Diffraction Techniques in Structural Biology

Martin Egli[1]

[1]Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, Tennessee

ABSTRACT

A detailed understanding of chemical and biological function and the mechanisms underlying the molecular activities ultimately requires atomic-resolution structural data. Diffraction-based techniques such as single-crystal X-ray crystallography, electron microscopy, and neutron diffraction are well established and they have paved the road to the stunning successes of modern-day structural biology. The major advances achieved in the last 20 years in all aspects of structural research, including sample preparation, crystallization, the construction of synchrotron and spallation sources, phasing approaches, and high-speed computing and visualization, now provide specialists and nonspecialists alike with a steady flow of molecular images of unprecedented detail. The present unit combines a general overview of diffraction methods with a detailed description of the process of a single-crystal X-ray structure determination experiment, from chemical synthesis or expression to phasing and refinement, analysis, and quality control. For novices it may serve as a stepping-stone to more in-depth treatises of the individual topics. Readers relying on structural information for interpreting functional data may find it a useful consumer guide. *Curr. Protoc. Nucleic Acid Chem.* 41:7.13.1-7.13.35. © 2010 by John Wiley & Sons, Inc.

Keywords: structural biology ● diffraction ● crystallization ● electron microscopy ● neutron scattering ● X-ray crystallography

INTRODUCTION

There are numerous approaches that furnish insight into the conformational properties of biopolymers such as proteins and nucleic acids. Among these, diffraction-based techniques occupy a unique place due to the atomic-resolution picture that they can reveal. Thus, provided a single crystal of a receptor, virus, or RNA diffracts X-rays to very high resolution, conformation, molecular interactions, and water structure can be visualized in stunning detail. A few selected examples of recent successes in the crystallographic structure determination of macromolecular assemblies, receptors, molecular machines, and viruses are depicted in Figure 7.13.1.

In the last decade, we have witnessed an unprecedented increase in the number of new crystal structures. Online databases such as the Research Collaboratory for Structural Biology/Protein Data Bank (RCSB/PDB; *http://www.rcsb.org*; 64,781 structures as of April, 20, 2010; Berman et al., 2000) and the Nucleic Acid Database (NDB; *http://ndbserver.rutgers.edu*; 4704 structures deposited as of April 14, 2010; Berman et al., 1992) now boast large numbers of entries.

Before long, the number of new PDB entries per year may surpass 10,000. Indeed, with the advent of structural genomics, the old adage that structure determination is preceded by a thorough understanding of function has given way to structure-driven initiatives that promise insights into function from structure (i.e., the Protein Structure Initiative funded by the U.S. National Institutes of Health: *http://www.nigms.nih.gov/Initiatives/PSI/*; Chandonia and Brenner, 2006; Terwilliger et al., 2009).

For some 100 hundred years, diffraction techniques have shaped our perception of the structure of condensed matter. An overview of the Nobel prizes awarded to scientists behind discoveries related to diffraction and their application to physics, chemistry, biology and medicine provides evidence for the wide-ranging scientific impact of diffraction phenomena (Table 7.13.1). The explosive growth in the number of crystal structures during the last years followed dramatic advances in practically all areas of X-ray crystallography, including crystallization (sparse matrix screens and robotics; Jancarik and Kim, 1991; Doudna et al., 1993; Scott et al., 1995), crystal handling (flash freezing;
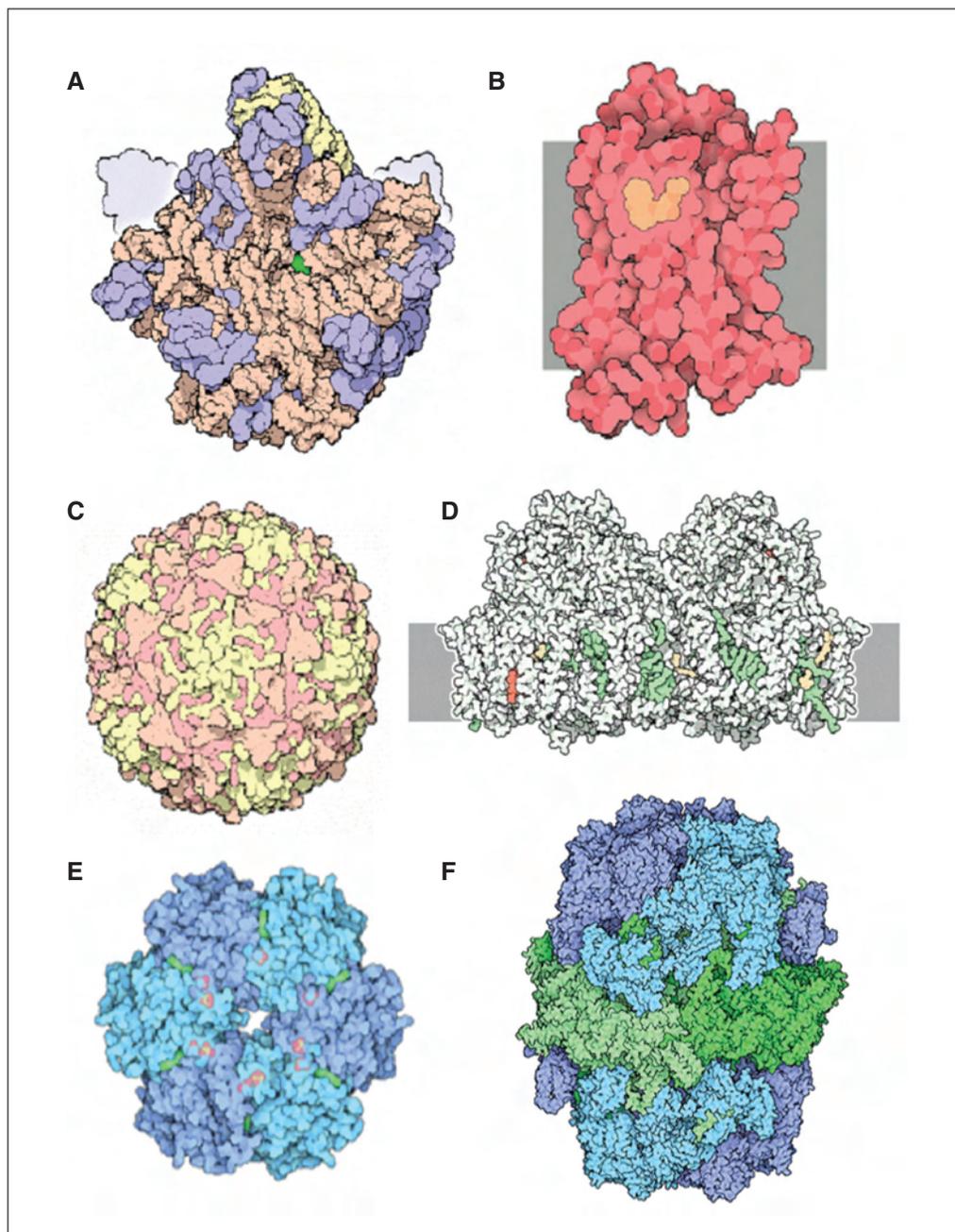
Biophysical Analysis of Nucleic Acids

Figure 7.13.1 Recent triumphs of structural biology. (**A**) The ribosome (large subunit; PDB entry code 1ffk); (**B**) Adrenergic receptor (GPCR; PDB entry code 2rh1); (**C**) Poliovirus (PDB entry code 2plv); (**D**) Photosystem II (PDB entry code 1s5l); (**E**) Cyanobacterial master clock protein KaiC (PDB entry code 2gbl); (**F**) Fatty acid synthase (PDB entry codes 2uvb and 2uvc). Reproduced from Molecule of the Month Illustrations (*info@rcsb.org*; *http://www.rcsb.org/pdb/static.do?p = education_discussion/molecule_of_the_month/index.html*) by David S. Goodsell with permission from RCSB Protein Data Bank (PDB). For the color version of this figure go to *http://www. currentprotocols.com/protocol/nc0713*.

Garman and Owen, 2006), data collection and resolution (synchrotron sources and fast CCD detectors; Hendrickson, 2000), phasing (single- and multi-wavelength anomalous dispersion; Terwilliger and Berendzen, 1999; Weeks et al., 2003), electron density map interpretation and model building (automatic chain tracing; CCP4, 1994; Abola et al., 2000), and structure refinement (increased computer power, simulated annealing and maximum likelihood refinement; Murshudov et al., 1999; Brunger and Adams, 2002).

Table 7.13.1 Nobel Prizes Related to Diffraction and Crystallography

Year	Nobel laureate(s)	Field	Discovery
1901	Wilhelm C. Roentgen	Physics	Recognition of the extraordinary services he has rendered by the discovery of the remarkable rays subsequently named after him
1914	Max von Laue	Physics	Discovery of the diffraction of X-rays by crystals
1915	William H. Bragg and William L. Bragg	Physics	Their services in the analysis of crystal structure by means of X-rays
1927	Arthur H. Compton and Charles T. R. Wilson	Physics	Discovery of the effect named after him, method of making the paths of electrically charged particles visible by condensation of vapor
1936	Peter J. W. Debye	Chemistry	Contributions to our knowledge of molecular structure through his investigations on dipole moments and on the diffraction of X-rays and electrons in gases
1937	Clinton J. Davisson and George P. Thompson	Physics	Experimental discovery of the diffraction of electrons by crystals
1954	Linus C. Pauling	Chemistry	Research into the nature of the chemical bond and its application to the elucidation of the structure of complex substances
1962	John Kendrew and Max Perutz	Chemistry	Studies of the structures of globular proteins
1962	James D. Watson, Francis H. C. Crick, and Maurice H. F. Wilkins	Medicine	Discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material
1964	Dorothy Hodgkin	Chemistry	Determinations by X-ray techniques of the structures of important biochemical substances
1976	William N. Lipscomb	Chemistry	Studies on the structure of boranes illuminating problems of chemical bonding
1982	Aaron Klug	Chemistry	Development of crystallographic electron microscopy and his structural elucidation of biologically important nucleic acid-protein complexes
1985	Herbert A. Hauptman and Jerome Karle	Chemistry	Outstanding achievements in the development of direct methods for the determination of crystal structures
1987	Robert Huber, Johann Deisenhofer, and Hartmut Michel	Chemistry	Determination of the three-dimensional structure of a photosynthetic reaction center
1994	Betram Brockhouse and Clifford Shull	Physics	Development of neutron spectroscopy, development of the neutron diffraction technique
2003	Peter Agre and Roderick MacKinnon	Chemistry	Discovery of water channels, structural and mechanistic studies of ion channels
2006	Roger Kornberg	Chemistry	Studies of the molecular basis of eukaryotic transcription
2009	Thomas A. Steitz, Venkatraman Ramakrishnan, and Ada E. Yonath	Chemistry	For studies of the structure and function of the ribosome

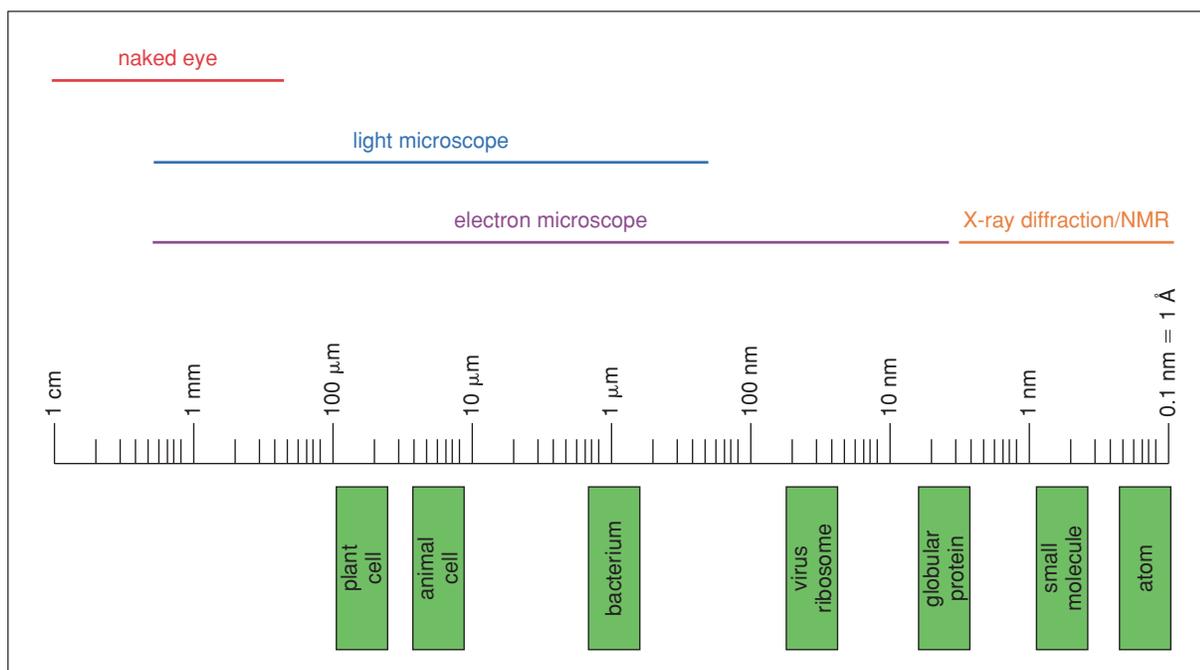**Biophysical Analysis of Nucleic Acids**

7.13.3

Figure 7.13.2 From the visible to the invisible. The diagram depicts the rough sizes of cells and their components on a logarithmic scale and illustrates the range of objects that can be visualized with different techniques.

It is now feasible to mount a protein crystal in the morning and end up with a preliminary, partially refined structure in the afternoon.

However, all these breakthroughs do not change the fact that crystallography can be a tedious business. Crystallization and phasing represent common bottlenecks on the way to a structure, and what is many times a straight-forward exercise can become a make-or-break effort that lasts months or years in some cases. Although it is impossible a priori to identify problem cases, empirical evidence exists supporting the notions that membrane proteins are hard to crystallize, that sampling proteins from various organisms increases the chances of obtaining diffraction-quality crystals, and that derivatization and phasing approaches ideally suited for proteins in the 15- to 50-kDa range are frequently inadequate to crack the structures of large macromolecular assemblies. Particularly as far as the latter are concerned, electron microscopy (EM) represents a powerful approach for structure and function studies at the intermediate 10- to 30-Å resolution range. In favorable cases and with averaging of ≥1 million subunits, near-atomic resolution can be achieved (Fig. 7.13.2) (Baumeister and Steven, 2000; Zhou, 2008). Moreover, hybrid structural approaches, marrying EM and X-ray crystallography or crystallography and solution NMR are becoming ever more popular.

This unit gives an overview of some of the major techniques in structural biology, particularly those that rely on diffraction, by briefly summarizing the benefits and limitations of individual methods and comparing them to each other. It will then describe in some detail the main stages of structure determinations by single-crystal X-ray crystallography, from crystallization to structure refinement, analysis, and quality control. It is by no means the intent of the author to provide an exhaustive account of the topic of X-ray diffraction and macromolecular structure determination. The interested reader may turn to some of the additional reading material listed at the end of the unit for a more in-depth treatment of the individual topics touched upon in this brief review.

MAJOR TECHNIQUES IN STRUCTURAL BIOLOGY

The following methods are considered to be of primary importance for experimental, three-dimensional structure determination: X-ray crystallography, X-ray fiber diffraction, electron diffraction, electron microscopy, neutron diffraction, and nuclear magnetic resonance (NMR). There are additional techniques that can provide insight into the shape of macromolecules, such as for example small angle X-ray scattering (SAXS; Putnam et al., 2007)
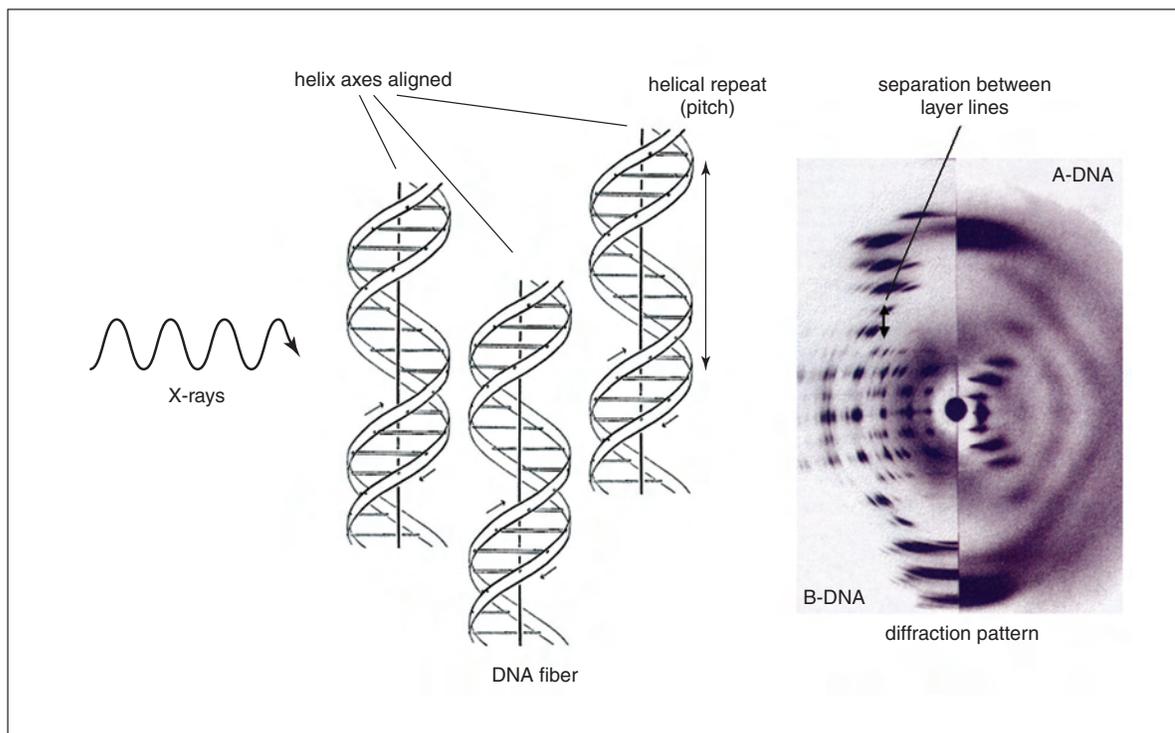
Figure 7.13.3 Principles of fiber diffraction. The diffraction pattern resulting from aligned helical structures in fibers exposed to X-rays exhibit characteristic cross-like shapes. The drawing of the DNA duplex was originally created by Odile Crick and is adapted from Kemp (2003) with permission from the Nature Publishing Group.
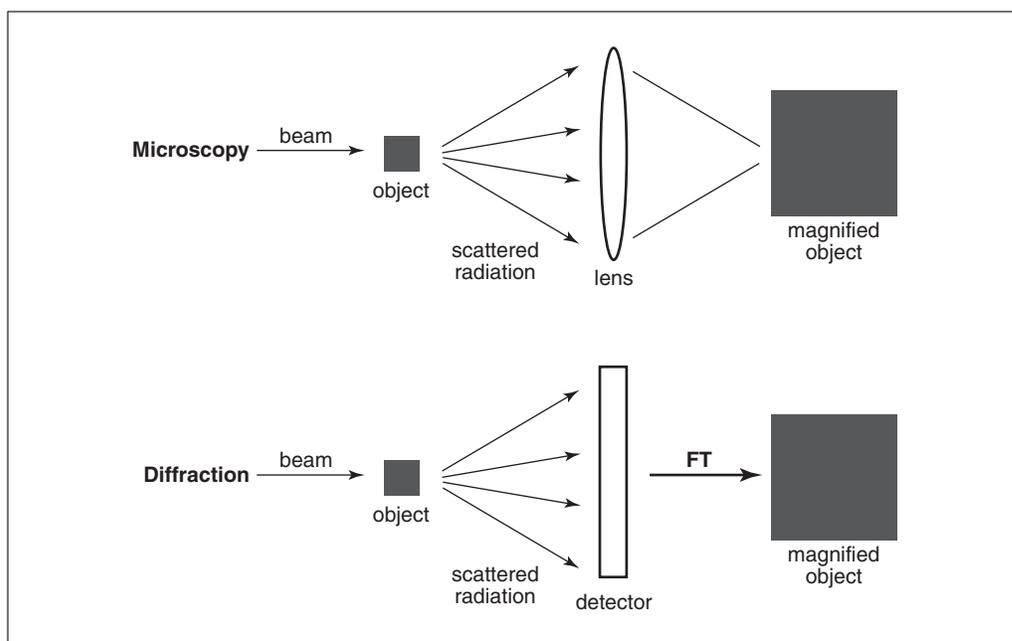

Figure 7.13.4 Light microscopy versus diffraction. Structure determination by X-ray diffraction entails the use of a mathematical lens, Fourier Transformation (FT), to "focus" the scattered radiation.

and fluorescence resonance energy transfer (FRET; Lilley and Wilson, 2000; Schuler and Eaton, 2008). Although these and others are very useful in combination with any of the above approaches and can shed light on the dynamic behavior of molecular systems, they will not be considered further here.

