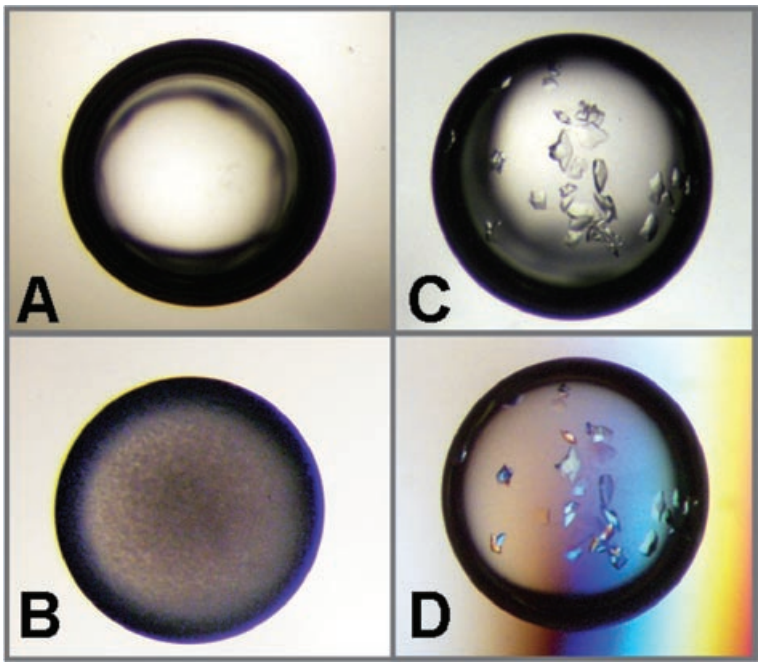

Figure 10.4 Images of hanging drop crystallization experiments. (*a*) Clear 1-μL drop at the outset of the crystallization experiment. (*b*) Precipitate. (*c*) Crystals of lysozyme inside a hanging drop. (*d*) Hanging drop with birefringent lysozyme crystals, imaged under cross-polarization setting. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

acid linkers were employed [24]. Alternatively, these tags may be removed with specific proteolytic enzymes that cleave appropriately engineered linker sites between the tag and the host protein. Once the protein is purified and concentrated, it may be stored via rapid freezing in liquid nitrogen. A cooling procedure proved beneficial that employs protein solution volumes below 50 μL and 0.2 mL ultrathin-walled polymerase chain reaction (PCR) tubes [25]. Fast thawing to room temperature was critical in order to prevent precipitation.

Besides precipitant type and concentration, further factors that affect crystallization include temperature, buffer type and concentration, the presence of additives, crystallization format, geometry, and other environmental parameters. It is not possible to screen all of these factors systematically. Several hundred crystallization experiments are therefore usually carried out varying the temperature (4 and 20°C) and formulations of precipitating agents, while all other parameters are kept constant. Subsequent fine screening may then be accomplished by systematically screening other factors. Rather exotic factors such as electric and magnetic fields have been identified to affect crystal quality. Their systematic use in the crystallography laboratory, however, is limited.

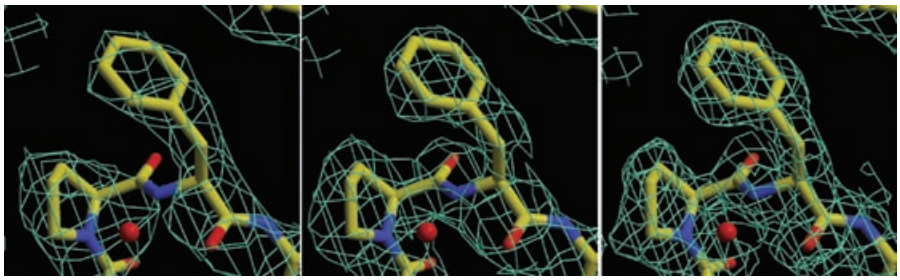

Figure 10.12 Electron density maps calculated for a protein at increasing resolutions: *(Left)* 3.5 Å, *(middle)* 2.8 Å, and *(right)* 2.1 Å. The electron density is shown in cyan chicken wire. Electron density is usually displayed as chicken wire contour lines around the protein model. The chicken wire contours correspond to ordered electron density in the crystal structure, and should therefore be superimposable on the model of the protein. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

general tendency for the intensities of the spots to weaken as their positions move farther from the center of the image. For any given protein crystal, the higher resolution reflections are more difficult to collect, if not unattainable, compared with the lower resolution reflections.

