PART I

INTRODUCTION AND OVERVIEW

1

Introduction

Biological macromolecules are the machinery of life; visualizing their three-dimensional structure helps us to fully understand their function. Visual observation with a light microscope is not possible as their sizes are well below the wavelength of visible light. While X-rays and neutrons allow visualization, they cannot be focused, so diffraction techniques have to be used. An understanding of three-dimensional macromolecular structure gives us a deeper understanding of basic biological concepts and processes, reveals the causes of diseases, assists rational pharmaceutical design and can lead to the design of macromolecules with novel properties. Visualizing these macromolecules is a complex ballet involving diverse but interrelated fields of endeavour. In this book, we aim to describe in some detail these complementary techniques, which include crystallization, diffraction and analysis of the data to obtain atomic structure from crystals of macromolecules. We cover areas where problems can occur and potential solutions to those problems. Finally, we touch on some of the developments in the not so distant future.

When we use the term crystal in the context of a biological macromolecule, we are describing an ordered array of macromolecules in an environment that keeps them stable. Biological macromolecules are predominantly made up of low atomic weight atoms, including hydrogen, carbon, nitrogen, oxygen, sulphur and phosphorous. Unlike inorganic crystals, a significant proportion (30–70%) of a macromolecular crystal is water (Matthews, 1968). This makes biological crystallography challenging; the process of crystallization is very complex and the crystals themselves diffract very weakly in comparison to inorganic crystals due to the low atomic weight content and disordered water making up a large proportion of the crystal volume. The biochemical and biophysical process of crystallization is still a largely empirical process. High-throughput technologies have been employed to speed up the crystallization process, allowing for more experiments to be set up using less sample, but their greater power may be realized if we achieve the array of experiments needed to better understand this complex process to develop crystallization.

Once we get the crystals, technologies and methods have advanced significantly and transformed our capabilities for structure analysis. In the X-ray field, highlights of the last three decades have included synchrotron radiation, detectors and cryo-crystallography. In a synergistic development neutron Laue methods, along with very large area neutron-sensitive image plates, and new spallation source

developments, are enhancing the potential for complete, i.e. with-H atoms (as deuteriums), structures. Complementary to these are developments in molecular biology techniques that go beyond the simple cloning and expression (production) of the target sample. Rational mutagenesis is providing functional information and being used to improve crystal quality, while fully perdeuterated protein production is enabling neutron studies.

1.1 Crystal Growth

The history of crystallization has been described by McPherson (1991; 1999a). The first published observation of crystallization, haemoglobin, was noted by Hünefeld (1840) when the blood of an earthworm was pressed between two microscope slides. This was followed by a slow growth in a number of other crystallized macromolecules until the 1930s when the pace quickened. Crystallization was initially used as a method of isolation with Northrop, Sumner and Stanley sharing the 1946 Nobel Prize for chemistry for the isolation and crystallization of proteins and viruses. The first biomolecular crystal-structure determined was that of vitamin B-12 in 1957 by Dorothy Crowfoot Hodgkin (1957) who subsequently won the Nobel prize for chemistry as a result of the work. Initially there were no set rules or recipes as to where to start to crystallize a macromolecule. Macromolecules were solubilized and then treated with arrays of precipitants in order to find favourable conditions for crystallization.

Crystallization is still largely empirical with many experiments sampling a large range of possible crystallization conditions. This idea of a designed sampling of many conditions, or screening, was introduced in 1979 (Carter and Carter). The numerous experiments required repetitive pipetting, a laborious, time-consuming and tedious task but the principal difficulty was attributing a quantitative score to the results to enable a meaningful mathematical analysis. In the late 1980s and early 1990s, the development of automatic means of dispensing crystallization trials (Chayen *et al*., 1992; Chayen *et al*., 1994; Chayen *et al*., 1990; Cox and Wever, 1987; Oldfield *et al*., 1991; Rubin *et al*., 1991; Sadaoui *et al*., 1994; Soriano and Fontecillacamps, 1993; Ward *et al*., 1988) showed the promise of designed screening. However, it was not until the commercialization of a crystallization screen developed by Jancarik and Kim (1991) that systematic screening became a standard laboratory technique.