A key difference between optical or electron microscopy and X-ray diffraction is that, unlike light or electron beams, X-rays cannot
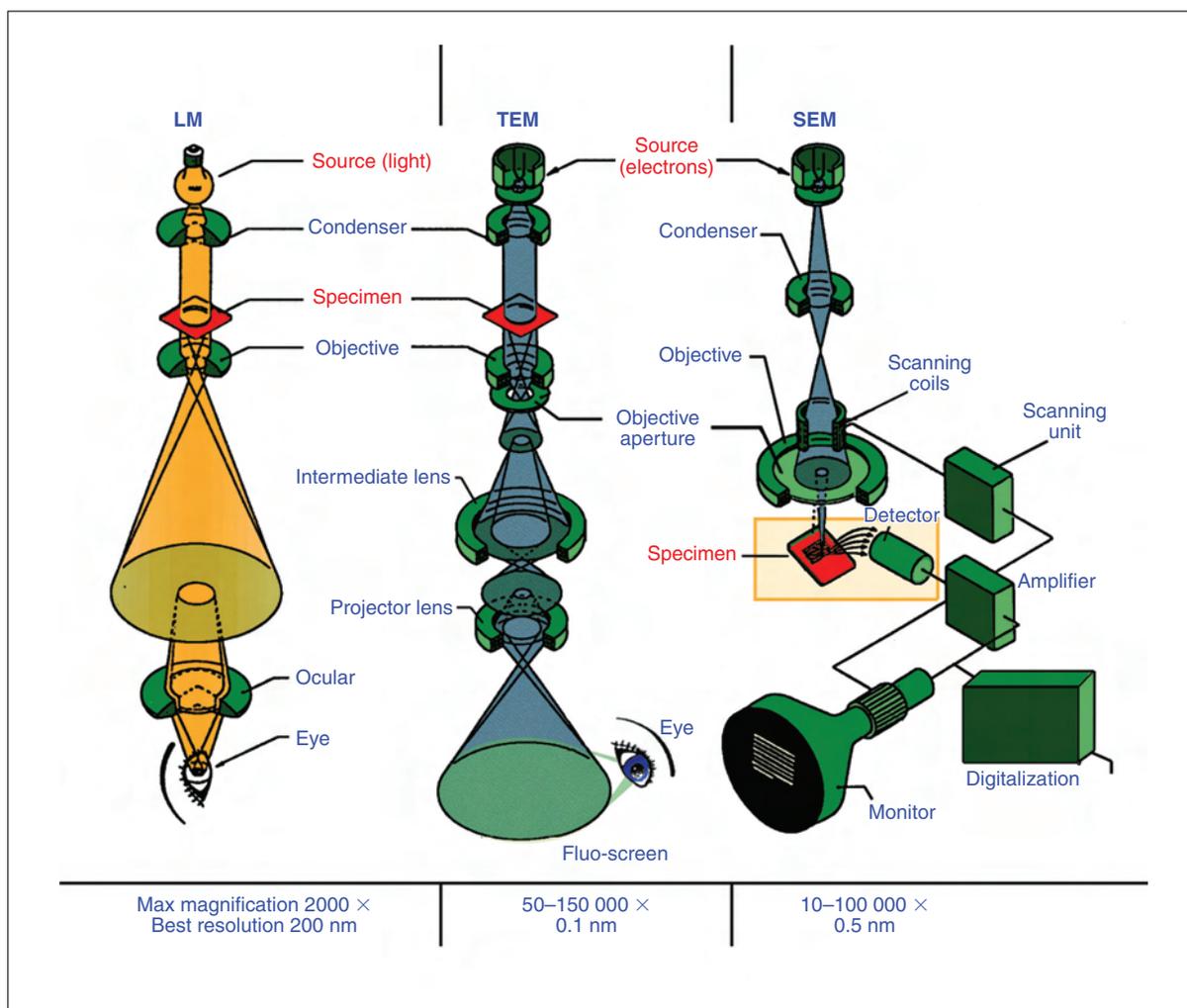
Biophysical Analysis of Nucleic Acids

7.13.5

Figure 7.13.5 Light microscopy versus electron microscopy. Lenses allow reconstruction of the image in both techniques, but to focus electron beams electromagnetic lenses are required. Standard light microscope (LM, left), transmission electron microscope (TEM, center), and scanning electron microscope (SEM, right). Source: *http://www.vcbio.science.ru.nl/images/fesem_beam_zoom.jpg*. Original illustration: Jeol Instruments. Redrawn by vcbio.science.ru.nl, Radboud University Nijmegen. Used with permission.

be focused (Fig. 7.13.3). The X-ray crystallographic visualization of a molecule requires a mathematical lens—Fourier transformation—that generates a 3D structure from the amplitudes of the scattered radiation (the structure factors) and the phases. The phase information is lost in the diffraction experiment, but several methods allow one to recover the phases and we will get back to the so-called phase problem in X-ray crystallography (see Phasing Approaches below).

X-Ray Fiber Diffraction

Fiber diffraction can give key insights into the geometry of nucleic acids or fibrous proteins (i.e., collagen) and its golden era coincides with the discovery of the structure of DNA. Very long double-helical DNA molecules tend to be packed side by side in an ordered manner inside fibers. The helical structure gives rise to cross-shaped diffraction patterns with various separations between layer lines (Fig. 7.13.4). The spacing of layer lines is determined by the helical repeat, and as the repeat distance increases, the layer lines move closer together. The DNA diffraction pattern depicted in Figure 7.13.4 shows different numbers of spots, and the pattern from A-DNA indicates a higher degree of regularity in the packing arrangement of fibers (there are more spots). The B-form and A-form DNA duplexes differ in their helical repeats (34 and 28 Å, respectively). The larger separation of stacked bases along the helical direction in B-DNA compared with A-DNA can be deduced from the smaller separation of diffraction spots in the B-DNA fiber diffraction pattern. From the helical repeat and the
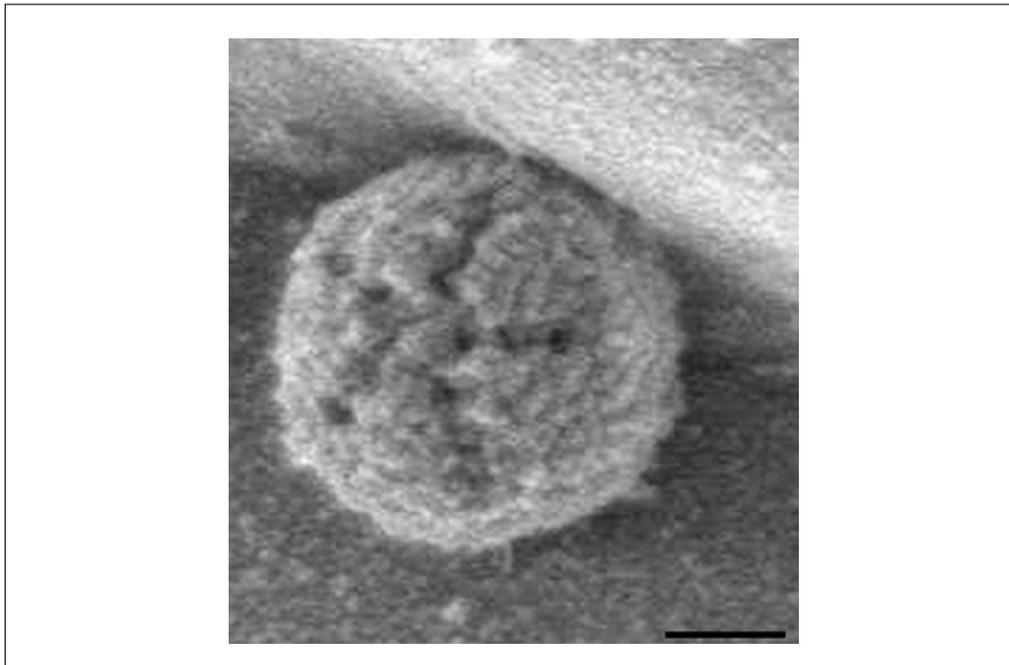
Figure 7.13.6 Example of an SEM image. The star-shaped structure in a mature extracellular *Acanthamoeba polyphaga* mimivirus, an icosahedral double-stranded DNA virus. The scale bar measures 200 nm. Reproduced from Zauberman et al. (2008).
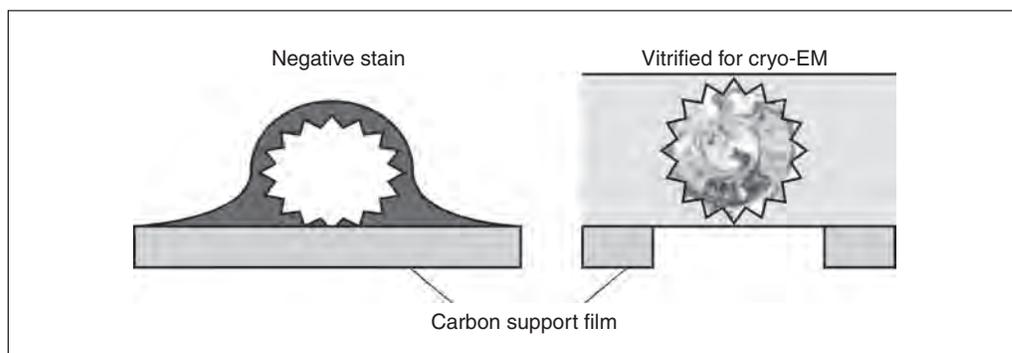

Figure 7.13.7 Negative-stain and cryo-EM. Left: A virus particle is outlined with good contrast by heavy-metal stain but is somewhat flattened due to dehydration. Right: By comparison, it is preserved in the native state in the cryo-EM sample, but the protein-ice contrast is very low. The particle is therefore imaged over holes in the carbon support to maximize the contrast. Reprinted from Saibil (2000) with permission from the International Union of Crystallography.

inclination of the arms in the cross, it is possible to derive an approximate radius for the double helix. Moreover, the orientation of the dyad in the diffraction pattern allowed Watson and Crick to conclude that the two strands in the DNA duplex run in opposite directions. X-ray fiber diffraction is still used today but has gradually given way to single crystal studies (Tsuruta and Irving, 2008). For further information, please see the official Web site for small-angle scattering and fiber-diffraction studies (*http://www.small-angle.ac.uk*).

Electron Diffraction

In terms of the theoretical framework, electron diffraction is similar to X-ray diffraction. However, there are a number of differences that have a significant impact on the practical aspects. Electrons interact strongly with matter and cause serious radiation damage. Thus, the method is typically only applicable to thin layers (2D crystals). Therefore, electron diffraction is useful for certain membrane proteins that may easily form 2D but not 3D crystals. An electron's wavelength decreases as its
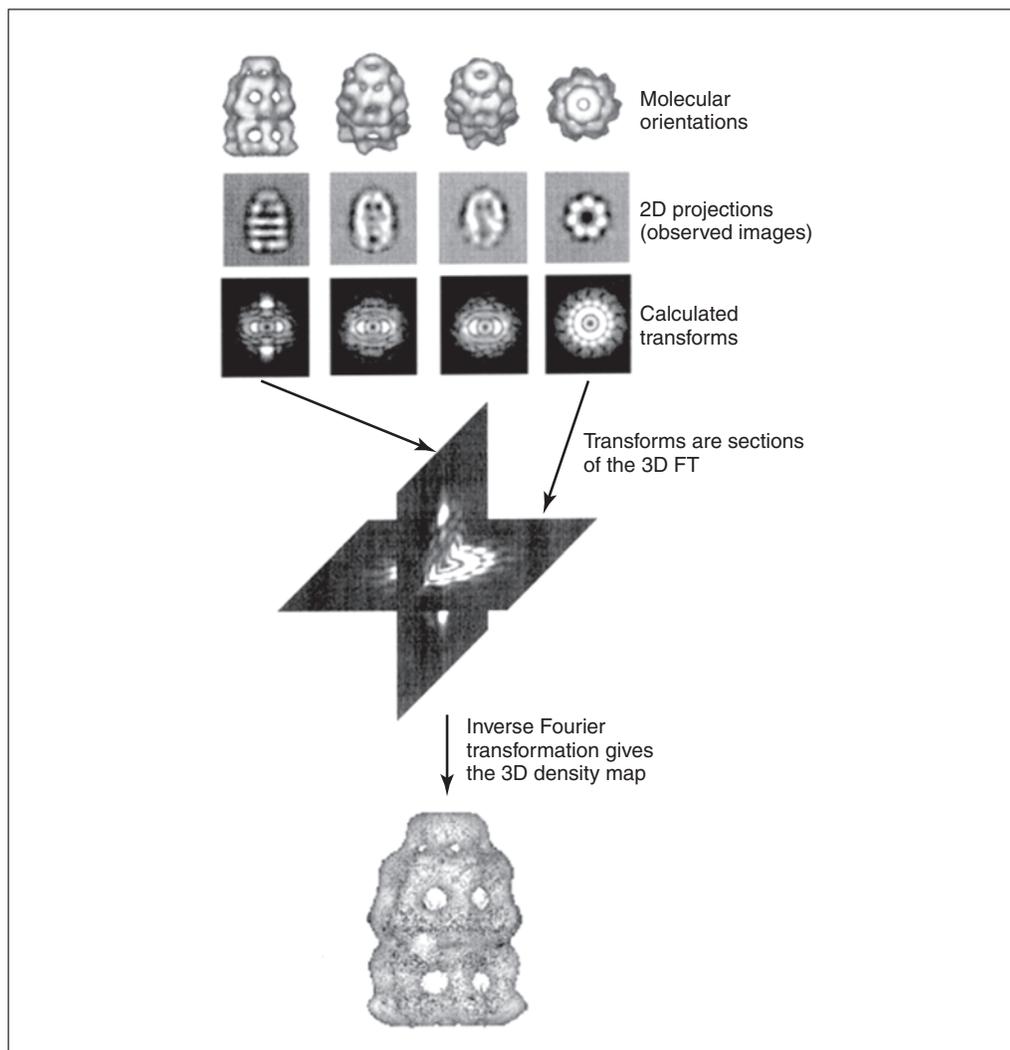
Biophysical Analysis of Nucleic Acids

7.13.7

Figure 7.13.8 Single-particle EM 3D-reconstruction from 2D-projections. A set of 2D-projections (four in this case) is depicted along rendered iso-surfaces. The Fourier transform of a 2D-projection is equivalent to a central section in the 3D-FT of a molecule. Once a sufficient number of sections are available, the complete 3D-transform can be generated and inverse-transformed into a 3D-density map (bottom). Reprinted from Saibil (2000) with permission from the International Union of Crystallography.

velocity increases; in a typical electron microscope the wavelength is around 0.04 Å and thus much lower than the X-ray wavelength used for single crystal diffraction experiments (1 to 2 Å). However, the damage to biological samples caused by the electron beam is such that the effective resolution is often reduced to 10 to 20 Å.

Electron Microscopy

Unlike with X-rays, electromagnetic lenses can be used with electrons to reconstruct the image as in a traditional light microscope. Hence, there is no phase problem. A comparison between a standard light microscope and transmission and scanning electron microscopes (TEM and SEM, respectively) is depicted in Figure 7.13.5. Samples for EM have to be carefully prepared: (i) they need to be exposed to high vacuum and therefore fixed with special chemicals or frozen; (ii) extremely thin sections are required, as electrons have limited penetrating power; and (iii) samples are often exposed to heavy metals (staining) because the contrast depends on the atomic number. In SEM, the specimen is dried and coated with a thin layer of heavy metal. The technique allows visualization of secondary electrons that are scattered or emitted from the specimen surface. SEM provides great depth of focus but only surface features can be examined and the resolution is not very high (around 100 Å).
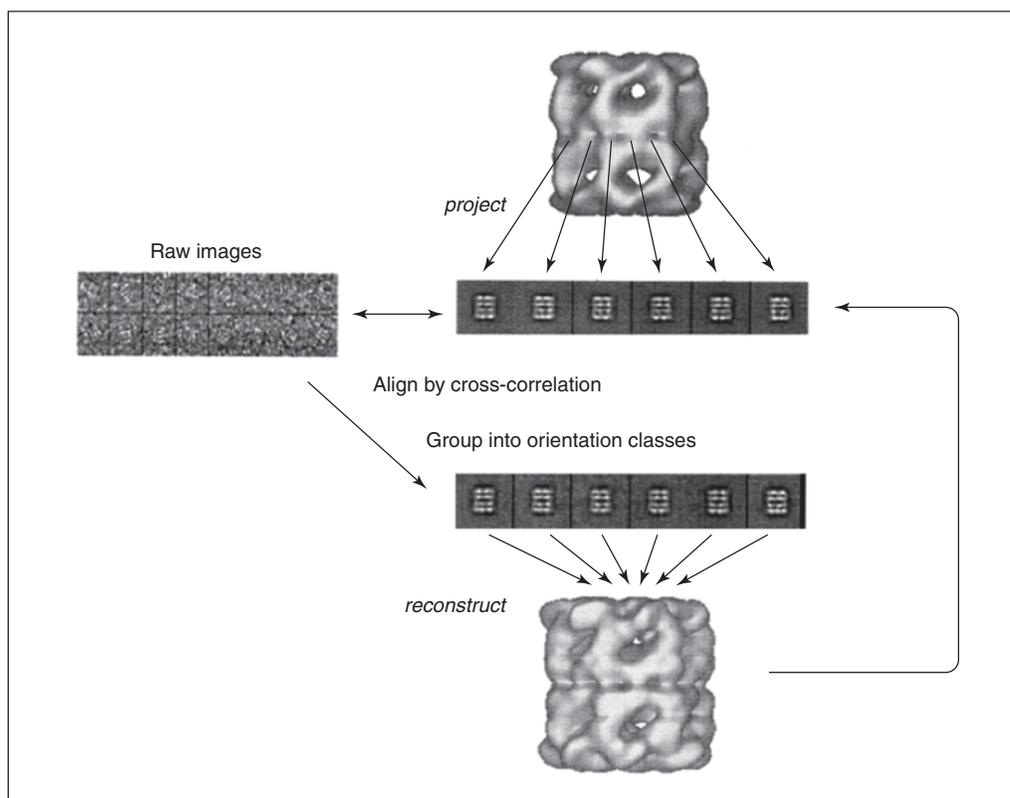
Figure 7.13.9 Refinement by projection matching. Reference images are created by projecting a 3D-map into a set of different orientations (center). Each raw image from the data set (left) is then rotationally and translationally aligned to individual reference images and given the orientation with the highest correlation coefficient. Images aligned in this fashion are grouped and averaged once again to create an improved 3D map (bottom). Reprinted with permission from Saibil (2000) with permission from the International Union of Crystallography.