Similar to the example of the square wave reconstruction in Figure 10.9, adding more, higher resolution, reflection terms to the electron density Fourier transform [Eq. (10.1)], gives a more accurate representation of the electron density for the target protein. Figure 10.12 shows the effects of using higher resolution reflections in the calculation of electron density maps. At lower resolutions, the electron density gives the general shape of the target protein's electron density. At higher resolutions, the electron density reveals more details of the target protein structure.

X-ray Diffraction Data Collection in Practice

The collection of protein X-ray diffraction data requires three basic hardware components: X-ray source, goniometer, and X-ray detector. There are numerous configurations of these components and additional X-ray optical components that can be used. Figure 10.13 shows a schematic of an X-ray diffraction data collection system. An intense, monochromatic X-ray beam produced by an X-ray source strikes the sample protein crystal. The X-rays diffracted by the crystal are measured by an X-ray detector, on the opposite side of the crystal as the X-ray source. The protein crystal is mounted on a goniometer, which allows for very precise angular adjustments of the crystal position. During a data collection, the crystal will be rotated about one or more axes by the goniometer. Usually, the crystal is rotated about the ϕ axis, which is

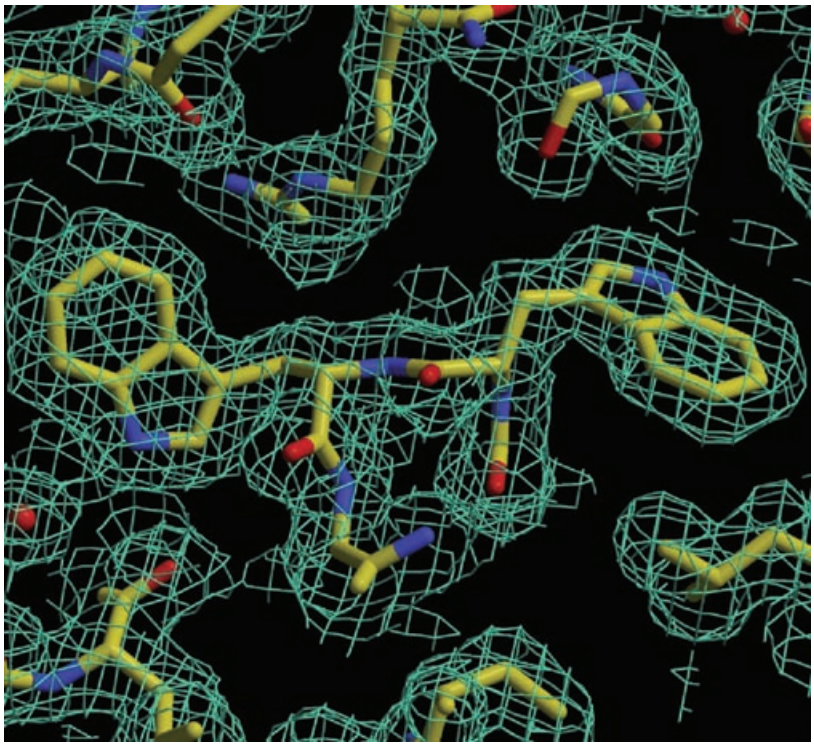

Figure 10.17 A $2F_o$–F_c electron density map calculated according to Eq. (10.13). (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

mined for the measured reflections, incorporating these phases into the Fourier summation for the electron density [Eq. (10.3)] will give an electron density map, into which the model of the protein can be built.