In the 1980s an effort was initiated to turn crystallization from an art into a science; the first of a continuing series of international conferences on the subject occurred (McPherson and Giege, 2007) and the term ‘crystallogenesis’ was coined (Giege *et al*., 1986). The purpose of crystallogenesis was to understand the fundamental principles of the crystallization process, to quantitatively measure the biophysical and chemical parameters that are involved in crystal growth, and to use that knowledge to design experiments for obtaining better diffracting crystals. Systematic studies were performed mostly on lysozyme and other model proteins that crystallized with ease.

The aim was to extrapolate that information to target proteins that were proving difficult to crystallize. Progress was slow due to lack of suitable equipment for monitoring the crystallization process and to the complexity of the problem. Even the crystallization process for model proteins was not straightforward given the number of variables that were involved. Considerable development effort has been ongoing in the last 20 years on crystal-growth methods as well as more sophisticated monitoring and characterization, such as use of light scattering, interferometry and other techniques.

Much of the success of the last decade has come by way of automating and miniaturizing crystallization trials (Kuhn *et al*., 2002; Luft *et al*., 2003; Walter *et al*., 2003) and by way of the development of diagnostic apparatus to study the crystallogenesis aspect of crystallization (Dierks *et al*., 2008; Yeh and Beale, 2007). The ability to dispense trials consisting of nanolitre volumes in a high-throughput mode has cut the time of setting up experiments from weeks to minutes, and reduced sample requirements by an order of magnitude, a scenario that was unimaginable even in the recent past. While high-throughput approaches and miniaturization do not elicit a better understanding of crystallization, the analysis of these systematic and highly reproducible trials will improve our comprehension of the crystallization process, and application of these methods to specifically understand this process will enable us to answer and then ask many more questions.

Having a well-diffracting single crystal is a first step, but is not necessarily sufficient to solve the macromolecular structure. Detectors are able to measure the position and intensity (amplitude) of scattered reflections, but do not record phase information. In order to use a Fourier transform to go from a diffraction pattern, to an interpretable electron-density map that can be used to model the structure, we need phase information. We can use several approaches to provide an initial set of phases. If the protein has fewer than 1,000 non-hydrogen atoms, and the resolution of the diffraction data is near atomic resolution, we can use *ab initio* phasing (direct methods) to solve the structure (Hauptman, 1997; Uson and Sheldrick, 1999), or if there is significant sequence homology to other known structures, then molecular replacement may be used to solve the structure without modifying the sample.

If not, sample modification will likely be required. This can be accomplished by soaking, or co-crystallizing the sample with heavy atoms to provide phase information for a sub-structure (Islam *et al*., 1998). Another approach, exploits differences in diffraction intensities caused by anomalous scattering, absorbance of X-rays by elements at wavelengths at, or near a particular element's absorption edge. The most common application of this method uses molecular biology to replace naturally occurring sulphur atoms in the protein's methionine residues with selenium atoms, creating a selenomethionine variant (Hendrickson *et al*., 1990). The derivative is crystallized, and diffraction data collected at several wavelengths near the selenium absorption edge; this is an example of MAD (multiple anomalous dispersion) phasing. SAD (single anomalous dispersion) phasing is similar, but uses only one wavelength, making this a better-suited method for radiation-sensitive crystals (Gonzalez, 2007).

A special aspect of this history is the growth of macromolecular crystals for neutron crystallography. Neutrons are weakly scattered and neutron fluxes are low compared to X-ray sources. Before the advent of area detectors at neutron sources and the use of polychromatic beams, this challenge was particularly acute, whereby crystal volumes on the order of $\sim$30 mm^3 were required and that measurable diffraction often could not be obtained within a reasonable timeframe. To achieve these large crystals, a crystal of 'typical size' would often be transferred into fresh growth solution (Lennart Sjolin, personal communication). Specialized image plates at reactor neutron sources have made a big impact on the efficiency of data collection compared to the point-by-point or linear detector diffractometers. The use of Laue techniques, e.g. a wavelength bandpass of 2.5 to 3.5 Å has increased the efficiency of measuring many reflections at once and also helped by allowing such long wavelengths to enhance the scattering efficiency for each spot via a λ^2 effect.