An example of an SEM image is shown in Figure 7.13.6. TEM uses electrons that have passed through a specimen to form an image.

Specimens are usually fixed, embedded, sectioned, and stained with an electron-dense material. Various techniques can be differentiated, one of them being metal shadowing that allows visualization of surface structures or cell components. Another technique is freeze fracture or freeze etch, used for studying membranes and the cell interior. Finally, negative staining and cryo-electron microscopy (Fig. 7.13.7) can be applied to unfixed biological samples. Thus, these techniques are useful to visualize large macromolecular assemblies such as viruses or ribosomes.

A single protein molecule gives only a weak and ill-defined image in the electron microscope. Increasing the signal by using higher intensity beams or longer exposure only increases the radiation damage. Therefore, it is necessary to combine the information from many molecules so as to average out random errors in the single images. This is more easily achieved when the molecule or particle fea-tures high symmetry, a key property of many viruses (Chiu et al., 1997). It is possible to apply averaging techniques and reconstruction analysis also to nonsymmetric molecules (Saibil, 2000). Images of randomly oriented molecules are collected and classes of similar particles are generated (Fig. 7.13.8). Angles are then assigned to each class and a 3D averaging procedure is carried out. The process can be further refined by projecting the image obtained, and using the projections to break the original classes into smaller ones and then assigning more precise angles (Fig. 7.13.9).

In favorable cases, cryo-EM can reach near-atomic resolution, and if more detailed structures of components of a particle are available from X-ray crystallography or solution NMR, these can be built into the cryo-EM molecular envelope (Zhou, 2008; Fig. 7.13.10). Therefore, EM and X-ray crystallography are complementary techniques. When they compete directly, crystallography delivers far more detailed information (e.g., ribosome, RNA polymerase). Nevertheless, EM is an extremely
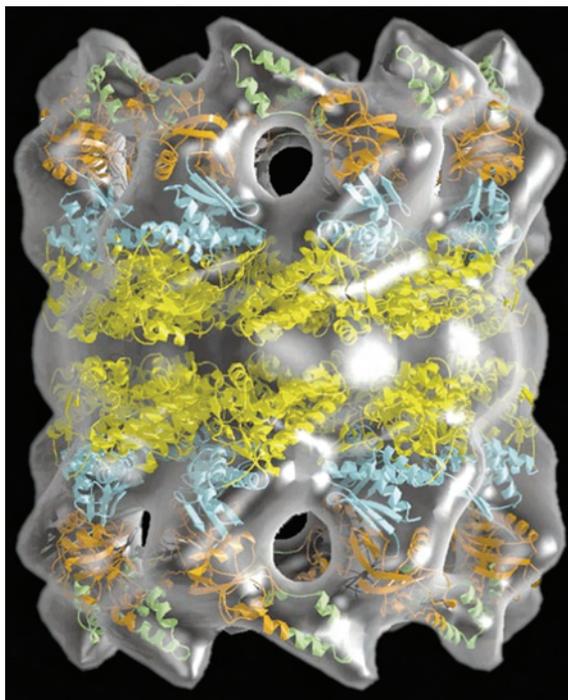
Biophysical Analysis of Nucleic Acids

7.13.9

Figure 7.13.10 3D model of the archaeal thermosome holoenzyme. Crystal structures of the subunits (in color) are modeled into the EM-molecular envelope of the hexadecameric chaperone. Reprinted from Baumeister and Steven (2000) with permission from Elsevier. For the color version of this figure go to *http://www.currentprotocols.com/protocol/nc0713*.

Table 7.13.2 Electron Microscopy versus X-Ray Crystallography

Variable	EM	Crystallography
Sample size	Relatively low amounts of material needed	Milligram quantities required
Crystals	No need for single crystals	Crystallization constitutes a bottleneck on the way to structure determination
Molecular size	Typically applied to large-size macromolecules or assemblies, ($>$300 kDa), but the technique has been successfully applied to a 78-kDa DNA nanostructure (Kato et al., 2009)	No intrinsic size limitation (structures up to the MDa size range have been determined); however, large molecules can be difficult to crystallize
Resolution	Typically 10 Å or less; in favorable cases, near-atomic resolution is possible (Zhou, 2008)	Near-atomic resolution can be achieved even with very large molecules, permitting detailed insights into recognition and mechanistic aspects

useful technique for studying macromolecule assemblies that are difficult to crystallize or in cases where the production of large amounts of materials is problematic. A more detailed comparison of the similarities and differences between EM and X-ray crystallography is provided in Table 7.13.2.

Neutron Diffraction

A fundamental difference between diffraction of X-rays (photons) and neutrons is that the former are scattered by electrons and the latter by protons. Neutrons are highly penetrating and unlike X-rays they are nondestructive, and crystals of macromolecules do not decay

in neutron beams even after lengthy exposure times. X-rays are typically blind to hydrogen atoms in crystals of macromolecules, unless diffraction data are available to extremely high resolution ($\ll 1$ Å). Even in those cases, the hydrogen atoms of water molecules in well-ordered solvent networks (first and second shell hydration) normally remain invisible.

The atomic form factor f in X-ray scattering (a measure of the scattering intensity of a wave by an isolated atom) is replaced by the scattering length b in neutron diffraction. The scattering length varies randomly across the periodic table and its magnitude can differ significantly even with isotopes of the same element, as in the case of hydrogen (^{1}H) and deuterium (^{2}H). The atomic form factors ($f \equiv Z$) and scattering lengths (unit 10^{-15} m, fm) for selected elements and isotopes are: hydrogen ($f = 1$; $b = -3.8$), deuterium ($f = 1$; $b = 6.5$), carbon ($f = 6$, $b = 6.6$), nitrogen ($f = 7$, $b = 9.4$), oxygen ($f = 8$, $b = 5.8$), sulfur ($f = 16$, $b = 3.1$), and iron ($f = 26$, $b = 9.6$). For a full list, please see *http://www.ncnr.nist. gov/resources/n-lengths*). Thus, deuterium and carbon exhibit very similar scattering lengths and the light element can be observed in the presence of the heavier carbon, oxygen, nitrogen, and sulfur atoms (Fig. 7.13.11). Deuterium also displays much weaker incoherent scattering than hydrogen. Therefore, visual-

ization of the positions of hydrogen atoms in neutron crystallographic experiments requires perdeuteration of proteins.

There are a number of advantages of neutron macromolecular crystallography (NMC) for structural biology (Blakeley et al., 2008). The positions of hydrogen atoms can be located even at resolutions of around 2 Å. Thus, NMC is complementary to ultrahigh resolution X-ray macromolecular crystallography (XMC). The protonation and ionization states of atoms can be determined, thus yielding atomic charges and pK_as. Insights can be gained into hydrogen-bonding pattern because NMC allows one to determine the orientation of hydroxyl and amide groups (Hanson et al., 2004). Similarly, the conformations of methyl groups and side chains can be established in neutron density maps, thus providing details on packing arrangements. Because it is possible to observe hydrogen atoms in neutron structures, the orientations of water molecules can be determined, effectively revealing donor and acceptor patterns in water networks. This will contribute to a better understanding of the role of water molecules at active sites and the effects on conformation and stability of solvation shells. Further advantages of NMC concern the monitoring of hydrogen/deuterium (H/D) exchange, permitting insight into solvent accessibility, dynamics, and folding
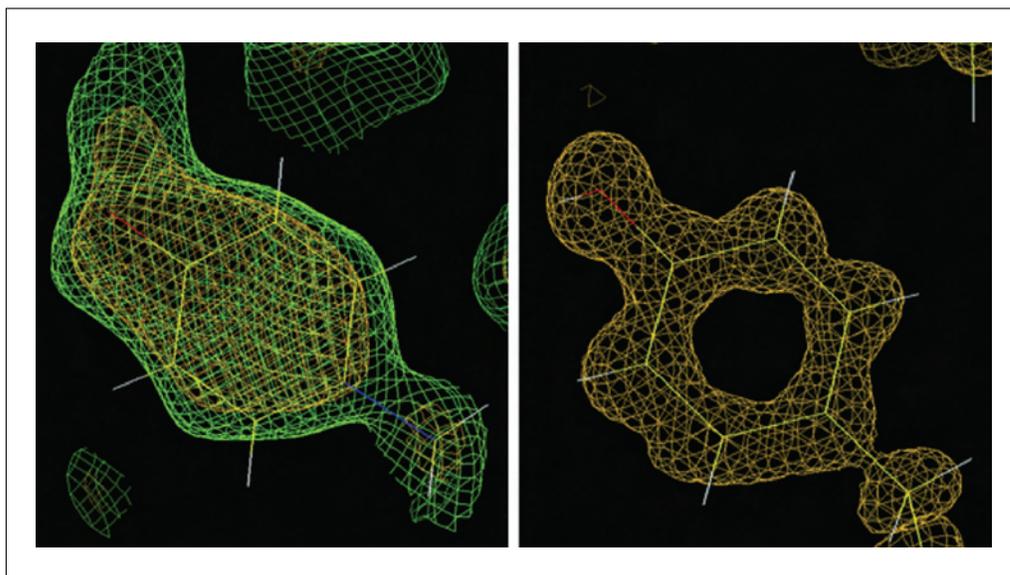

Figure 7.13.11 Neutron versus X-ray macromolecular crystallography. Left: The neutron density for Tyr137 in the structure of D-xylose isomerase contoured at 1.5σ (green) and 2.0σ (yellow) clearly reveals the orientation of the deuteron on the O atom of tyrosine. Right: The protonation state of Tyr254 remains unclear from electron density maps in the X-ray crystal structure of the same enzyme determined to 0.94-Å resolution at -170°C. Reprinted from Hanson et al. (2004) with permission from the International Union of Crystallography. For the color version of this figure go to *http://www.currentprotocols.com/protocol/nc0713*.

Biophysical Analysis of Nucleic Acids

7.13.11

patterns. Finally, NMC allows one to discriminate between metals at active sites due to unique neutron scattering cross-sections, i.e., Mn(25) = −3.6 fm, Fe(26) = 9.5 fm, and Zn(30) = 5.6 fm.

More widespread applications of NMC have traditionally suffered from the high cost of the instrumentation required (either a nuclear reactor or a spallation neutron source, SNS; the complexity and cost of neutron detectors also exceed by far those of state-of-the-art X-ray CCDs) and the need for large crystals ($\sim$1 mm^3). However, the availability of SNSs in Europe, Japan, and the USA (some of these are still under construction; Fig. 7.13.12) that produce high-intensity beams has sparked a renewed interest in applications of neutron scattering and promises a renaissance of NMC. The design criteria for the Macromolecular Neutron Diffractometer (MaNDi) on the SNS at Oak Ridge National Laboratory (ORNL, Oak Ridge, Tennessee, USA) anticipate resolution limits of between 1.5 and 2.0 Å for crystals with a lattice constant of up to 150 Å (2.5 to 3.0 Å for constants of 150 to 300 Å). Moreover, the time spent to collect data from a crystal with a volume of 0.125 mm^3 and unit cell constants of maximum 100 Å is expected to be 24 hr for a resolution of $\sim$2 Å.

Nuclear Magnetic Resonance Spectroscopy

Certain nuclei, such as for example ^{1}H, ^{13}C, ^{15}N, and ^{31}P possess an angular momentum. The energy levels associated with nuclei of different spin angular momentums can be separated in high magnetic fields. The spin will align along the field and absorption of electromagnetic radiation of the appropriate frequency (radio waves), then induce a transition. When the nuclei revert to their equilibrium state, they emit radiation that can be measured. Most importantly, the precise frequency of the emitted radiation is dependent on the environment of the individual nuclei. The environment of a particular nucleus affects the frequency of the emitted radiation. These different frequencies are referred to as chemical shifts. NMR spectra are further complicated by scalar coupling between neighboring nuclei that is apparent from the splitting of individual signals (Fig. 7.13.13; Keeler, 2005).

Protein NMR spectra contain a large number of overlapping peaks, and it is impossible
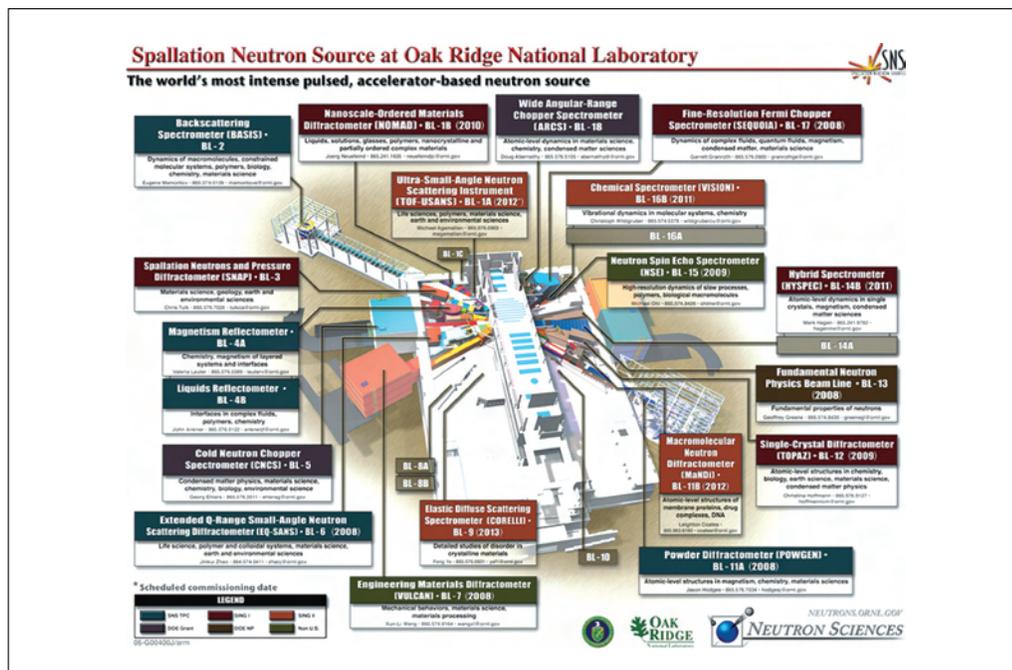

Figure 7.13.12 The Spallation Neutron Source (SNS) at Oak Ridge National Laboratory (ORNL, Oak Ridge, Tennessee, USA). The SNS instrument hall currently under construction will eventually contain 24 instruments on 18 beam lines. The Macromolecular Neutron Diffractometer (MaNDi, BL-11B) and the Single-Crystal Diffractometer (TOPAZ, BL-12), to be completed in 2012 and 2009, respectively, are of particular interest for neutron macromolecular crystallography research. Source: *http://neutrons.ornl.gov/instrument_systems/beamline.shtml*; please visit the original Web site and click on the individual boxes for details of the instrument layout and capabilities.
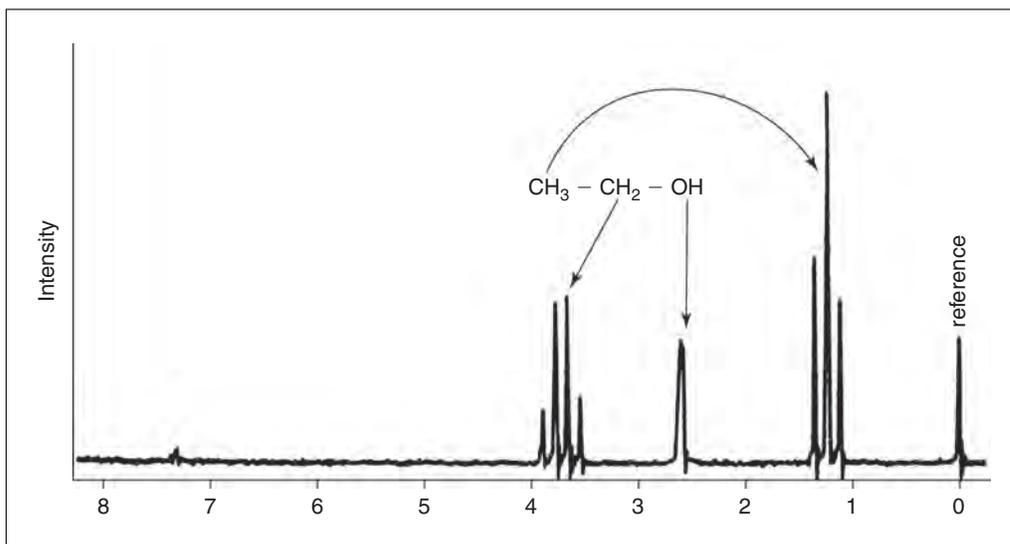
Figure 7.13.13 One-dimensional ^{1}H-NMR spectrum of ethanol. The three groups of protons in this small molecule, (C)H$_3$, (C)H$_2$, and (O)H, all exhibit different chemical shifts relative to the protons in reference molecule, tetramethylsilane (TMS). The characteristic splitting of the signals arising from the methyl (1:2:1) and methylene (1:3:3:1) protons is the result of through-bond coupling between neighboring nuclei.

Figure 7.13.14 Two-dimensional heteronuclear NMR spectroscopy. ^{15}N-HSQC spectrum of the circadian clock protein KaiB from the cyanobacterium *Synechococcus elongatus* recorded on an 800-MHz spectrometer.