After determination of the phases and building in the protein model, electron density maps are generally calculated by Fourier transforms using coefficients $2F_o - F_c$. The explicit expression for the electron density map calculation is [52]:

$$\rho(x, y, z) = (1/V)\sum_{hkl}(2|F_o| - |F_c|)\cos[2\pi(hx + ky + lz) - \alpha(hkl)] \qquad (10.13)$$

In Eq. (10.13), $|F_o|$ and $|F_c|$ are, respectively, the amplitudes of the reflections from the experimental diffraction data and the structure model built into the electron density, and α is the calculated phase for the reflection with Miller indices *hkl*. This type of electron density map shows the electron density of the calculated model, and the difference electron density of the target protein structure and the calculated model [52]. An example of an electron density map calculated according to Eq. (10.13) is shown in Figure 10.17. The electron density is usually represented in chicken-wire contours into which the protein model can be built.

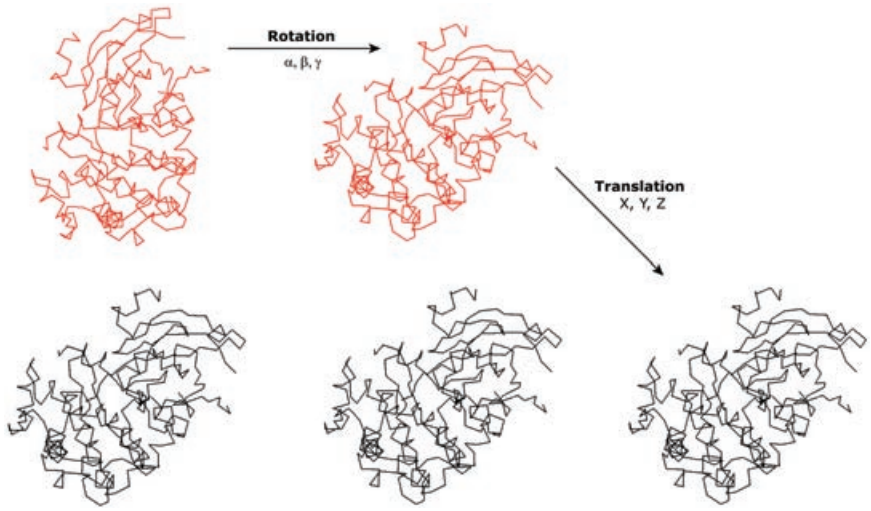

Figure 10.19 Schematic of an MR structure determination. (left) The target protein structure is drawn in black lines. The search model protein structure in an arbitrary orientation and position with respect to the target protein is drawn in red lines. (middle) The search model is rotated so that its orientation corresponds to that of the target protein structure as indicated by the intramolecular Patterson vectors in the experimental diffraction data. (right) The search model is translated so that it is correctly placed in the unit cell of the experimental diffraction data. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

After the rotational search, the correct placement of the oriented search model in the target protein crystal unit cell is determined by a translational search. In the translational search, the oriented search model is translated within the unit cell of the crystal until the crystal packing of the search model matches that of the experimental diffraction data. There are various methods for the translational search. In one method, the longer intermolecular Patterson vectors are compared with those of the experimental data at each translation of the search model. The highest correlation between the sets of intermolecular Patterson vectors is generally the correct solution for the crystal structure of the target protein.

Compared with isomorphous replacement and anomalous dispersion methods, crystal structure determination by MR has much less experimental requirements. No special derivatization of the sample crystal with heavy atoms or anomalous scatterers is required. Only a single data set from the crystal of the target protein is required. Also, no particular radiation wavelength is required for data collection. The ability to use diffraction data collected at any wavelength makes structure determination by MR readily accessible since all of the necessary diffraction data can be collected on a standard laboratory X-ray source (assuming adequate diffraction power of the crystals).

At this point, one may wonder why go through the effort of determining the "crystal" structure of the target protein by MR if the "protein" structure