Production of fully deuterated protein for neutron protein crystallography, initiated by Gamble *et al.*, (1994) has led to a further reduction in crystal sample volumes due to two effects. First, for coherent neutron scattering, that produces Bragg reflections, deuterium scatters neutrons as strongly as carbons; there is a doubling of the strongly scattering atoms since there are so many hydrogens in a protein. Secondly, the incoherent scattering cross-section of hydrogen is anomalously large, causing high background scattering from hydrogen. This is reduced significantly when hydrogen is replaced by deuterium. The combination of these two factors provides for nearly an order of magnitude improvement in the neutron protein crystallography signal-to-noise ratio (Myles, 2006).

Finally, source developments help neutron diffraction. The new spallation neutron sources offer peak flux improvements and time-of-flight measurements that allow the background being accumulated over a diffracted spot to be kept to the minimum. All these developments have radically reduced the required crystal volume but it is still much larger than that required for X-rays.

While neutrons have required larger crystals, developments in intense and tunable X-ray beams have steadily allowed smaller and smaller crystal volumes to produce useable data with now down to micrometre sized beams and sample volumes. Plans for future upgrades and beamlines (e.g. the ESRF Upgrade) offer sub-micrometre and indeed 'nano-' sized X-ray beams for possible further reductions in sample volume.

1.2 Diffraction Techniques

1.2.1 X-rays

X-rays were discovered in 1895 by Röntgen and in 1912 Friedrich, Knipping and von Laue discovered that they were diffracted by crystals of small molecules. There is a large literature available on the development of X-ray single-crystal techniques and that literature, critical to understanding how crystal quality is important, is summarized

here. For the oscillation method in single-crystal X-ray diffraction the integrated intensity of a reflection (*h*,*k*,*l*) for atoms at rest, $E(h,k,l)$, can be calculated from:

$$E(h,k,l) = \frac{\lambda^3}{\omega.V^2}\left(\frac{e^2}{mc^2}\right)^2 V_{cr} I_o LPA|F(h,k,l)|^2 \tag{1.1}$$

where λ is the wavelength, ω the angular rotation velocity, V the unit-cell volume, V_{cr} the illuminated volume of the crystal, I_o the incident intensity, L the Lorentz factor, P the polarization, A the absorption and F the structure factor (Woolfson, 1997). The structure factor is given by

$$F(h,k,l) = V \int_{\text{cell}} \rho(x,y,z) e^{[2\pi i(hx+ky+lz)]} dx\, dy\, dz \tag{1.2}$$

where x, y and z are the fractional coordinates in the unit cell and $\rho(x,y,z)$ is the electron density at those coordinates. From this we can see how diffraction intensity decreases as the square of the unit-cell volume; large macromolecules or complex systems will diffract poorly compared to smaller counterparts. The dependence on wavelength can be misleading as absorption and practical considerations may limit signal improvements that can be gained by changing wavelength.

All contemporary laboratory sources operate on a principle of electrons striking an anode with radiation given off from the anode at a characteristic wavelength associated with the anode material, for example Cu=1.54 Å, Co=1.80 Å, *etc*.

Synchrotrons have been one of the most dramatic developments in X-ray crystallography. Electrons (or positrons) are accelerated to high velocity and stored in a circular orbit. As they lose energy, the stored particles give off electromagnetic energy, some of it in the X-ray spectrum. The first synchrotrons started life as experimental machines for particle physics and were used by crystallographers in a parasitic mode; the loss of energy was considered detrimental. These synchrotrons were superseded by a second generation of machines designed and dedicated solely to provide synchrotron radiation for experimental use. The current third-generation synchrotron source are machines designed to incorporate insertion devices.