Table 7.13.3 X-Ray Crystallography versus Solution NMR Spectroscopy
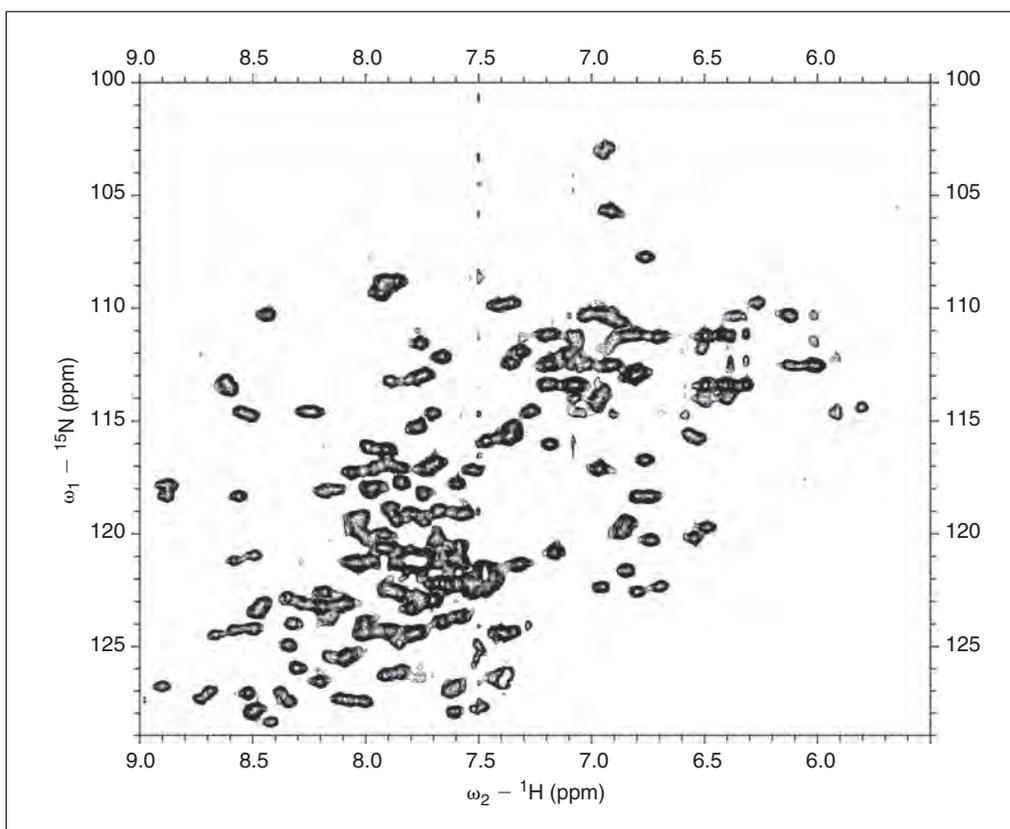
Variable	X-ray crystallography	NMR spectroscopy
Amount and purity of material (30-kDa protein)	10-50 mg, very pure, stable at room temperature	10-20 mg, 95% pure, stable at room temperature; if >10 kDa must be labeled with ^{15}N and/or ^{13}C
Studied sample	Crystals with high content of non-crystalline solvent, ($\sim$50% protein)	$\sim$1 mM protein in solution, ($\sim$1% protein)
Physiological relevance	Artifacts due to crystal packing forces	Artifacts due to use of isolated fragments or domains
Experimental variations of sample conditions	Very difficult, crystallization conditions have to be maintained	Straightforward, can change temperature, pH, and others
Size limitations	Virtually none (40-MDa structures determined)	$\sim$50 kDa at present (40-kDa structures solved)
Meaning of the single data point	None: one spot has contributions from the whole unit cell	Single inter-atomic interactions interpreted as distance or angle constraint
Meaning of all data points	After FT with proper phases, direct definition of the electron density within one unit cell	All observable inter-atomic interactions interpreted as distances and torsion angles
Interpretation of experimental data	Relatively quantitative	Relatively qualitative
Most time consuming	Varies: crystallization, phase generation, model building	Resonance assignments
Final result	One model that minimizes R factor; model variations and uncertainties "hidden" in B values	Many models (ensemble) satisfying constraints; variations explicit in multiple models
Use of stereo-chemical constraints	Necessary (exception: ultra-high resolution data available)	Necessary
Classes of proteins amenable to study	Stable tertiary structure (fold) throughout most of the polypeptide; non-aggregating	Folded and unfolded regions can occur; non-aggregating
Observation of dynamic processes	Very difficult, but in principle enzyme reactions can be followed in crystals	Very straightforward
Measure of accuracy	Biochemical data, threading calculation, φ–ψ plots (potential energy)	Biochemical data, threading calculation, φ–ψ plots (potential energy)
Measure of precision	Resolution, R factors, rms deviations from standard bond lengths and angles	Average root-mean-square deviation (rmsd) among structures in calculated family
Comparable resolution	2 Å resolution, R factor <25%	rmsd <1.0 Å for most of the polypeptide length

to interpret a one-dimensional (1D) spectrum. However, it is possible to design 2D NMR experiments and to plot the results into a *xy*-diagram, i.e., a so-called 2D homonuclear COSY (correlation spectroscopy) experiment. In this 2D representation, the diagonal corresponds to the common 1D spectrum. Off-diagonal peaks arise from the interactions between hydrogen atoms that are relatively closely spaced. Another common type of NMR experiment with proteins concerns the heteronuclear single quantum correlation (HSQC), i.e., between the nitrogen atom of an NH_x group with the attached proton. Therefore, each signal in a ^{15}N-HSQC spectrum represents a signal from a single amino acid. In addition to the signals from the H^N protons in the backbone, the HSQC spectrum also contains signals from the amino groups of the side chains of Asn and Gln and the aromatic N-H groups in the His and Trp side chains. However, unlike a 2D homonuclear spectrum, a heteronuclear ^{13}C- or ^{15}N-HSQC spectrum does not contain a diagonal (Fig. 7.13.14; Wüthrich, 1986).

Relaxation processes are very sensitive to both geometry and motion, but only interactions between atoms that are <5 Å apart can typically be detected. Therefore, NMR spectroscopy allows us to map the distances between pairs of atoms by specifying which pairs are close together in space, NMR spectra contain information about the 3D structure of protein molecules. In reality, it is far from trivial to assign the peaks in a spectrum to a specific H atom in the protein sequence. Kurt Wüthrich worked out a solution to the assignment problem in the 1980s and he was co-awarded the 2002 Nobel Prize in Chemistry for the development of NMR spectroscopy for determining the 3D structure of biological macromolecules in solution. Both solution NMR and X-ray crystallography provide insight into the 3D structures of macromolecules. In many ways, the two techniques are complementary, with the most significant limitation of NMR and crystallography being size (<40 kDa) and the need for single crystals, respectively. A more detailed comparison of these two key techniques in structural biology is provided in Table 7.13.3.

SINGLE CRYSTAL X-RAY CRYSTALLOGRAPHY

Overview

The following sections are dedicated to arguably the most powerful "weapon" in the structural biology arsenal: X-ray crystallography. This technique can provide more detailed models than any of the other approaches available to study macromolecules. In principle, there is no limitation as far as size is concerned: the basic principles remain the same independently of whether one is working out the structure of an oligopeptide with a molecular weight of a few kDa or that of a virus a thousand times larger. Individual steps of a structure determination are outlined in Figure 7.13.15. Among them, crystallization and phasing constitute the biggest hurdles. Despite the fact that impressive advances have been made in recent years to increase the chances of obtaining protein or nucleic acid crystals, crystallization has remained a trial-and-error approach that frequently fails when only a single construct is available. It can easily escalate into a potentially costly and time-consuming battle when various constructs and/or homologous proteins from different organisms are screened (McPherson, 1998). However, the end—be it a detailed 3-dimensional model of an enzyme, receptor, RNA, or protein-DNA complex and the biological insights gained from it—generally justifies the means.

Sample Preparation

Crystallography requires large, milligram, amounts of pure material, precluding in most cases isolation of enzymes or receptors for crystallization from tissues. Instead, proteins based on recombinant DNA technology are used for the structural studies. The DNA is subcloned from a cDNA library or, alternatively, the gene is synthesized. A battery of expression vectors is commercially available and, while *E. coli* still represents the most common organism for over-expression, insect cells, yeast, and human cell lines are becoming ever more popular for producing recombinant proteins. In addition, cell-free expression should also be considered as an alternative approach.

Molecules for crystallization need to be reasonably well structured and not floppy. Therefore, it is important to consider possibly unstructured or flexible regions, i.e., at the N- or C-terminus, in the design of the construct. Constructs amenable to crystallization can often be identified by limited proteolysis (Dong et al., 2007). In many cases, only domains can be crystallized, or it is necessary to resort to the homologous proteins from a thermophilic organism for successful crystallization. Induced-fit binding of a ligand may

STEP	METHOD
Protein, nucleic acid production	Isolation/expression (proteins), synthesis/in vitro transcription (nucleic acids)
Purification	Affinity chromatography (including fusion proteins), ion-exchange chromatography, gel electrophoresis, antibodies
Crystallization	Various sparse matrix crystallization screens, sitting or hanging drop vapor diffusion techniques; basically trial and error – often requires repurification or alternative constructs
Crystal characterization	Diffraction? If yes, resolution limit, unit cell constants, space group, number of molecules per asymmetric unit, solvent content, etc.
Data collection	In house X-ray diffraction setups, often X-ray synchrotron, CCDs
Phase determination	Multiple isomorphous replacement (MIR), molecular replacement (MR), single- or multiwavelength anomalous dispersion (SAD or MAD, resp.), direct methods (DM)
Electron density map	Fourier transformation of phased diffraction data
Fitting of model to map	Automatic or manual chain tracing using computer graphics; iterative rounds of building and refinement guided by electron density, chemical rationale, R-free
Structure analysis	Ramachandran plot, root mean square deviations (r.m.s.d.'s) for key parameters, such as bond lengths and angles, omit electron density maps etc.

(Refinement spans from Phase determination through Fitting of model to map)

Figure 7.13.15 Individual stages of a macromolecular X-ray crystal structure determination. Selected methods are listed on the right. Approaches for refining structures include least squares fitting and simulated annealing. Adapted from Ringe and Petsko (1996).

render the protein with the ligand bound more likely to crystallize than protein alone. It is also worthwhile to consider whether there are many charged residues solvent exposed. This is because reduction of surface entropy by mutation of Lys to Ala or other strategies can dramatically increase the chances of obtaining crystals or of producing higher quality crystals (Czepas et al., 2004). Another important aspect concerns the size of the protein: Is the target a small protein (less than ~70 amino acids) or a polypeptide? In that case, crystallization of the small protein as a fusion with a larger and well-characterized protein, such as glutathione-S-transferase (GST), should be tried (Smyth et al., 2003). This often improves solubility and allows for phasing by molecular replacement of the GST.

Fusion with a variety of tags or proteins also facilitates purification via affinity chromatography (Structural Genomics Consortia, 2008). Some popular ones include the $(His)_6$ tag, GST, maltose-binding protein (MBP), and small ubiquitin-like modifier (SUMO) protein. Further purification steps may involve gel filtration and/or ion-exchange chromatography. Procedures that should be avoided are ammonium sulfate precipitation and lyophilization, and care should be applied when combining various fractions following column chromatography or different batches of protein. In general, the purification should be carried out quickly and proteins need to be handled gently and maintained at reduced temperature. Turbid samples need to be centrifuged and, for filtrations, cartridges with minimal dead volume should be used and one should check for adsorption (OD/activity) after filtering. As a rule of thumb, the purity of a protein should be 90% to 95% by SDS-polyacrylamide gel electrophoresis (PAGE) with Coomassie stain. The purified protein can be further characterized with native PAGE, light scattering, isoelectric focusing (to determine the pI), mass spectrometry, circular dichroism (CD) spectroscopy, and other techniques. Proteins of low solubility (less than 1 mg/mL) are typically not suitable for crystallization experiments, and a search for other constructs or mutation via in vitro directed evolution may be advisable in such cases.

DNA is produced by solid-phase chemical synthesis using suitably protected phosphoramidite building blocks (Gait, 1984). Two basic methods exist for producing RNAs of sufficient quality suitable for crystallization and X-ray structure determination. Longer fragments (>50 nucleotides) can be generated by in vitro transcription using the DNA-dependent T7 RNA polymerase (Milligan and Uhlenbeck, 1989; Wyatt et al., 1991). For shorter RNA oligonucleotides, the method of choice is chemical synthesis, usually by the solid-phase phosphoramidite technique. Due to the presence of the 2′-hydroxyl group in the furanose sugar, chemical synthesis of RNA is more complicated compared with DNA. Common protection groups for the 2′-OH moiety are the tertiary butyl dimethyl silyl (TB-DMS; Scaringe et al., 1990; Wincott et al., 1995) group, the 2′-acetoxy ethyl orthoester (2′-ACE; Scaringe et al., 1998), and the triisopropylsilyloxymethyl functionality (TOM; Pitsch et al., 2001). The latter approach has allowed production of RNAs as long as 100 residues, a size range that includes many biologically interesting RNA motifs. Once deprotected and cleaved from the solid support, DNA and RNA oligonucleotides are typically purified via trityl-on reverse phase HPLC or ion-exchange chromatography. However, column chromatography is not suitable for the purification of longer fragments. Instead, large RNAs need to be purified by denatured PAGE and desalted following elution from the gel (Wyatt et al., 1991).

Crystallization

There are a number of crystallization techniques commonly used with proteins or nucleic acids: Hanging-drop and sitting-drop vapor diffusion, batch/microbatch under oil, free interface diffusion employing either integrated fluidic circuits (i.e., the Topaz crystallization system) or the Zeppezauer tube, and dialysis (Ducruix and Giegé, 1992; Carter and Sweet, 1997a,b; McPherson, 1998; McRee, 1999; Carter, 2003a,b; Rhodes, 2006; Drenth, 2007). The first two techniques are illustrated schematically in Figure 7.13.16. Both are fast and easy to set up and versatile for both screening and optimization. The droplets can be viewed through glass (hanging drop) or either a plastic lid or a transparent tape (sitting drop) under a microscope. The drop size can vary but the volume of hanging drops is usually limited to ∼5 μL. In both cases, the concentration of the particular precipitant in the reservoir exceeds that in the drop. As a result, water will diffuse from the drop to the reservoir, thus increasing the concentration of the precipitant in the drop over time and slowly lowering the solubility of the protein. Ideally, the protein solution will change from the unsaturated region (in terms of a phase diagram) to a labile, supersaturated region, where stable nuclei spontaneously form and grow. The advantage of the sitting drop method is that it can be automated and used in combination with crystallization robots.

Microbatch crystallizations using petroleum oil or silicon oil are also easily set up and can be automated to some degree as well. By comparison, crystallizations using dialysis are somewhat more time consuming to set up, but the method allows for a greater control of the individual parameters that affect crystallization. Moreover, dialysis is ideal for replacing the crystallization buffer by a cryo solution, required for flash freezing crystals. Free interface diffusion in a Zeppezauer tube works better in microgravity, but crystallization experiments in space are expensive and not likely to be available in the foreseeable future.

Crystallization remains a trial and error—mostly error—approach and there is no general recipe for overcoming the nucleation barrier, i.e., a universal nucleant. There are many ways to achieve supersaturation in principle, including adding protein directly to precipitant, altering the temperature, increasing the salt concentration (salt out), decreasing the salt concentration (salt in), adding a ligand that changes the solubility of the protein, altering the dielectric constant of the medium, evaporating water, adding polymer (i.e., polyethylene glycols, PEGs) to produce volume exclusion, adding a cross-linking agent, concentrating the macromolecule, and removing a solubilizing agent. Success in crystallization is to a large degree dependent on crystal packing interactions and these remain unpredictable. Lattice contacts are noncovalent and entail various classes of hydrogen bonds (direct bonds between polar, uncharged groups such as OH, NH_2, =O; direct bonds between one or more charged groups, so-called salt bridges; two polar or charged groups bridged by a water molecule; bridging of two moieties by a chain of two or more waters) and van der Waals interactions. Optimal packing requires electrostatic and shape complementarity.

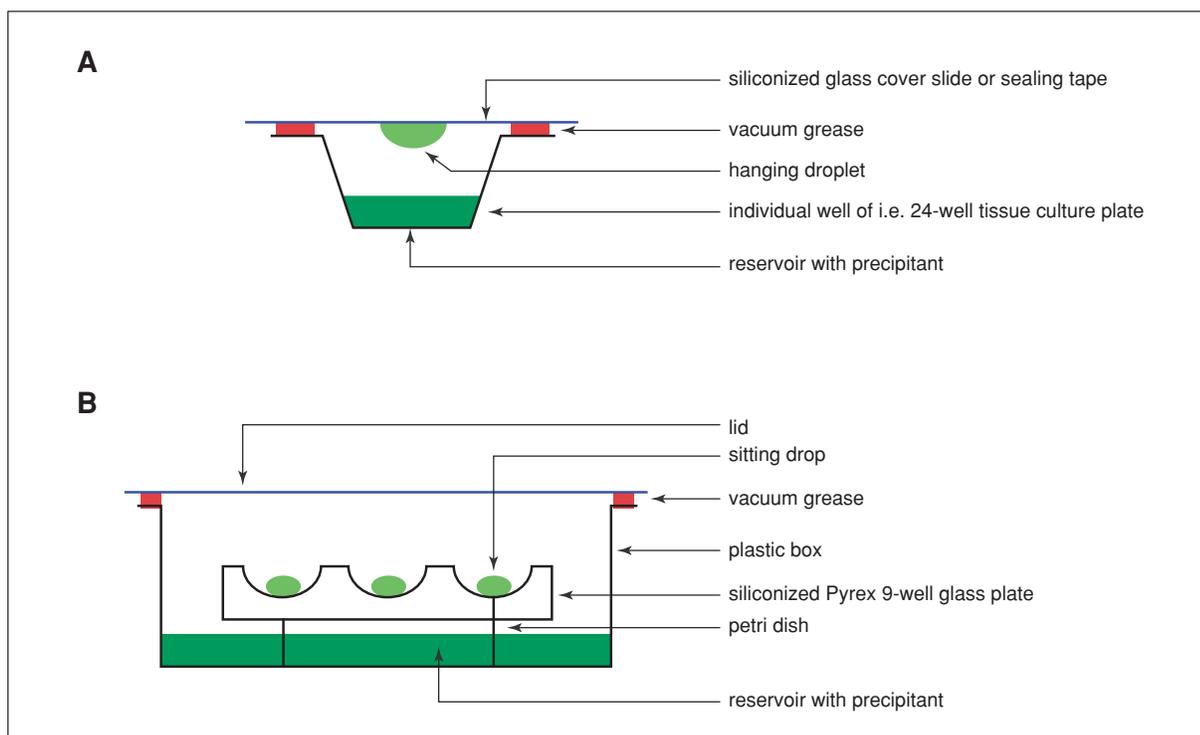It is now common to resort to so-called sparse matrix crystallization screens

7.13.17

Figure 7.13.16 Two related methods for growing single crystals of biomacromolecules. Schematic depictions of the (**A**) hanging and (**B**) sitting drop vapor diffusion techniques. The volume of the droplets is in the nL (nanodrop setting robots) to μL range.

to increase the chances of obtaining crystals. Such screening kits are commercially available (see for example, *http://www.hamptonresearch.com*) and they come in a variety of flavors, suitable for proteins, protein-protein complexes, membrane proteins, DNA oligonucleotides, RNA, and so forth. The initial set of protein crystallization solutions compiled by Jancarik and Kim in the early 1990s is shown in Figure 7.13.17 (Jancarik and Kim, 1991). Individual solutions typically feature a salt, a particular precipitant, and a buffer. The pH of the buffers ranges from ~4 to 9 and ammonium sulfate figures prominently in the list of salts or precipitants. Similarly, various classes of PEGs are favorites among the precipitants. The recipes for many of these screens are largely based on empirical data that demonstrate, for example, that many proteins can be crystallized from ammonium sulfate solutions. However, not all salts are the same and in the Hofmeister series one can distinguish between stabilizing kosmotropes (weakly hydrated cations such as NH_4^+ or Cs^+ and strongly hydrated anions such as citrate or sulfate) and destabilizing chaotropes (strongly hydrated cations such as Mg^{2+} or Al^{3+} and weakly hydrated anions such as nitrate or perchlorate; Collins, 2004). The use of PEGs in protein crystallization is based on the tendency of the random coil, water-soluble polymers to reduce protein solubility by volume exclusion (PEG and protein cannot occupy the same space at the same time). This mutual exclusion is mainly dependent on size and shape as well as on concentration.