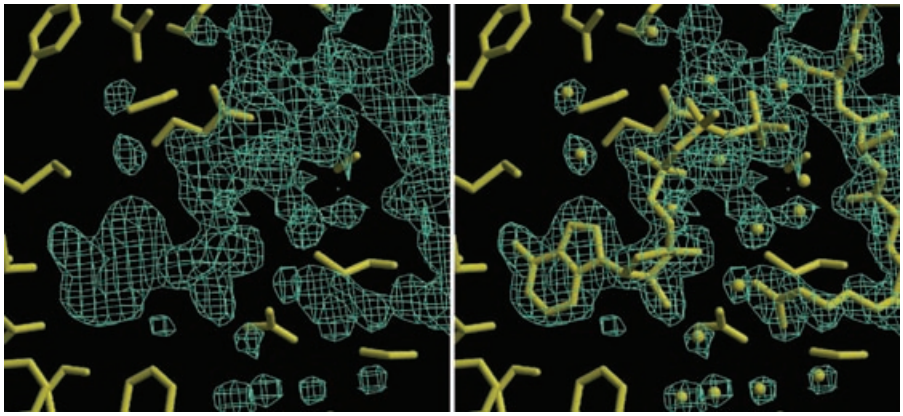

Figure 10.20 F_o–F_c electron density map calculated from an MR structure determination. *(Left)* The structure of a protein–ligand complex has been determined by MR using the protein-only structure as the search model. The F_o–F_c electron density map calculated from the protein-only MR solution reveals electron density (blue chicken wire contours) which does not correspond to the search model structure (yellow bonds). *(Right)* The structure of the bound ligand is built into the difference electron density in the F_o–F_c electron density map. Additional protein residues not present in the search model and bound waters are also built into the difference electron density. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

is already known or assumed to be that of the search model. MR crystal structure determination is required because the search model may not be complete. Although the overall structure of the target protein may already be known, the search model will lack certain structural features of the experimental data, such as amino acid mutations, insertions, and deletions, and the binding modes of ligands bound to the target protein. These additional features are not incorporated into the search model during MR calculations. The electron densities of these additional structures can only be calculated after determining the correct crystal structure of the target protein, and using phases determined from the correct placement of the protein in the crystal.

Once the search model has been correctly rotated and translated with respect to the diffraction data, difference electron density maps, using Fourier coefficients $F_o - F_c$, reveal structures that were not part of the search model. The explicit expression for this type of electron density map is

$$\rho(x, y, z) = (1/V)\sum_{hkl}(|F_o| - |F_c|)\cos[2\pi(hx + ky + lz) - \alpha(hkl)] \qquad (10.16)$$

Compare this expression with that of the $2F_o - F_c$ electron density map [Eq. (10.13)]. The $F_o - F_c$ will show electron density in regions where the search model and the target protein structures differ and no electron density where the two structures are the same. Figure 10.20 illustrates the MR structure determination of a protein–compound complex. The experimental diffraction

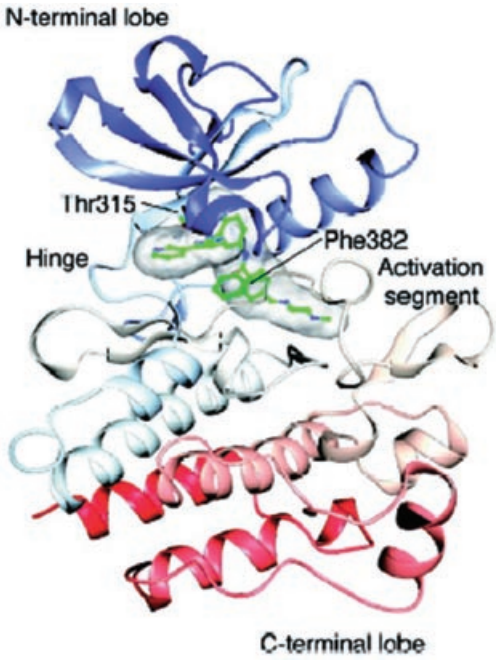

Figure 10.26 Overview of the structure of a kinase with an inhibitor bound. Kinases offer the possibility to design inhibitors based on stabilization of inactive conformations. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

Antistructures

As discussed, the main applications of X-ray crystallography in drug discovery and optimization projects are based on the analysis of the target and the interactions of targets with their ligands. The design efforts are aimed at strengthening the resulting complexes. The opposite approach, however, weakening the interaction of drug leads with some proteins may be used to one's advantage. The interacting proteins are not the actual targets but those that cause detrimental effects in drug efficacy. Such proteins are sometimes called *antitargets* and may be related enzymes with similar substrate binding pockets but with very different function, such as kinases or phosphatases.

