA, Kirian RA, Aquila A, et al. Femtosecond X-ray protein nanocrystallography. Nature 2011; 470(7332):73–7.