Setting up hundreds or perhaps thousands of crystallization trials is a tedious task and the screening process is nowadays facilitated by crystallization robotics. An example of a crystallization robot is depicted in Figure 7.13.18. Robotics can be used to generate crystallization screens (so-called liquid handlers), to set up sitting-drop crystallization plates (the 96-well format is quite common), and to barcode, store, retrieve, and image at regular intervals of one's choice the plates. Epifluorescence microscopy can be used to differentiate between crystals of salt and protein; phosphate buffer should be avoided as phosphate tends to crystallize readily and such crystals are then often mistaken for crystals of a macromolecule. Initial leads can be further optimized by manual crystallization setups and the size optimized by seeding. Microseeding uses seed beads from crushed crystals in a serial dilution to seed fresh drops in the hope that the introduction of a few seed nuclei into a metastable solution will produce

Crystal Screen™   HR2-110 Reagent Formulation

Tube #	Salt	Tube #	Buffer ◊	Tube #	Precipitant
1.	0.02 M Calcium chloride dihydrate	1.	0.1 M Sodium acetate trihydrate pH 4.6	1.	30% v/v (+/-)-2-Methyl-2,4-pentanediol
2.	None	2.	None	2.	0.4 M Potassium sodium tartrate tetrahydrate
3.	None	3.	None	3.	0.4 M Ammonium phosphate monobasic
4.	None	4.	0.1 M TRIS hydrochloride pH 8.5	4.	2.0 M Ammonium sulfate
5.	0.2 M Sodium citrate tribasic dihydrate	5.	0.1 M HEPES sodium pH 7.5	5.	30% v/v (+/-)-2-Methyl-2,4-pentanediol
6.	0.2 M Magnesium chloride hexahydrate	6.	0.1 M TRIS hydrochloride pH 8.5	6.	30% w/v Polyethylene glycol 4,000
7.	None	7.	0.1 M Sodium cacodylate trihydrate pH 6.5	7.	1.4 M Sodium acetate trihydrate
8.	0.2 M Sodium citrate tribasic dihydrate	8.	0.1 M Sodium cacodylate trihydrate pH 6.5	8.	30% v/v 2-Propanol
9.	0.2 M Ammonium acetate	9.	0.1 M Sodium citrate tribasic dihydrate pH 5.6	9.	30% w/v Polyethylene glycol 4,000
10.	0.2 M Ammonium acetate	10.	0.1 M Sodium acetate trihydrate pH 4.6	10.	30% w/v Polyethylene glycol 4,000
11.	None	11.	0.1 M Sodium citrate tribasic dihydrate pH 5.6	11.	1.0 M Ammonium phosphate monobasic
12.	0.2 M Magnesium chloride hexahydrate	12.	0.1 M HEPES sodium pH 7.5	12.	30% v/v 2-Propanol
13.	0.2 M Sodium citrate tribasic dihydrate	13.	0.1 M TRIS hydrochloride pH 8.5	13.	30% v/v Polyethylene glycol 400
14.	0.2 M Calcium chloride dihydrate	14.	0.1 M HEPES sodium pH 7.5	14.	28% v/v Polyethylene glycol 400
15.	0.2 M Ammonium sulfate	15.	0.1 M Sodium cacodylate trihydrate pH 6.5	15.	30% w/v Polyethylene glycol 8,000
16.	None	16.	0.1 M HEPES sodium pH 7.5	16.	1.5 M Lithium sulfate monohydrate
17.	0.2 M Lithium sulfate monohydrate	17.	0.1 M TRIS hydrochloride pH 8.5	17.	30% w/v Polyethylene glycol 4,000
18.	0.2 M Magnesium acetate tetrahydrate	18.	0.1 M Sodium cacodylate trihydrate pH 6.5	18.	20% w/v Polyethylene glycol 8,000
19.	0.2 M Ammonium acetate	19.	0.1 M TRIS hydrochloride pH 8.5	19.	30% v/v 2-Propanol
20.	0.2 M Ammonium sulfate	20.	0.1 M Sodium acetate trihydrate pH 4.6	20.	25% w/v Polyethylene glycol 4,000
21.	0.2 M Magnesium acetate tetrahydrate	21.	0.1 M Sodium cacodylate trihydrate pH 6.5	21.	30% v/v (+/-)-2-Methyl-2,4-pentanediol
22.	0.2 M Sodium acetate trihydrate	22.	0.1 M TRIS hydrochloride pH 8.5	22.	30% w/v Polyethylene glycol 4,000
23.	0.2 M Magnesium chloride hexahydrate	23.	0.1 M HEPES sodium pH 7.5	23.	30% v/v Polyethylene glycol 400
24.	0.2 M Calcium chloride dihydrate	24.	0.1 M Sodium acetate trihydrate pH 4.6	24.	20% v/v 2-Propanol
25.	None	25.	0.1 M Imidazole pH 6.5	25.	1.0 M Sodium acetate trihydrate
26.	0.2 M Ammonium acetate	26.	0.1 M Sodium citrate tribasic dihydrate pH 5.6	26.	30% v/v (+/-)-2-Methyl-2,4-pentanediol
27.	0.2 M Sodium citrate tribasic dihydrate	27.	0.1 M HEPES sodium pH 7.5	27.	20% v/v 2-Propanol
28.	0.2 M Sodium acetate trihydrate	28.	0.1 M Sodium cacodylate trihydrate pH 6.5	28.	30% w/v Polyethylene glycol 8,000
29.	None	29.	0.1 M HEPES sodium pH 7.5	29.	0.8 M Potassium sodium tartrate tetrahydrate
30.	0.2 M Ammonium sulfate	30.	None	30.	30% w/v Polyethylene glycol 8,000
31.	0.2 M Ammonium sulfate	31.	None	31.	30% w/v Polyethylene glycol 4,000
32.	None	32.	None	32.	2.0 M Ammonium sulfate
33.	None	33.	None	33.	4.0 M Sodium formate
34.	None	34.	0.1 M Sodium acetate trihydrate pH 4.6	34.	2.0 M Sodium formate
35.	None	35.	0.1 M HEPES sodium pH 7.5	35.	0.8 M Sodium phosphate monobasic monohydrate 0.8 M Potassium phosphate monobasic
36.	None	36.	0.1 M TRIS hydrochloride pH 8.5	36.	8% w/v Polyethylene glycol 8,000
37.	None	37.	0.1 M Sodium acetate trihydrate pH 4.6	37.	8% w/v Polyethylene glycol 4,000
38.	None	38.	0.1 M HEPES sodium pH 7.5	38.	1.4 M Sodium citrate tribasic dihydrate
39.	None	39.	0.1 M HEPES sodium pH 7.5	39.	2% v/v Polyethylene glycol 400 2.0 M Ammonium sulfate
40.	None	40.	0.1 M Sodium citrate tribasic dihydrate pH 5.6	40.	20% v/v 2-Propanol 20% w/v Polyethylene glycol 4,000
41.	None	41.	0.1 M HEPES sodium pH 7.5	41.	10% v/v 2-Propanol 20% w/v Polyethylene glycol 4,000
42.	0.05 M Potassium phosphate monobasic	42.	None	42.	20% w/v Polyethylene glycol 8,000
43.	None	43.	None	43.	30% w/v Polyethylene glycol 1,500
44.	None	44.	None	44.	0.2 M Magnesium formate dihydrate
45.	0.2 M Zinc acetate dihydrate	45.	0.1 M Sodium cacodylate trihydrate pH 6.5	45.	18% w/v Polyethylene glycol 8,000
46.	0.2 M Calcium acetate hydrate	46.	0.1 M Sodium cacodylate trihydrate pH 6.5	46.	18% w/v Polyethylene glycol 8,000
47.	None	47.	0.1 M Sodium acetate trihydrate pH 4.6	47.	2.0 M Ammonium sulfate
48.	None	48.	0.1 M TRIS hydrochloride pH 8.5	48.	2.0 M Ammonium phosphate monobasic
49.	1.0 M Lithium sulfate monohydrate	49.	None	49.	2% w/v Polyethylene glycol 8,000
50.	0.5 M Lithium sulfate monohydrate	50.	None	50.	15% w/v Polyethylene glycol 8,000

◊ Buffer pH is that of a 1.0 M stock prior to dilution with other reagent components: pH with HCl or NaOH.

Crystal Screen contains fifty unique reagents. To determine the formulation of each reagent, simply read across the page.

Figure 7.13.17 Example of a sparse matrix crystallization screen. Composition of the 50 solutions in the so-called Crystal Screen that is commercially available from Hampton Research. Reprinted with permission from Hampton Research (*http://hamptonresearch.com/product_detail.aspx?cid = 1&sid = 17&pid = 1*).

Biophysical Analysis of Nucleic Acids
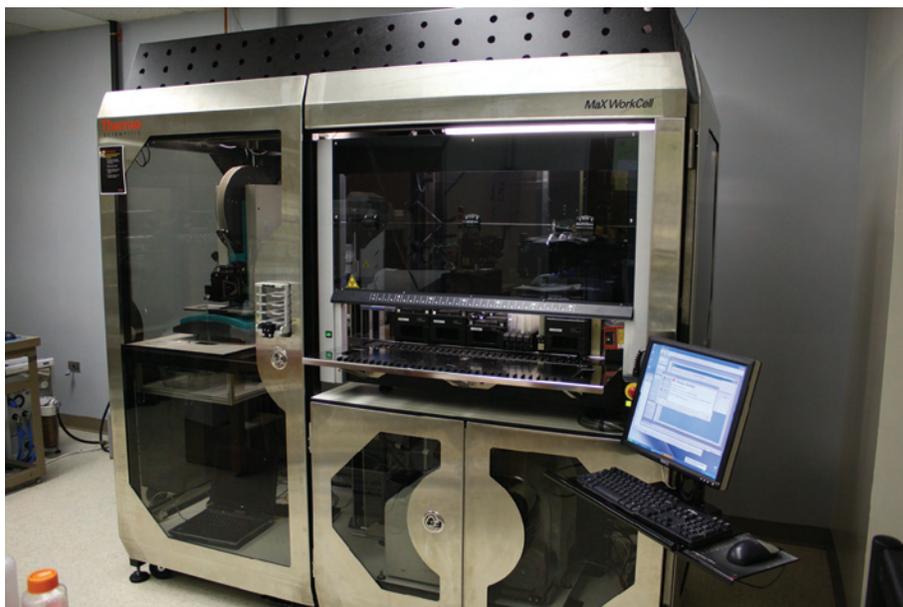
7.13.19

Figure 7.13.18 Automation of crystallization experiments. Crystallization robot in the laboratory of the author, the MaX WorkCell by Thermo Fisher Scientific. The WorkCell integrates nanodrop setting ("Mosquito," Molecular Dimensions; on the left), liquid handling, and screen preparation ("Starlet," Hamilton Company; on the right), and sealing of crystallization plates (bottom right). The robot can handle a wide range of crystallization plates and formats and is typically combined with so-called storage hotels and a plate imager (not shown) that permit automated, periodic access to bar-coded plates and digital photography of individual droplets, respectively.

larger crystals. Streak seeding is similar to micro-seeding but quicker in that a whisker is used to pull off seeds from a crystal in order to then streak it through a fresh drop. Finally, macro-seeding consists of partially dissolving the surface layers of a crystal and then placing it into a fresh metastable solution for growth (*http://xray.bmc.uu.se/~terese/crystallization/tutorials/tutorial4.html*).

There are some differences between the crystallizations of proteins and nucleic acids, owing to the polyanionic nature of the latter. Thus, many DNA or RNA oligonucleotides can be crystallized in the presence of either magnesium chloride or polyamines (e.g., spermine tetrahydrochloride; Berger et al., 1996). Other alkaline earth metal ions such as Ca^{2+}, Sr^{2+}, and Ba^{2+} are also quite widespread, as are Na^+, K^+, and Rb^+. Sodium cacodylate represents a very common buffer and 2-methyl-2,4-pentanediol (MPD), ammonium sulfate, and PEGs are probably the most commonly used precipitants (Baeyens et al., 1994). When all attempts to crystallize a protein fail, it is a good idea to resort to a different construct or to try a homologue from a different organism. Similarly, the key to success in nucleic acid crystallization is to try multiple sequences and to include overhanging

bases at the 5′- or 3′-termini. Another option in RNA crystallography is helix engineering, for example by incorporating a tetraloop at the end of a stem (double helical) region and a tetraloop receptor elsewhere (Ferré-D'Amaré et al., 1998a). The pairing of such motifs often mediates stabilizing intermolecular contacts. A related approach to potentially generate a stable lattice is the use of mutagenized RNAs with a binding site for a particular protein. An example of this is constituted by a hepatitis delta virus ribozyme that contains the high-affinity binding site for the basic RNA binding domain of the U1A spliceosomal protein (Ferré-D'Amaré et al., 1998b).

A note of caution at the end of this section: although it is exciting to see crystals under a microscope, it turns out that many crystals do not diffract X-rays at all, or only very weakly. Before letting the excitement build up too much, it is therefore a good idea to test the crystals for diffraction on an in-house X-ray setup.

Data Collection and Processing

X-rays are high-energy photons and the wavelengths of those used in macromolecular crystallography experiments lie in the 0.5 to 1.8 Å range (Blundell and Johnson,

1976; Woolfson, 1997; Rhodes, 2006). X-rays can, for example, be generated in sealed high-voltage tubes where an anode (Cu, Mb, Fe, etc.) is bombarded with electrons from a heated cathode filament. An electron is hitting the anode material, and, as it passes within proximity of an atom, the electron is attracted to the nucleus by the Coulombic force. This alters the trajectory of the electron, and the closer the electron to the nucleus, the greater the change in its trajectory. To conserve momentum, a photon is created, whereby the photon's energy depends on the degree to which the electron's trajectory was changed. The energy released in the form of photons is referred to as Brems-Strahlung ("braking radiation" or "white radiation"). Every now and then, an electron that hits the anode target is of sufficiently high energy to displace an electron from an inner shell (i.e., the K shell) and an electron from a higher shell (L, M, etc.) then takes its place, with the energy difference between them being emitted as monochromatic X-ray radiation. Normally X-rays are polychromatic, but monochromatic radiation can be obtained by way of a monochromator, for example a graphite crystal.

However, most of the energy is generated as heat and not "light," and X-rays from a sealed-tube setup (Fig. 7.13.19A) are typically not of high enough intensity for data collection with weakly diffracting macromolecular crystals. By comparison, so-called rotating anode units (Fig. 7.13.19B) feature an effective increase in the area of the anode target bombarded by accelerated electrons. However, the advantage in terms of higher intensity X-rays comes at a cost: rotating anode generators require more maintenance than sealed-tube setups as parts need to be replaced (cathode filament), cleaned (rotating anode), or rebuilt (ferrofluidic seal).

Today, most diffraction data collections are conducted at X-ray synchrotrons, where electron or positron beams are circling close to the speed of light in a storage ring (Fig. 7.13.20). X-rays are emitted in a tangential fashion when the beam is deflected by extremely strong electromagnets, so-called wigglers or undulators (Helliwell, 1992). Unlike the above sealed-tube or rotating anode generators that produce X-rays of a particular wavelength (i.e., CuKα = 1.5418 Å), the wavelength of the X-ray beam at synchrotrons is tunable. The availability of synchrotrons has had a major impact on structural biology and has impacted many other areas of research in a dramatic fashion

(Table 7.13.4; Hendrickson, 2000). The higher intensity of X-rays at synchrotrons leads to significant improvements in the resolution of diffraction data (>0.5 Å and more), but also causes radiation damage of crystals. Damage inflicted over the long run on a rotating anode source can occur in minutes on an unattenuated undulator beamline. Primary radiation damage is due to the large absorption cross-section of heavier atoms such as sulfur or selenium and secondary damage is caused by free radicals and photoelectrons.

To preserve crystals in the beam, they need to be flash-frozen and maintained near liquid nitrogen temperature in a cold stream during data collection (Fig. 7.13.21; Harp et al., 1998; Garman and Owen, 2006). Crystals mounted in capillaries (possible for neutron data collection; see *http://www.mitegen.com/* for rapid room temperature mounting) will not last very long in the beam. For flash-freezing, crystals are scooped up from a droplet with a nylon loop and then swiped through a cryoprotectant before being plunged into liquid nitrogen. The choice of cryoprotectant is important, as ice inside the loop formed during freezing will lead to diffuse scattering and powder pattern rings in diffraction images. Popular protectants are glycerol, sucrose, ethylene glycol, propylene glycol, low-molecular weight PEGs, MPD, and 2,3-butanediol. Very high concentrations of salts such as sodium malonate have also been reported to be suitable for cryoprotection. Crystals are then shipped to the synchrotron source in the frozen state inside so-called dry-shippers. Most macromolecular crystallography synchrotron beamlines are now equipped with automatic sample changers and some feature remote access, allowing users to collect data without leaving the office or the laboratory.

Prior to the actual data collection, a single or multiple test frames (Fig. 7.13.22) are recorded and indexed and the orientation matrix determined and refined. Once Bravais lattice type and Laue group are assigned, one needs to decide on the best data acquisition protocol. Important parameters are the angle of rotation (around the phi axis in most cases), exposure time, and the crystal-to-detector distance. In terms of the correct rotation angle, fine phi slicing guarantees a reduced background whereas coarse phi slicing is more suitable for rapid data collection. In cases where crystals diffract to very high resolution, it is necessary to collect separate low-, medium-, and high-resolution data sets, whereby proper
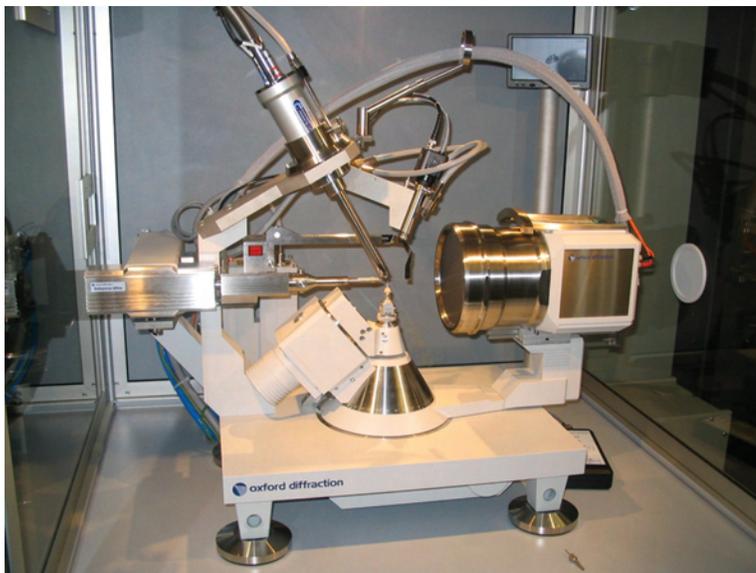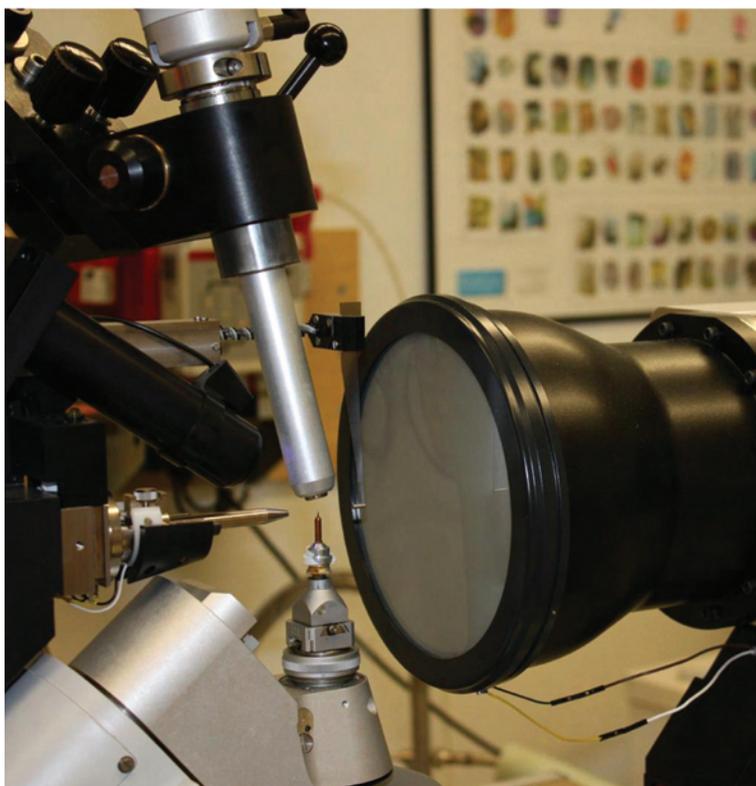
Biophysical Analysis of Nucleic Acids

7.13.21

Figure 7.13.19 X-ray generators and detectors. Two 4-circle, kappa-geometry X-ray diffraction setups currently used by researchers at Vanderbilt University. (**A**) The sealed tube Oxford Xcalibur PX2 Ultra (Oxford Diffraction), and (**B**) the rotating anode Bruker Microstar (Bruker AXS). Tube housing (Xcalibur), beam collimator, beam stop, CCD detector, crystal cooler, goniostat, goniometer head, and telescope are clearly visible.