Generally, weak binding of small-molecule drugs to serum albumin and detoxification proteins such as cytochrome P450 is a desired property. P450 enzymes are involved in the oxidative metabolization of most drugs and are often the source of drug-related side effects or their toxicity. Several P450 structures are available [73] and may be used for in silico docking studies and the published crystallization methods may be used to grow crystals for soaking or co-crystallization studies. The goal of such projects is to increase lead effi-

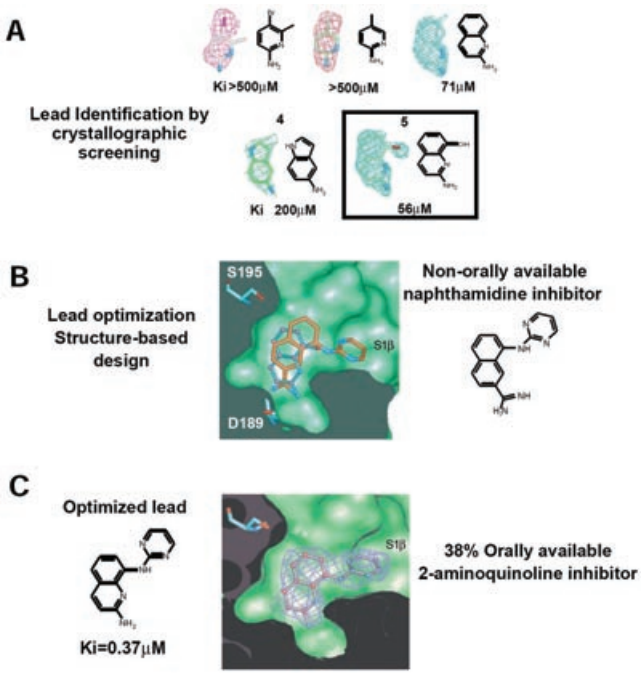

Figure 10.27 Urokinase lead identification via crystallographic screening and optimization [1]. (*a*) Initial F_o–F_c electron density maps for ligands that were identified from compound cocktail-soaked urokinase crystals. (*b*) Crystal structures of 8-aminopyrimidyl-2-naphtamidine (orange) and a 2-aminoquinoline lead (blue). (*c*) Structure and $2F_o$–F_c electron density map for the optimized lead compound 8-aminopyrimidyl-2-aminoquinoline. (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

included and lead to the development of 8-aminopyrimidyl-2-aminoquinoline, a ligand with a ca. 100-fold increased inhibitor potency (K_i = 0.37 μM) and a 38 percent oral availability, as determined by in vivo pharmacokinetic tests. This type of process is capable of identifying weaker binding ligands (1 mM) and is applicable where apo-crystals are available and tolerate soaking. Crystallographic screening may also be used to facilitate the validation of new targets, the development of assays and assist in assigning biochemical function to orphan targets.

Crystallographic Fragment Screening

A variation of this theme is crystallographic fragment screening. Here crystals are soaked with cocktails that contain small druglike fragments rather than complete leadlike compounds. Once several fragments are identified crystal-

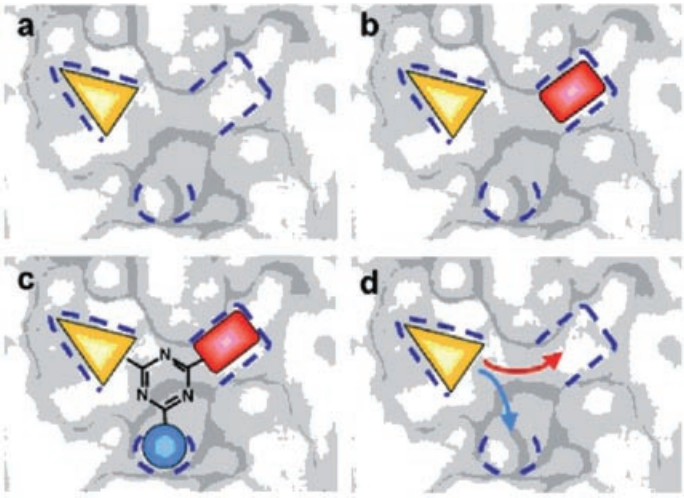

Figure 10.28 Schematic crystallographic fragment screening. Once fragments are identified (*a*, *b*) they can be joined (*c*) resulting in a leadlike compound or fragments may be developed along the lines of conventional structure-based drug design. (Figure taken from Blundell et al., 2002 [80].) (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