Table 7.13.4 Impact of X-Ray Synchrotron Radiation

Dramatic improvements in resolution (<1 Å in some cases)

High-throughput X-ray crystallography (structural genomics)

Structural analysis of multi-component macromolecular machines

Micro- and 2D-crystallography (small crystals and membranes)

New phasing strategies (MAD)

Ultrafast time-resolved crystallography (ns resolution)

X-ray absorption spectroscopy (XAS, EXAFS)

Non-crystalline diffraction and small angle scattering

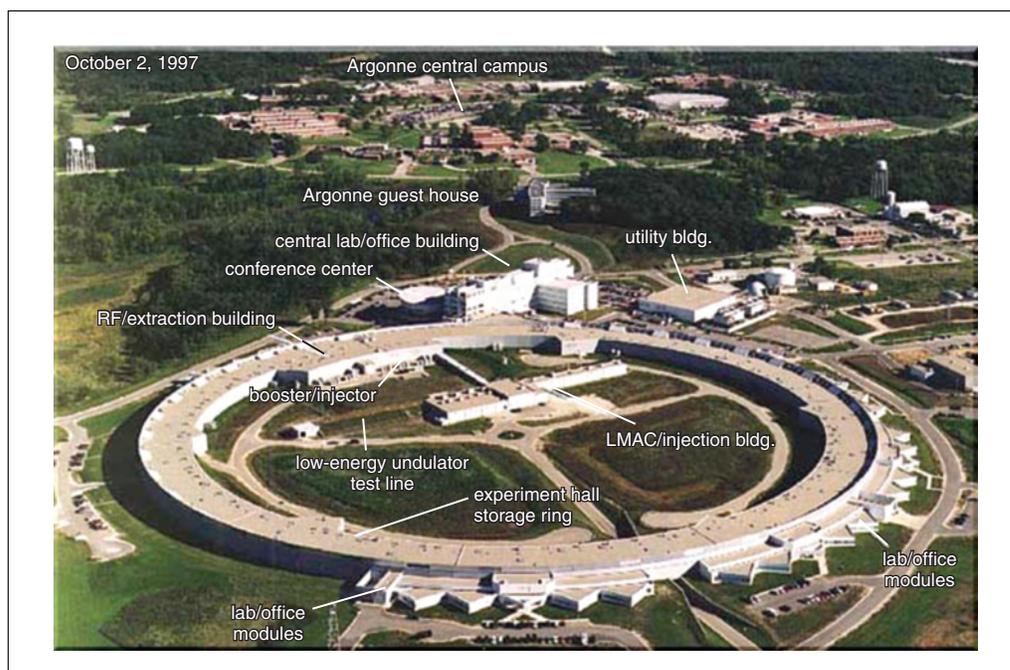X-ray microscopy of whole hydrated single cells

Material sciences

Figure 7.13.20 X-ray Synchrotron. Aerial view of the Advanced Photon Source (APS) at Argonne National Laboratory, a so-called 3rd generation X-ray synchrotron.

acquisition of low-resolution reflections may require an attenuated beam. In general, data collection is now a matter of hours, and as long as the crystal survives, it is better to collect too much data than too little. CCD detectors are used to record individual diffraction frames (Figs. 7.13.19, 7.13.21, and 7.13.22). These detectors offer several advantages over multiwire proportional counters or image plate area detectors, i.e., a linear response and high dynamic range, rapid readout, and high spatial resolution. Unlike standard data collections that use X-rays with a discrete wavelength in the rotation mode, Laue diffraction experiments employ "white" or polychromatic radiation with exposures in as little as 50 psec

for time-resolved structural studies. Such experiments are complicated by multiple intensities, variations in the absorption coefficient, an uneven detector response at varying wavelengths, and reflection spot overlaps, among others.

While the data collection is ongoing, the experimenter starts the data reduction. The reflections (spots) in the individual images or frames are indexed and the crystal and detector parameters are refined before the diffraction peaks are integrated, i.e., their intensities extracted. After establishing the relative scale factors between measurements, these parameters are once more refined using the total data set. Finally, the frames are merged and a
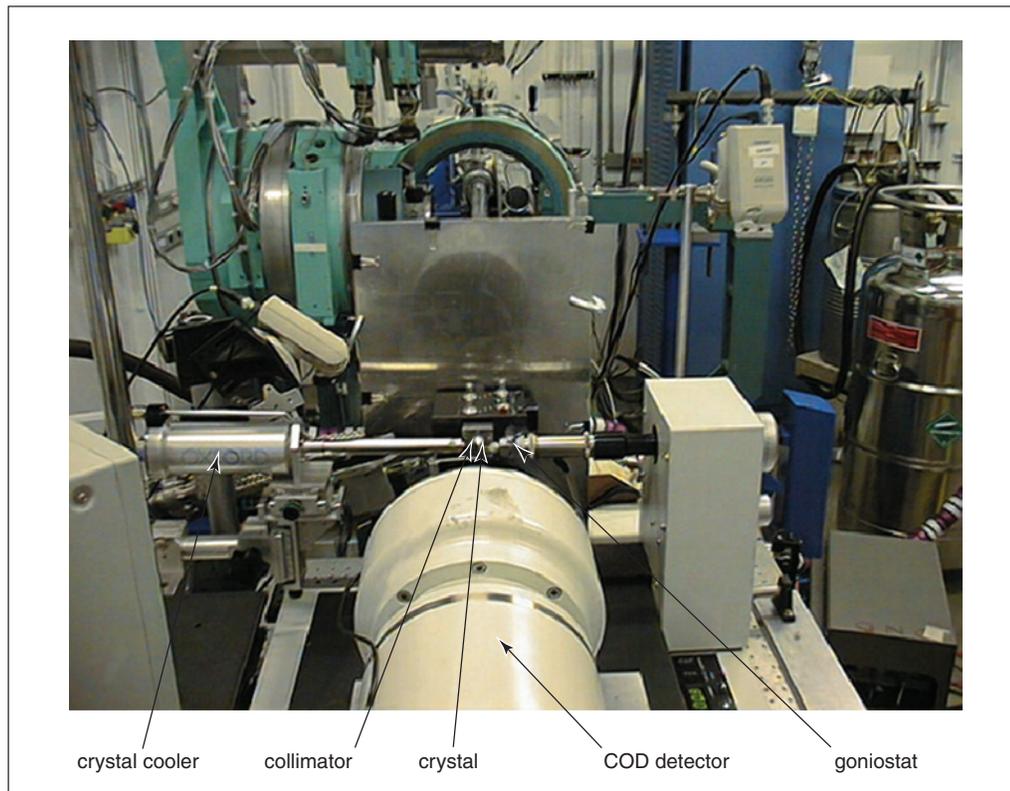
Biophysical Analysis of Nucleic Acids

7.13.23

| crystal cooler | collimator | crystal | COD detector | goniostat |

Figure 7.13.21 Inside a Synchrotron experimental station. The Marresearch charge-coupled device detector (MARCCD 225; foreground, *http://www.marresearch.com*) mounted on the MAR desktop beamline (DTB) at the insertion device beamline (5-ID-D hutch) of the DuPont-Northwestern-Dow collaborative access team (DND-CAT), located at sector 5 of the APS. The view is into the beam that is transported along the tube visible in the center of the upper half of the photograph. The instrumentation colored light blue in the background is not part of the macro-molecular crystallography setup. Work conducted at the DND-CAT now focuses more on surface and interface science, nano-materials, catalysis, and environmental science. The macromolecular crystallography efforts have moved to the new Life Sciences (LS-CAT) at sector 21 of the APS that offers four ID lines and is jointly run by Michigan institutions, Northwestern University, the University of Illinois at Urbana-Champaign, the University of Wisconsin, and Vanderbilt University. Further consortia that operate ID and/or Bending Magnet (BM) beamlines for macromolecular crystallography at the APS include BioCARS-CAT (sector 14), IMCA-CAT (sector 17), SBC-CAT (sector 19), SER-CAT (sector 22), GM/CA-CAT (sector 23), and NE-CAT (sector 24).

statistical analysis of reflections based on the space group symmetry is computed. An example of the completeness and quality of a diffraction data set broken down into resolution shells or bins is shown in Table 7.13.5. The final product of the diffraction experiment is a file with the amplitudes of individual reflections (the so-called structure factors, F_{obs}) and their standard deviations $\sigma(F_{obs})$. The R_{sym} represents the spread of equivalent reflections (the smaller the better) and the resolution limit can be estimated from the mean$[I/\sigma(I)]$ ratio (the highest resolution shell included should have a mean$[I/\sigma(I)] \geq 2$) and/or the completeness of the data in a higher shell (i.e., $> 70\%$ in the outermost shell).

Phasing Approaches

Unfortunately, the measured structure factor amplitudes alone are insufficient for building a structural model. The Fourier transformation of the diffraction pattern that is needed to generate the crystal structure (expressed in terms of an electron density distribution) requires both the amplitudes and the phases of structure factors (Blundell and Johnson, 1976; McRee, 1999; Woolfson, 1997; Rhodes, 2006; Drenth, 2007). However, the phase information is lost in the diffraction experiment. Contrary to data collections that are rapid and more or less automatic, the determination of a structure can therefore still be a time-consuming challenge. There are four basic techniques
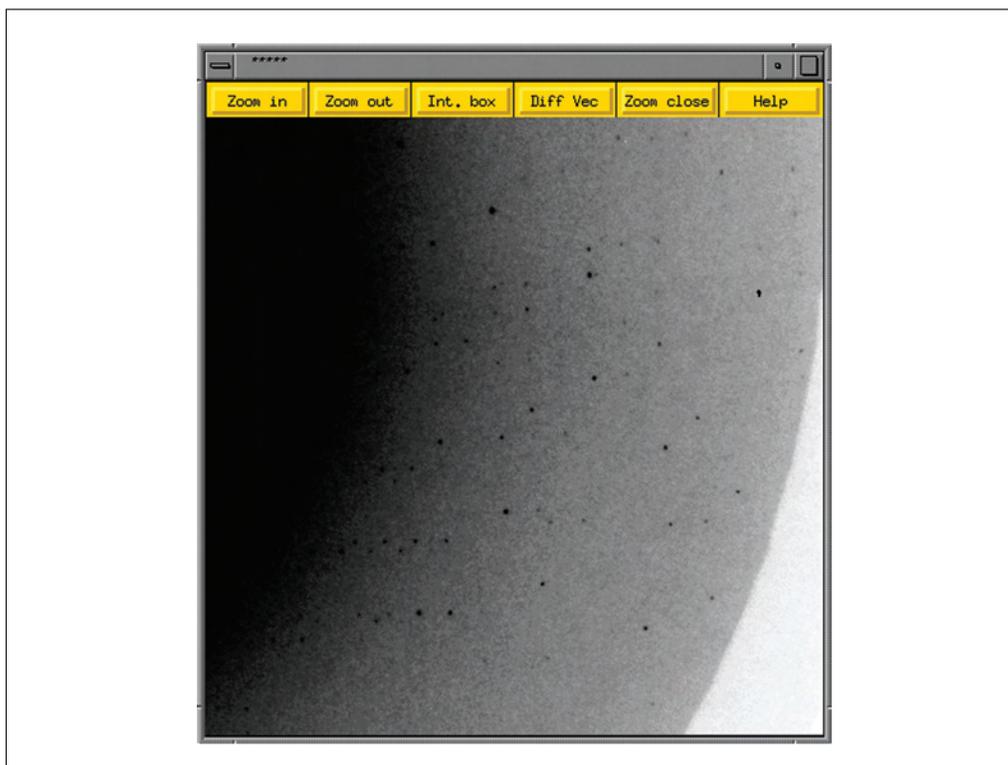
Figure 7.13.22 Diffraction data collection. Close-up of a 1° (ΔPhi) diffraction image obtained from a single crystal of the so-called Dickerson Drew Dodecamer (DDD; B-form DNA of sequence CGCGAATTCGCG). The dark spots represent individual reflections and the diffraction limit is around 1 Å. Data statistics for this particular crystal of the DDD are listed in Table 7.13.5.

Table 7.13.5 Diffraction Data Quality: Breakdown into Resolution Bins

Resolution [Å]	N (unique)	Mean [I/σ (I)]	Completeness [%]	$R_{sym}{}^a$
20.00-3.00	1436	22.4	98.8	0.067
3.00-2.50	997	26.1	99.8	0.059
2.50-2.00	2218	24.5	99.5	0.063
2.00-1.80	1626	18.5	97.4	0.049
1.80-1.60	2578	18.5	99.0	0.049
1.60-1.40	4282	16.0	100.0	0.064
1.40-1.20	7524	13.7	100.0	0.085
1.20-1.10	6060	8.5	99.7	0.154
All data	26,721	15.5	99.5	0.064

$^a R_{sym} = \Sigma_{hkl}\Sigma_i |I(hkl)_i - <I(hkl)>|/\Sigma_{hkl}\Sigma_i <I(hkl)_i>$

The R_{sym} is a measure for the similarity of the intensities of symmetry-equivalent reflections. It should be a small as possible, typically around 5% or below. Another parameter, R_{merge}, is used to characterize the similarity of corresponding reflections in different data sets (i.e. low- and high-resolution data sets) or in data sets from different crystals.

for solving the phase problem with crystals of macromolecules: multiple isomorphous replacement (MIR), multi-wavelength anomalous dispersion (MAD) and a combination of the two (MIRAS), molecular replacement (MR), and direct methods (DM). Molecular replacement requires a good model structure and it is the method of choice for complexes of the same enzyme with different ligands (i.e., inhibitors) or multi-domain proteins for which the structure of a domain is available (i.e., fusion proteins). Particularly with crystals of oligonucleotide duplexes, one is often tempted to perform rotation and translation searches

Biophysical Analysis of Nucleic Acids

7.13.25

using A- or B-form models. However, the failure rate is quite high and relatively small deviations between the conformations of the model and the actual structure are sufficient to derail the search.

Direct methods are model-independent, but will only work in cases for which diffraction data to very high resolution are available (<1.0 Å). In addition, there is a size limit and the structure of a 100-kDa protein is unlikely to be phased by DM even with crystals diffracting to atomic resolution. Of the 50,000 or so structures of proteins currently deposited in the Protein Data Bank, $<0.5\%$ were determined at resolutions of 1 Å and higher. Unlike with crystal structures of small molecules that are mostly solved by DM, the approach is not likely to replace MAD or MIR as the standard phasing techniques for new macromolecular structures in the near future (Terwilliger and Berendzen, 1999; Weeks et al. 2003).

Both MIR and MAD require derivatization of a macromolecule, which is the introduction of heavy atoms into the crystal lattice. Heavy atoms can be bound covalently or by coordination and can be incorporated synthetically (nucleic acids), covalently during protein expression (selenium), by co-crystallization, soaking of native crystals, or in a pressure cell (xenon). A key difference between MAD and MIR is the requirement with the latter that native crystal and derivative crystals (two but better more derivatives are needed for MIR) are closely isomorphous. In this context, it is noteworthy that highly similar unit cell constants are not necessarily an indication that the orientations of the protein or nucleic acid in two crystals are identical. The classic approach for introduction of heavy atoms is soaking, and a resource for heavy-atom derivatization can be found at *http://www.sbg.bio.ic.ac.uk/had/heavyatom.html* (Heavy Atom Databank). Among the favorites for proteins are mercurial compounds (binding to free cysteines or methionine) and platinum compounds [binding mainly to methionine, histidine, and cysteine; $Pt(CN)_2$ binds to positively charged residues; Petsko, 1985; Rould, 1997; Garman and Murray, 2003]. The heavier the atom the better since the scattering amplitude is proportional to the number of electrons.

Several classes of heavy atoms can be differentiated: single metal ions are bound electrostatically, endogenous metal ions such as zinc in zinc fingers and iron in heme that can be used directly for phasing or substituted to obtain a larger signal (i.e., Sr^{2+} for Ca^{2+}),

compounds requiring a chemical reaction, multi-metal complexes for larger molecules (for example the tantalum bromide cluster), xenon and krypton, and anions such as halides or triiodide. For nucleic acids, Rb^+, Sr^{2+}, Ba^{2+} (Tereshko et al., 2001), and Tl^+ (Conn et al., 2002) are particularly useful and helix engineering for generating a coordination site for Co (III) hexamine has been used for large RNAs (Keel et al., 2007), as have lanthanides (Holbrook and Kim, 1985; Kim et al., 1985). For soaking, it is important to establish a suitable stabilizing solution or artificial mother liquor. The crystal is then transferred to the stabilizing solution that contains the heavy atom at a concentration of typically <1 to 10 mM. Occasionally, cracks or ragged edges develop and crystals need to be tested for diffraction at various time intervals, whereby it is useful to have a diffraction image prior to soaking for comparison. There are various ways to determine whether the heavy atom is indeed bound. A color change or cracking may be taken as evidence for binding. Mass spectrometry or MicroPIXE (particle induced X-ray emission microprobe) can also be used to confirm successful derivatization. Ultimately, the experimental determination of difference Patterson peaks (there are various means to retrieve the locations of heavy atoms) is the best proof for a useful derivative that paves the way to an interpretable electron density map.

Contrary to MIR, MAD phasing can be accomplished with a single derivative and the technique has gained widespread popularity in the past 15 years and now accounts for the majority of newly determined structures (Hendrickson, 2000). However, since diffraction data need to be collected at the absorption peak of a particular anomalously scattering atom (Fig. 7.13.23), MAD or the related SAD (single-wavelength anomalous dispersion) experiments have to be performed at a synchrotron source.

The most common anomalous scatterer for proteins is selenium, which can be incorporated as Se-methionine in *E. coli* using an auxotrophic strain or metabolic inhibition (Hendrickson et al., 1990; Doublié, 1997). High concentrations of isoleucine, lysine, and threonine are known to block methionine synthesis in *E. coli* by inhibiting aspartokinases. In addition, phenylalanine and leucine act in synergy with lysine. Thus, growth in a medium lacking methionine but supplemented with Se-methionine and plenty of the methionine pathway inhibitors allows for efficient incorporation of the Met analog. Se-Met
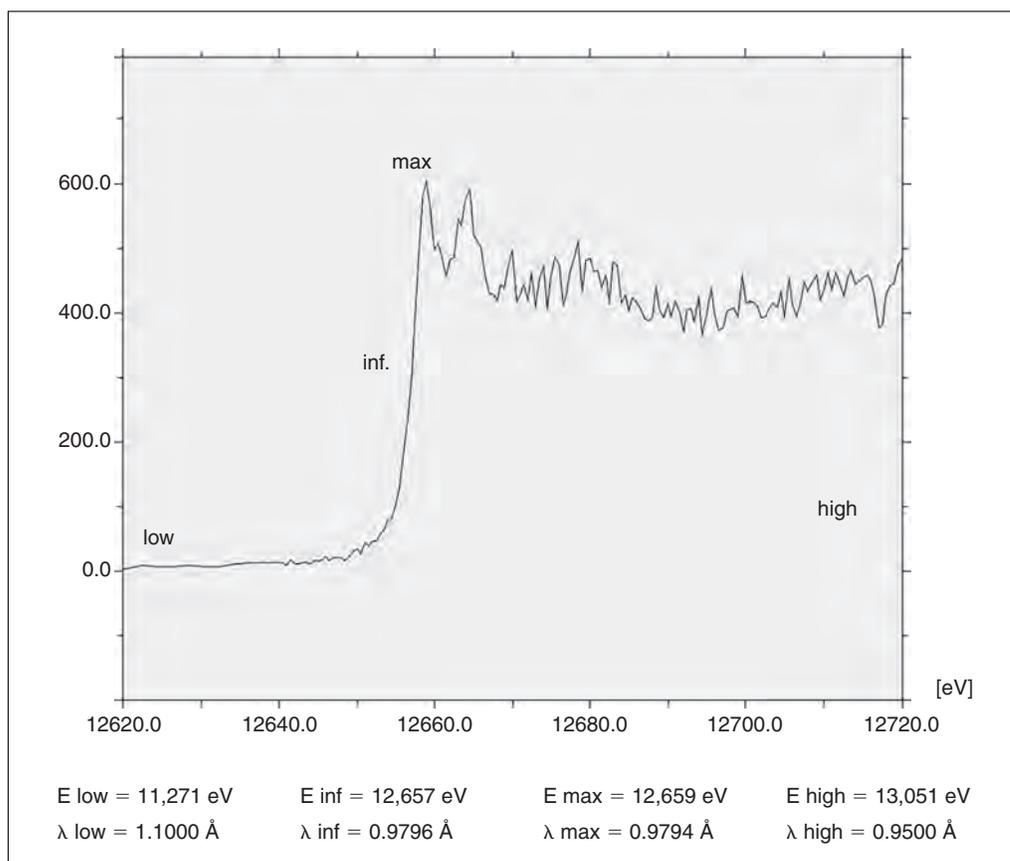
Diffraction Techniques in Structural Biology

7.13.26

Figure 7.13.23 Multiwavelength anomalous dispersion (MAD) experiment. Example of an X-ray fluorescence spectrum from a protein crystal that contains Se-methionine (Se-Met) in place of Met (Maf protein from *Bacillus subtilis*; 189 amino acids and 6 Se atoms per protein molecule). The theoretical K absorption edge of selenium lies at 12.6578 keV or 0.9795 Å (an energy of 12.398 keV corresponds to 1.0 Å; *http://skuld.bmsc.washington.edu/scatter/*). In a typical MAD experiment, diffraction data of high redundancy from the same crystal are collected at three or four wavelengths (i.e., reference below the edge, low, inflection point, inf., peak, max, and reference above the edge, high).

derivatization does not always work and a number of caveats need to be considered.

Selenium is toxic and so cells will not grow as fast. Se-Met derivatized proteins are often less soluble and the altered solubility can affect crystallization. Selenium is also easily oxidized and this may blur the absorption edge or render phasing more difficult. Moreover, it is crucial to precisely determine the peak of the anomalous absorption signal with a particular crystal on the beamline using a fluorescence detector (Fig. 7.13.23). Very tiny deviations from the maximum may subsequently lead to failure in locating the anomalous scatterers or adversely affect the quality of the MAD electron density map (for a successful example, see Fig. 7.13.24). Bromine is the most popular anomalous scatterer for derivatization of crystals of nucleic acids or protein-nucleic acid complexes, and can be covalently incorporated in the form of Br^5U or Br^5C. Naturally, many other heavy atoms are not only useful for MIR but can also serve as anomalous scatterers. For example, most crystals of oligonucleotides are grown in the presence of alkaline earth metal ions and it is advisable to always collect MAD data with crystals that contain Sr^{2+} or Ba^{2+}. This is because the common assumption that structures of oligonucleotide fragments typically yield to phasing by MR is incorrect. Selenium has also been covalently incorporated into nucleic acids for structure determination via SAD or MAD (Wilds et al., 2002; Pallan and Egli, 2007a,b).

Accurate phases are very important, as they influence the quality of the experimental electron density and without accurate density, it is impossible to build a model. MAD phasing has the advantage that the derivative does not have to be isomorphous with the native crystal. Once the model based on the, say, Se-Met protein structure is built and refined, it can be used
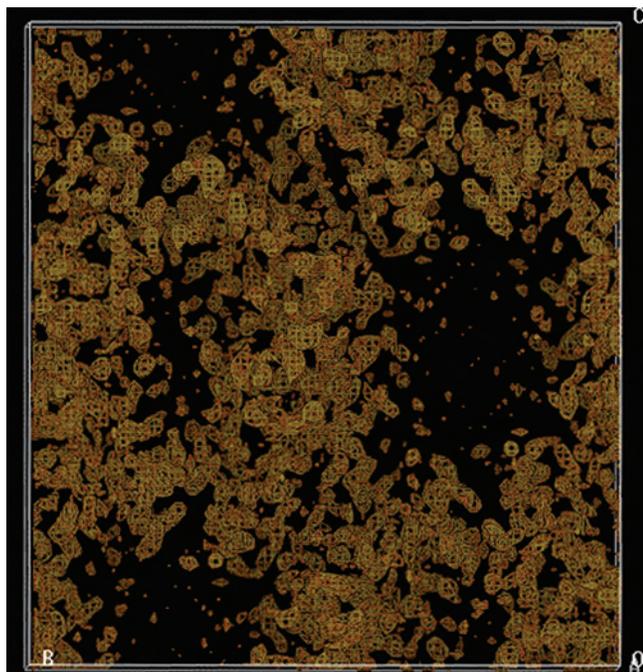
Biophysical Analysis of Nucleic Acids

7.13.27

Figure 7.13.24 MAD phasing. Experimental electron density map based on five Se sites obtained from a Se-Met crystal of the Maf protein (2.7 Å resolution, no solvent flattening), calculated with the program SOLVE (Terwilliger and Berendzen, 1999). The map displays clear boundaries and reveals large solvent-filled channels (black regions). For the color version of this figure go to *http://www.currentprotocol.com/protocol/nc0713*.

to solve the native crystal structure via MR if the two are not isomorphous. However, that is not always necessary, and one may decide to just use the structure of the Se-Met protein unless the native dataset is of higher resolution. MAD electron density maps are often of excellent quality, making it possible to automatically trace the protein and build an initial model.

Thus, it is not uncommon to end up with a preliminary model of a protein within hours of completing data collection; however, in most cases the initial electron density needs to be improved. This is achieved by improving the phases, since they are the terms with the largest amount of error in the Fourier transformation. Inaccuracies in the phases dominate those in the amplitudes with regard to the quality of the electron density. The general approach to improving the phase information is to apply constraints in real space; this is referred to as density modification. Density modification methods commonly used are solvent flattening (and flipping), noncrystallographic symmetry averaging (multiple molecules per asymmetric unit that are not related by crystallographic symmetry, e.g., in viruses), histogram matching, phase combination and extension, and

the maximum likelihood approach (Carter and Sweet, 1997a,b; Carter, 2003a,b).

Refinement and Analysis

The model built into the experimental density typically represents just a rough approximation, and to arrive at a final structure it is necessary to refine it. Each atom in the model is represented by coordinates x, y and z, an occupancy parameter ($q \leq 1$), and a temperature factor (B-factor). The atomic coordinates are stored in a file of a particular format, i.e., the so-called PDB format. The objective of crystallographic refinement is to apply changes to the atomic model such that the difference between model (represented by calculated structure factors F_{calc}) and the observed structure factors F_{obs} are minimized. The R-factor is a measure for the deviations between the calculated and observed amplitudes:

$$\text{R-factor} = \Sigma_{hkl} \; || F(hkl)_{obs} | - | F(hkl)_{calc} ||/ \Sigma_{hkl} |F(hkl)_{obs}|$$

whereby h, k, and l represent the Miller indices, the coordinates of reflections in reciprocal space.

Refinement is an iterative process that entails the following basic steps: manual

building and (re)fitting, automatic constrained least squares optimization taking into account both X-ray data and geometric constraints of the physical model, and electron density map calculation from the improved model (so-called Fourier sum and difference electron density maps), followed by additional building and so forth. The model will profit from a large excess of reflection data over the number of parameters (x, y, z, q, B) that define the model. A ratio of, say, 10 would be considered excellent, and a ratio of 2 represents a poorly over-determined structure. To reduce the total number of parameters that need to be refined, stereochemical restraints are applied (i.e., bond length, bond angle, torsion angle, planarity, chirality, van der Waals distances). The restraints are entered as terms in the refinement target and are weighted so that the deviations from ideal values match those found in databases of high-resolution structures. Thus, the target function is an energy that consists of an X-ray (F_{obs}, F_{calc}) and an empirical term (bonds, angles, van der Waals contacts etc.), and optimization algorithms such as steepest descent or conjugate gradient are used to find the nearest minimum in the target function.

To escape local energy minima in the target function and to improve the radius of convergence, simulated annealing (molecular dynamics, MD) is used (Brunger and Adams, 2002). Atoms are given random starting velocities and their motion is modeled according to Newton's laws of motion (bond stretching and angle bending). The temperature of the system is increased (to 2000°C or more) with periodic cooling (annealing), followed by energy minimization. The MD equations are modified through addition of crystallographic residual to the empirical potential energy. Overall, the random element and the thermal motion help to overcome local minima in the target function. Another variant of the least squares optimization is maximum likelihood (Murshudov et al., 1999). Its basic premise is that refinement is not just a matter of making F_{calc} equal to F_{obs}, but also needs to consider the phases. To decide how to move an atom we need to take into account the overall accuracy of the model and the best model is consistent with all observations. Consistency is measured statistically by the probability that these observations would be made given the current model. The probabilities include all sources of error (including the model) and as the model gets better, errors get smaller and probabilities become sharper, which in turn increases the likelihood.

The R-factor serves as one guide for the status of the refinement. An R-factor of ~60% is consistent with a random relation between the observed and calculated amplitudes. A good starting model will have an R-factor of 40% to 45% and a final model of a macromolecular structure may exhibit an R-factor of ~20%. During the refinement ($2F_{obs}$-F_{calc}), sum electron density maps should look like the corrected model, although they can be biased by incorrect phases/models. On the other hand (F_{obs}-F_{calc}), difference electron density maps will indicate missing or incorrectly placed atoms. So-called omit maps can be used to remove phase bias that results from least-squares refinement using wrong coordinates. These are difference electron density maps calculated after removing a part of the model from the calculation of F_{calc} amplitudes. Because nearby atoms have been influenced by the incorrect portions, the "memory" associated with the omitted atoms needs to be removed. This is achieved by annealed omit maps that are calculated after removal of specific portions of the structure and additional MD. A composite omit map can be generated by placing a 3D grid over the entire unit cell and removing one grid box at the time, calculating the F_{calc}, and then repeating this for all grid boxes and summing over all grid points.

An independent measure of the quality of the fit is provided by the R-free, an R-factor that is based on a test data set, reflections (typically amounting to 5% of the total diffraction data) that are set aside and are not included in the refinement (Brunger, 1992). The R-free will be higher than the R-factor (i.e., by up to 5%) and an R-free of 30% with an R-factor (also called R-work) of 20% may indicate errors or over-refinement. Obviously, model building and refinement are easier with high-resolution data. Figure 7.13.25 depicts sum electron densities around an aromatic moiety at different resolutions and it is obvious that a map at 3 Å offers some challenges to the model builder. Other parameters beyond R-factors and resolution that need to be considered for judging quality and correctness of a structure are the root mean square deviations (r.m.s.d.s) of bond lengths and angles from standard values (should be <0.02 Å and 3°, respectively) and the B-factors (portions of a structure with atoms displaying B-factors >50 Å^2 indicate weak electron density). With crystal structures of proteins, the so-called Ramachandran plot (Fig. 7.13.26) can be used to
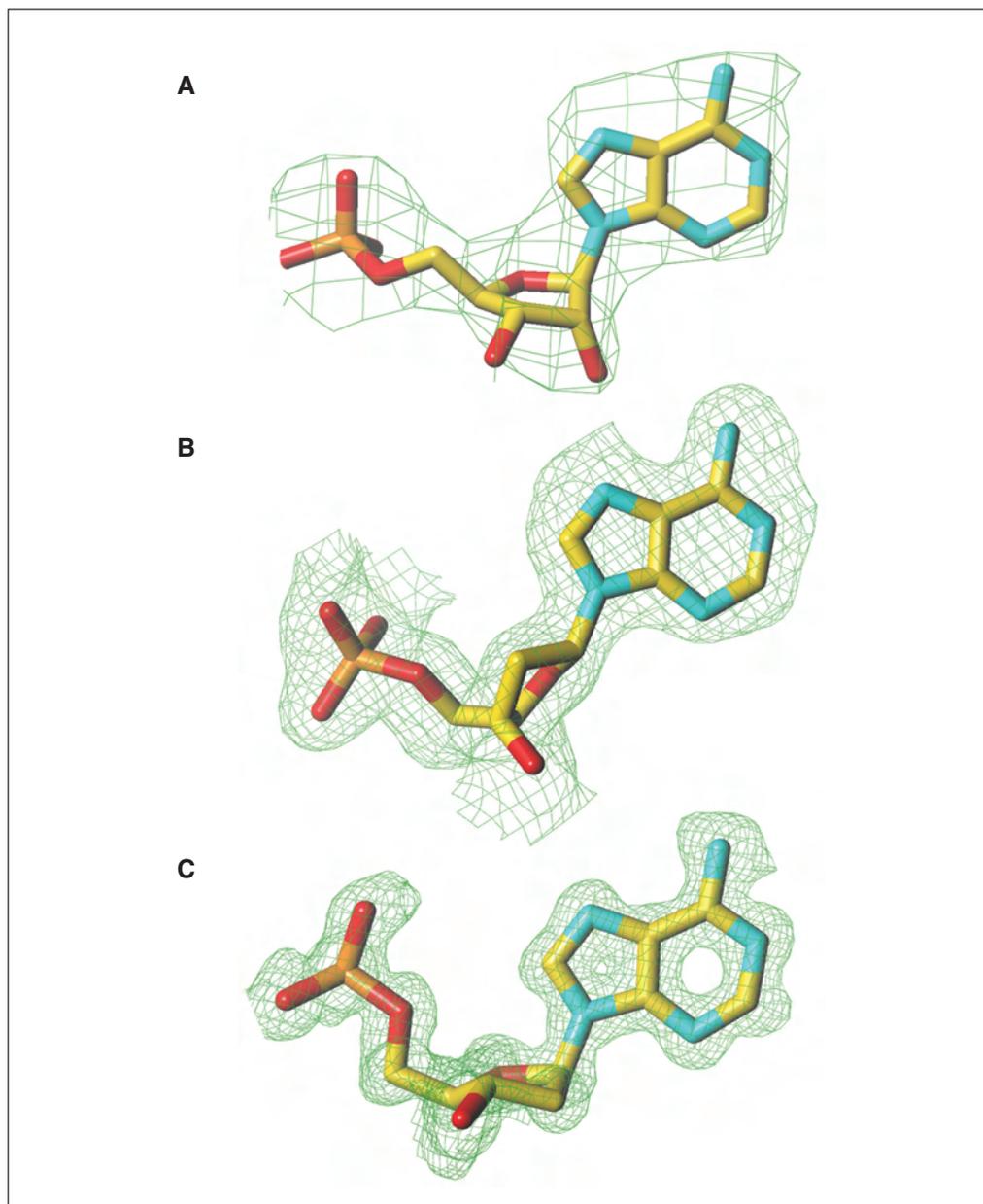
Figure 7.13.25 Resolution and quality of the electron density. Comparison of the quality of the Fourier ($2F_o$-F_c) sum electron density around adenosine monophosphate in crystal structures obtained at various resolutions (1σ threshold). (**A**) 2.85 Å, ATP in the crystal structure of the KaiC protein from *Synechococcus elongatus*; (**B**) 1.80 Å, A residue in the crystal structure of a B-form DNA; (**C**) 1.10 Å, atomic resolution, A residue in the crystal structure of an A-form DNA. For the color version of this figure go to *http://www.currentprotocol.com/protocol/nc0713*.

pinpoint problematic areas in a structure based on deviations of the backbone torsion angles from commonly encountered values.

It is important to realize that crystallographic models often lack parts of a protein or nucleic acid sequence. The N- and C-terminal portions of a protein are normally more flexible than the core as are the terminal nucleotides in DNA duplexes or single-stranded regions in RNAs. In the crystal structure of *E. coli* DNA polymerase I (Klenow fragment), ~10% of the

amino acids are missing because they could not be seen in the electron density map at ~2.5 Å. This indicates that proteins packed into a crystal lattice can still retain considerable flexibility. Indeed, some enzymes are active in the crystalline state and enzymatic reactions have been studied using Laue crystallography. Along with protein or nucleic acid, crystals contain a lot of water (in some cases crystals consist of 70% to 80% water), and the final model consists not just of the coordinates of
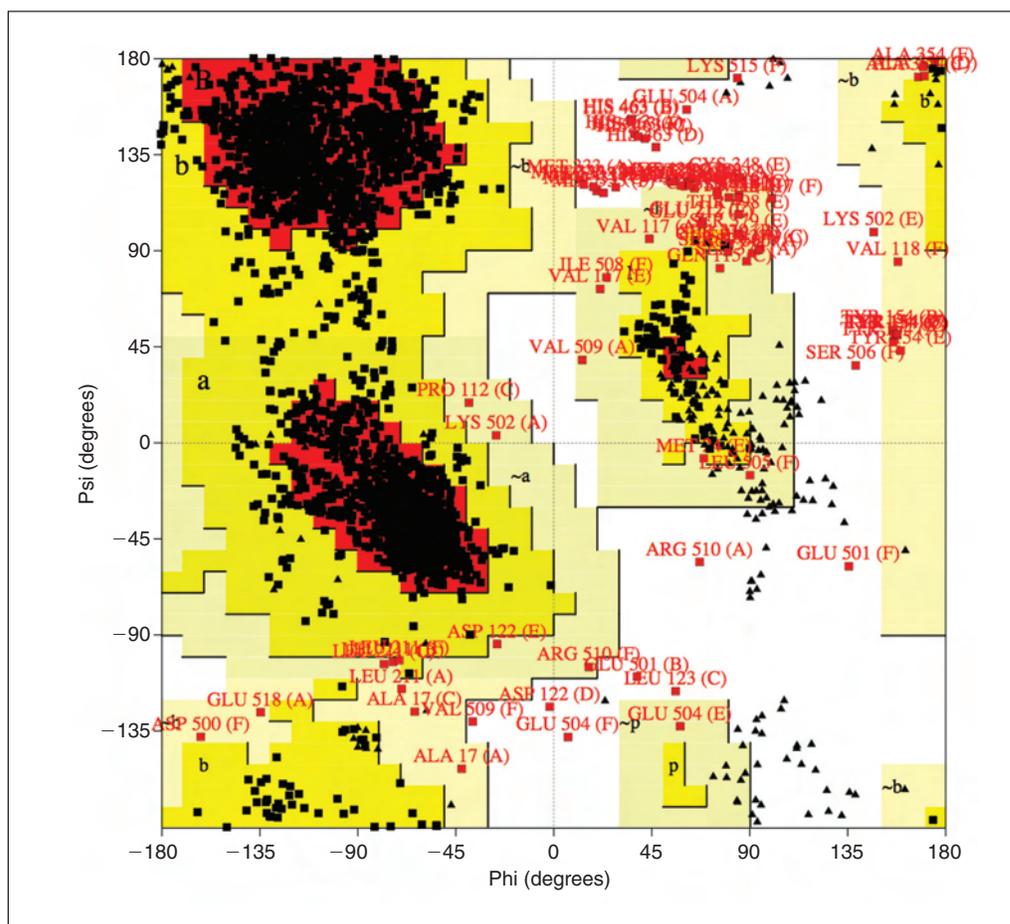
Figure 7.13.26 Structure refinement and quality control. Conformations of the Φ and Ψ backbone torsion angle pairs (Ramachandran plot) for amino acids in the crystal structure of the KaiC clock protein from *S. elongatus* (PDD ID code 2GBL) (CCP4, 1994). Individual angles fall into the most favored regions (A, B, L; red [84%]), additionally allowed regions (a, b, l, p; yellow [13.3%]), generously allowed regions (~a, ~b, ~l, ~p; faint yellow [1.8%]), or disallowed regions (white [1%]). There are 244 glycine (shown as triangles) and 74 proline residues. For the color version of this figure go to *http://www.currentprotocol.com/protocol/nc0713*.

protein atoms but many first and second shell water molecules, ions, and other cosolutes.