lographically, they can be developed into new lead compounds (Fig. 10.28). Curiously, low-affinity small fragments that bind adjacent binding pockets can be joined and result in a larger molecule with increased affinity. Typically fragment libraries consisting of only a few hundred to a thousand compounds are screened. Crystallographic fragment screening is a new and promising technology employed by several biotechnology companies; however, specific examples for drug discovery have not yet been published in the scientific literature.

Rees et al.[81] discuss 25 examples for the successful application of the fragment-based lead discovery approach, some of them aided by crystallographic screening. They also formulate a "Rule of three" in which the average fragment is characterized as (a) having a mass of less than 300 Da, (b) having less than or equal to 3 hydrogen bond donors, (c) having less than or equal to 3 hydrogen bond acceptors, and (d) having a *c* log *P* of 3. In addition, the number of rotatable bonds was on average less than or equal to 3 and the polar surface area was about 60 $Å^2$.

Site-Directed Leads via Fragment Tethering

An additional layer of complexity is added by generating site-directed leads via fragment tethering ([2]; Fig. 10.29). In a first step target proteins are covalently modified at a particular site on the surface. Mass-spectrometric detection allows the identification of weakly binding ligand precursors. In a second

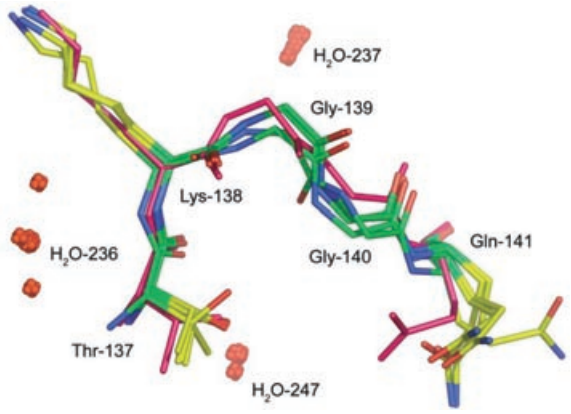

Figure 10.31 Structural heterogeneity in human interleukin1β. The ensemble of models displays considerable backbone variability, disordered side chains, and multiple locations of water molecules. The models were obtained from the same set of 2.3 Å resolution X-ray diffraction data, and refined to similar levels. (Image taken from DePristo et al. [87].) (This figure is available in full color at ftp://ftp.wiley.com/public/sci_tech_med/drug_discovery/.)

stabilizes a particular conformational state of the pool of many low-energy states that proteins can exist in at thermal equilibrium. In the process of SBDD this causes a high degree of unpredictability, making the method less useful. However, some proteins are less flexible, and their conformation hardly changes when ligands bind. These are the targets that are particularly susceptible to conventional SBDD efforts. There are only a few solutions to this fundamental predicament [89], notably the computationally intensive approach to treat the protein as a flexible entity and the tethering discovery approach. Understanding this limitation, however, may serve as the best antidote against the overuse of this tool.

Sanders et al. [90] point out a serious shortcoming of crystallographic screening. They described the discovery of competitive inhibitors for dihydroneopterin aldolase via crystallographic screening and demonstrated that several compounds with IC_{50} around 1 μM were negative in crystal soaking experiments. Apparently the conformational shift associated with the binding of these missed compounds did not allow association to the protein in the preformed crystal.

The deficiencies of current computational methods to properly quantify the interactions of proteins with ligands is one of the consequences of molecular flexibility. But even more fundamentally, our current understanding of the energetics of ligand–protein interaction and hence their proper quantification by scoring functions is limited [91]. A weak point remains, for instance, in the description of entropic terms for binding interactions, although progress is being made and an energetic penalty of 10 to 30 kcal/mol is estimated for protein reorganization due to binding [92].