Crystal packing forces obviously have an effect on the structure of a macromolecule and need to be considered in the conformational analysis of a protein. Rather than cursing them, lattice forces should be considered a blessing, as they can provide valuable information on the deformability of a loop region or particular features of the interface between a protein and its interacting partner. Apart from anisotropic B-factors from data at very high resolution, crystal structures typically provide mostly static information. Occasionally, two or more crystal forms are available, however, allowing one to sample multiple conformations of the same molecule. In such cases, it is possible to determine how packing forces affect the structure of a protein and to identify flexible regions and relative motions of domains.

SUMMARY AND OUTLOOK

Progress in structural biology over the last quarter of a century has been dramatic on all fronts, including instrumentation, mechanistic insights into ever-larger molecules and multiprotein complexes, and the automation of individual steps on the way to a structure determination, (for those involved in a crystal structure analysis, see Fig. 7.13.15). The increasing complexity of the problems being tackled has led to the recognition that one technique alone cannot possibly provide all the answers and has motivated researchers to apply hybrid structural approaches, i.e., combinations of single crystal X-ray crystallography and cryo-EM, crystallography and SAXS, or NMR and computational simulations (computational biology has not been discussed beyond applications in crystallography in this unit).

Biophysical Analysis of Nucleic Acids

7.13.31

Looking into the crystal ball, one can see significant developments in the area of X-ray synchrotron sources in the future, with the emergence of so-called compact light sources (tabletop synchrotron) based on the free-electron laser (FEL) process. In an FEL, electrons traveling at nearly the speed of light make their way through an undulator magnet where they are accelerated, resulting in the release of photons. Electrons continue to move in phase with the field of the light emitted and the fields add together in a coherent fashion. The wavelength of the resulting X-ray beam of high brilliance can be tuned by changes in the magnetic field strength of the undulators or the energy of the electron beam. This setup precludes the need for a large storage ring (Fig. 7.13.20) and the equipment could be housed on university campuses or in medical centers, thus allowing users local access to X-ray synchrotron radiation. Beyond crystallography, applications could include material science, single-molecule X-ray diffraction (Hajdu, 2000), imaging, and surgery. For additional information see the following Web sites: *http://www.lynceantech. com/sci_tech_cls.html* and *http://www.photon-production.co.jp/e/PPL-HomePage.html*).

Automation of protein expression and purification, crystallization, data collection and structure determination, and model building will continue, driven by the need for high-throughput crystallography as part of structural genomics projects and drug discovery. A decade of large-scale structure determination of proteins has had a major impact on technological advances that have clearly benefited traditional structural biology projects. However, the expectation that one may have had regarding potential outcomes of the PSI, namely that function could be gleaned from structure alone, has not been fulfilled in most cases (Chandonia and Brenner, 2006; Terwilliger et al., 2009). A more likely scenario is that structural information deposited in publicly accessible databases and improved data sharing in combination with biochemical, mutational, and genetic studies (that are perhaps initiated by the structural data) will allow the classification of proteins of unknown function at an increased pace.

The achievements made in terms of the structural characterization of soluble proteins, RNA, molecular machines, multisubunit complexes, and others cannot detract from the fact that there are areas where progress has been slower and significant challenges remain. An example that comes to mind is the membrane protein field. Several structures have been determined including photo systems, ion channels, and the first G-coupled protein receptors (GPCRs) in 2008. However, expression of stable constructs of membrane proteins in amounts suitable for structural characterization, solubilization, and crystallization still constitute formidable obstacles on the way to a more routine generation of structural data. Capturing dynamic systems involving formation of relatively labile protein-protein complexes represents another frontier of structural biology (Radaev and Sun, 2002; Dafforn, 2007). One such system studied in the laboratory of the author is the minimal circadian clock from the cyanobacterium *S. elongatus* that can be reconstituted in vitro from three proteins in the presence of ATP. The KaiA, KaiB, and KaiC proteins interact to form complexes of different compositions throughout the 24-hour cycle, whereby the concentrations of the free proteins and the respective complexes oscillate (Johnson et al., 2008). Clearly, only by using hybrid structural approaches such as those outlined above can one expect to make headway with regard to a structural dissection of the clock and a better understanding of its mechanism.

ACKNOWLEDGMENTS

LITERATURE CITED

Abola, E., Kuhn, P., Earnest, T., and Stevens, R.C. 2000. Automation of X-ray crystallography. *Nat. Struct. Biol.* 7:973-977.

Baeyens, K.J., Jancarik, J., and Holbrook, S.R. 1994. Use of low-molecular-weight polyethylene glycol in the crystallization of RNA oligomers. *Acta Cryst. D* 50:764-767.

Baumeister, W. and Steven, A.C. 2000. Macromolecular electron microscopy in the era of structural genomics. *Trends Biochem. Sci.* 25:624-631.

Berger, I., Kang, C.H., Sinha, N., Wolters, M., and Rich, A. 1996. A highly efficient 24-condition matrix for the crystallization of nucleic acid fragments. *Acta Cryst. D* 52:465-468.

Berman, H.M., Olson, W.K., Beveridge, D.L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A.R., and Schneider, B. 1992. The nucleic acid database: A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* 63:751-759.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The protein data bank. *Nucleic Acids Res.* 28:235-242.

Blakeley, M.P., Langan, P., Niimura, N., and Podjarny, A. 2008. Neutron crystallography: Opportunities, challenges and limitations. *Curr. Opin. Struct. Biol.* 18:593-600.

Blundell, T.L. and Johnson, L.N. 1976. Protein Crystallography. Academic Press Inc., New York.

Brunger, A.T. 1992. The free *R* value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355:472-474.

Brunger, A.T. and Adams, P.D. 2002. Molecular dynamics applied to X-ray structure refinement. *Acc. Chem. Res.* 35:404-412.

Carter, C.W. Jr. (ed.) 2003a. Macromolecular Crystallography, Part C. Methods Enzymology, Vol. 368. Academic Press Inc., New York, NY.

Carter, C.W. Jr. (ed.) 2003b. Macromolecular Crystallography, Part D. Methods Enzymology, Vol. 374. Academic Press Inc., New York, NY.

Carter, C.W. Jr. and Sweet, R.M. (eds.) 1997a. Macromolecular Crystallography, Part A. Methods Enzymology, Vol. 276. Academic Press Inc., New York, NY.

Carter, C.W. Jr. and Sweet, R.M. (eds.) 1997b. Macromolecular crystallography, part B. Methods Enzymology, Vol. 277. Academic Press Inc., New York, NY.

CCP4. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Cryst. D* 50:760-763.

Chandonia, J.M. and Brenner, S.E. 2006. The impact of structural genomics: Expectations and outcomes. *Science* 311:347-351.

Chiu, W., Burnett, R.M., and Garcea, R.L. (eds.) 1997. Structural Biology of Viruses. Oxford University Press, Oxford, UK.

Collins, K.D. 2004. Ions from the Hofmeister series and osmolytes: effects on proteins in solution and in the crystallization process. *Methods* 34:300-311.

Conn, G.L., Gittis, A.G., Lattman, E.E., Misra, V.K., and Draper, D.E. 2002. A compact RNA tertiary structure contains a buried backbone – K^+ complex. *J. Mol. Biol.* 318:963-973.

Czepas, J., Devedjiev, Y., Krowarsch, D., Derewenda, U., Otlewski, J., and Derewenda, Z.S. 2004. The impact of Lys→Arg surface mutations on the crystallization of the globular domain of RhoGDI. *Acta Cryst. D.* 60:275-280.

Dafforn, T.R. 2007. So how do you know you have a macromolecular complex? *Acta Cryst. D.* 63:17-25.

Dong, A., Xu, X., Edwards, A.M., Midwest Center for Structural Genomics; Structural Genomics Consortium, Chang, C., Chruszcz, M., Cuff, M., Cymborowski, M., Di Leo, R., Egorova, O., Evdokimova, E., Filippova, E., Gu, J., Guthrie, J., Ignatchenko, A., Joachimiak, A., Klostermann, N., Kim, Y., Korniyenko, Y., Minor, W., Que, Q., Savchenko, A., Skarina, T., Tan, K., Yakunin, A., Yee, A., Yim, V., Zhang, R., Zheng, H., Akutsu, M., Arrowsmith, C., Avvakumov, G.V., Bochkarev, A., Dahlgren,

L.G., Dhe-Paganon, S., Dimov, S., Dombrovski, L., Finerty, P. Jr., Flodin, S., Flores, A., Gräslund, S., Hammerström, M., Herman, M.D., Hong, B.S., Hui, R., Johansson, I., Liu, Y., Nilsson, M., Nedyalkova, L., Nordlund, P., Nyman, T., Min, J., Ouyang, H., Park, H.W., Qi, C., Rabeh, W., Shen, L., Shen, Y., Sukumard, D., Tempel, W., Tong, Y., Tresagues, L., Vedadi, M., Walker, J.R., Weigelt, J., Welin, M., Wu, H., Xiao, T., Zeng, H., and Zhu, H. 2007. *In situ* proteolysis for protein crystallization and structure determination. *Nat. Methods* 4:1019-1021.

Doublié, S. 1997. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* 276:523-530.

Doudna, J.A., Grosshans, C., Gooding, A., and Kundrot, C.E. 1993. Crystallization of ribozymes and small RNA motifs by a sparse matrix approach. *Proc. Natl. Acad. Sci. U.S.A.* 90:7829-7833.

Drenth, J. 2007. Principles of Protein X-Ray Crystallography, 3rd ed. Springer Science and Business Media LLC, New York.

Ducruix, A. and Giegé, R. (eds.) 1992. Crystallization of Nucleic Acids and Proteins: A Practical Approach, 2nd ed. Oxford University Press, Oxford, UK.

Ferré-D'Amaré, A.R., Zhou, K., and Doudna, J.M. 1998a. A general module for RNA crystallization. *J. Mol. Biol.* 279:621-631.

Ferré-D'Amaré, A.R., Zhou, K., and Doudna, J.M. 1998b. Crystal structure of a hepatitis delta virus ribozyme. *Nature* 395:567-574.

Gait, M.J. (ed.) 1984. Oligonucleotide synthesis: A practical approach. IRL Press Ltd., Oxford, UK.

Garman, E.F. and Murray, J.W. 2003. Heavy-atom derivatization. *Acta Cryst. D* 59:1903-1913.

Garman, E.F. and Owen, R.L. 2006. Cryocooling and radiation damage in macromolecular crystallography. *Acta Cryst. D* 62:32-47.

Hajdu, J. 2000. Single-molecule X-ray diffraction. *Curr. Opin. Struct. Biol.* 10:569-573.

Hanson, B.L., Langan, P., Katz, A., Li, X., Harp, J.M., Glusker, J.P., Schoenborn, B.P., and Bunick, G.J. 2004. A preliminary time-of-flight neutron diffraction study of *S. rubiginosus* D-xylose isomerase. *Acta Cryst. D.* 60:241-249.

Harp, J.M., Timm, D.E., and Bunick, G.J. 1998. Macromolecular crystal annealing: overcoming increased mosaicity associated with cryocrystallography. *Acta Cryst. D.* 54:622-628.

Helliwell, J.R. 1992. Macromolecular crystallography with synchrotron radiation. Cambridge University Press, Cambridge, UK.

Hendrickson, W.A. 2000. Synchrotron crystallography. *Trends Biochem. Sci.* 25:637-643.

Hendrickson, W.A., Horton, J.R., and LeMaster, D.M. 1990. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): A vehicle for direct

determination of three-dimensional structure. *EMBO J.* 9:1665-1672.

Holbrook, S.R. and Kim, S.-H. 1985. Crystallization and heavy-atom derivatives of polynucleotides. *Methods Enzymol.* 114:167-176.

Jancarik, J. and Kim, S.-H. 1991. Sparse matrix sampling: A screening method for crystallization of proteins. *J. Appl. Cryst.* 24:409-411.

Johnson, C.H., Egli, M., and Stewart, P.L. 2008. Structural insights into a circadian oscillator. *Science* 322:697-701.

Kato, T., Goodman, R.P., Erben, C.M., Turberfield, A.J., and Namba, K. 2009. High-resolution structural analysis of a DNA nanostructure by cryoEM. *Nano Lett.* 9:2747-2750.

Keel, A.Y., Rambo, R.P., Batey, R.T., and Kieft, J.S. 2007. A general strategy to solve the phase problem in RNA crystallography. *Structure* 15:761-772.

Keeler, J. 2005. Understanding NMR. 1st ed. John Wiley & Sons, Hoboken, N.J.

Kemp, M. 2003. The Mona Lisa of modern science. *Nature* 421:416-420.

Kim, S.-H., Shin, W.-C., and Warrant, R.W. 1985. Heavy metal ion-nucleic acid interaction. *Methods Enzymol.* 114:156-167.

Lilley, D.M.J. and Wilson, T.J. 2000. Fluorescence resonance energy transfer as a structural tool for nucleic acids. *Curr. Opin. Chem. Biol.* 4:507-517.

McPherson, A. 1998. Crystallization of Biological Macromolecules. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

McRee, D.E. 1999. Practical Protein Crystallography, 2nd ed. Academic Press Inc., San Diego, Calif.

Milligan, J.F. and Uhlenbeck, O.C. 1989. Synthesis of small RNAs using T7 RNA-polymerase. *Methods Enzymol.* 180:51-62.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. 1999. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst. D* 53:240-255.

Pallan, P.S. and Egli, M. 2007a. Selenium modification of nucleic acids. Preparation of phosphoroselenoate derivatives for crystallographic phasing of nucleic acid structures. *Nat. Protocols* 2:640-646.

Pallan, P.S. and Egli, M. 2007b. Selenium modification of nucleic acids. Preparation of oligonucleotides with incorporated 2′-SeMe-uridine for crystallographic phasing of nucleic acid structures. *Nat. Protocols* 2:647-651.

Petsko, G.A. 1985. Preparation of isomorphous heavy-atom derivatives. *Methods Enzymol.* 114:147-156.

Pitsch, S., Weiss, P.A., Jenny, L., Stutz, A., and Wu, X. 2001. Reliable chemical synthesis of oligoribonucleotides (RNA) with 2′-O-[(triisopropylsilyl)oxy]methyl (2′-O-tom) protected phosphoramidites. *Helv. Chim. Acta* 84:3773-3795.

Putnam, C.D., Hammel, M., Hura, G.L., and Tainer, J.A. 2007. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* 40:191-285.

Radaev, S. and Sun, P.D. 2002. Crystallization of protein-protein complexes. *J. Appl. Cryst.* 35:674-676.

Rhodes, G. 2006. Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models, 3rd ed. Academic Press Inc., San Diego.

Ringe, D. and Petsko, G.A. 1996. A consumer's guide to protein crystallography. *In* Protein Engineering and Design (P.R. Carey, ed.) pp. 205-229. Academic Press Inc., San Diego.

Rould, M.A. 1997. Screening for heavy-atom derivatives and obtaining accurate isomorphous differences. *Methods Enzymol.* 276:461-472.

Saibil, H.R. 2000. Macromolecular structure determination by cryo-electron microscopy. *Acta Cryst. D* 56:1215-1222.

Scaringe, S.A., Francklyn, C., and Usman, N. 1990. Chemical synthesis of biologically active oligoribonucleotides using β-cyanoethyl protected ribonucleoside phosphoramidites. *Nucleic Acids Res.* 18:5433-5441.

Scaringe, S.A., Wincott, F.E., and Caruthers, M.H. 1998. Novel RNA synthesis method using 5′-silyl-2′-orthoester protecting groups. *J. Am. Chem. Soc.* 120:11820-11821.

Schuler, B. and Eaton, W.A. 2008. Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* 18:16-26.

Scott, W.G., Finch, J.T., Grenfell, R., Fogg, J., Smith, T., Gait, M.J., and Klug, A. 1995. Rapid crystallization of chemically synthesized hammerhead RNAs using a double screening procedure. *J. Mol. Biol.* 250:327-332.

Smyth, D.R., Mrozkiewicz, M.K., McGrath, W.J., Listwan, P., and Kobe, B. 2003. Crystal structures of fusion proteins with large-affinity tags. *Protein Sci.* 12:1313-1322.

Structural Genomics Consortia. 2008. Protein production and purification. *Nat. Methods* 5:135-146.

Tereshko, V., Wilds, C.J., Minasov, G., Prakash, T.P., Maier, M.A., Howard, A., Wawrzak, Z., Manoharan, M. and Egli, M. 2001. Detection of alkali metal ions in DNA crystals using state-of-the-art X-ray diffraction experiments. *Nucleic Acids Res.* 29:1208-1215.

Terwilliger, T.C. and Berendzen, J. 1999. Automated MAD and MIR structure solution. *Acta Cryst. D.* 55:849-861.

Terwilliger, T.C., Stuart, D., and Yokoyama, S. 2009. Lessons from structural genomics. *Annu. Rev. Biophys.* 38:371-383.

Tsuruta, H. and Irving, T.C. 2008. Experimental approaches for solution X-ray scattering and fiber diffraction. *Curr. Opin. Struct. Biol.* 18:601-608.

Weeks, C.M., Adams, P.D., Berendzen, J., Brunger, A.T., Dodson, E.J., Grosse-Kunstleve, R.W., Schneider, T.R., Sheldrick, G.M., Terwilliger, T.C., Turkenburg, M.G., and Uson, I. 2003. Automatic solution of heavy-atom substructures. *Methods Enzymol.* 374:37-83.

Wilds, C.J., Pattanayek, R., Pan, C., Wawrzak, Z., and Egli, M. 2002. Selenium-assisted nucleic acid crystallography: Use of DNA phosphoroselenoates for MAD phasing. *J. Am. Chem. Soc.* 124:14910-14916.

Wincott, F., DiRenzo, A., Shaffer, C., Grimm, S., Tracz, D., Workman, C., Sweedler, D., Gonzalez, C., Scaringe, S., and Usman, N. 1995. Synthesis, deprotection, analysis and purification of RNA and ribozymes. *Nucleic Acids Res.* 23:2677-2684.

Woolfson, M.M. 1997. An Introduction to X-ray Crystallography, 2nd ed. Cambridge University Press, Cambridge.

Wüthrich, K. 1986. NMR of Proteins and Nucleic Acids. 1st ed. John Wiley & Sons, New York.

Wyatt, J.R., Chastain, M., and Puglisi, J.D. 1991. Synthesis and purification of large amounts of RNA oligonucleotides. *BioTechniques* 11:764-769.

Zauberman, N., Mutsafi, Y., Ben Halevy, D., Shimoni, E., Klein, E., Xiao, C., Sun, S., and Minsky, A. 2008. Distinct DNA exit and packaging portals in the virus *Acanthamoeba polyphaga mimivirus*. *PLoS Biol* 6:e114.

Zhou, Z.H. 2008. Towards atomic resolution structural determination by single-particle cryo-electron microscopy. *Curr. Opin. Struct. Biol.* 18:218-228.

Biophysical Analysis of Nucleic Acids

7.13.35