1.2.2 Neutrons

Neutrons were discovered later than X-rays, by Chadwick in 1932, and the first diffraction experiments took place in 1945 by Wollan. Neutrons, produced in a nuclear reactor, are slowed down such that their wavelength is similar to that of X-rays used for crystallography. X-rays interact with the electron cloud, while neutrons interact with the nucleus of the atom. In the case of X-rays, the X-rays scattered are, to a first approximation, proportional to the atomic number. For neutrons the contribution to the scattered intensity is different for each isotope, illustrated in Figure 1.1. In macromolecular crystallography we can make use of this to differentiate hydrogen from deuterium or to examine the protonation state of an amino acid residue (see e.g. Shu *et al*., (2000)). Neutrons are scattered by the nucleus causing relatively little

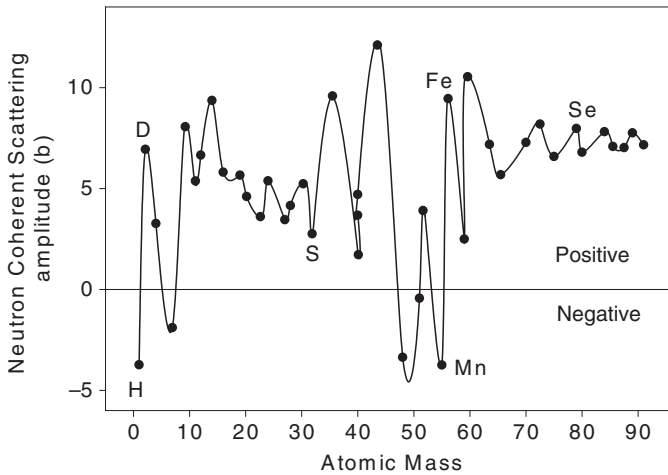

Fig. 1.1. Plot of neutron coherent scattering amplitude as a function of atomic number.

decrease in intensity with scattering angle, unlike X-rays that are scattered by the much larger electron charge cloud. Obviously ADPs (atomic displacement factors) (deleteriously) affect the fall off with scattering angle in both the neutron and X-ray cases. Neutrons do not cause radiation damage or suffer absorption, allowing ready use of larger crystals.

The original neutron source, the nuclear reactor, has also been joined by spallation sources. When a high-energy proton bombards a heavy atom nucleus some neutrons are 'spalled' or knocked out. For every proton hitting a nucleus some 20 to 30 neutrons are expelled.

Neutron sources have relatively low flux compared to synchrotron X-ray beamlines, in fact typical sources have a flux comparable to that of an X-ray tube. In addition, the scattering of neutrons is weak. To enable optimum experimental results, the neutron flux and diffraction signal have to be maximized and the noise minimized. Table 1.1 (Snell *et al*., 2006) summarizes several complementary pathways to achieve this goal with a qualitative estimate of the relative cost, time and chance of success.

Certain approaches, i.e. increasing the neutron flux or improving the detection of the signal, are beyond the resources of an individual user and are being addressed at the facility level, e.g. in the construction of dedicated instruments on new neutron sources such as the Spallation Neutron Source (SNS), USA, and the Japan Proton Accelerator Complex (J-PARC). Other techniques, for example improving the signal-to-noise ratio by perdeuteration (replacing hydrogen with deuterium atoms) may be routinely attempted in the laboratory (Hazemann *et al*., 2005; Shu *et al*., 2000) and can dramatically decrease the volume of the crystal needed for successful data collection (Hazemann *et al*., 2005; Shu *et al*., 2000).

Table 1.1. Summary of different approaches to improve reflection intensity from neutron diffraction divided into increasing neutron flux and brilliance, improving the instrument, reducing the noise or improving the sample (beyond optimization in the crystallization step). All the approaches are complementary and cumulative. They are assessed by the relative cost, time and success of implementation (Snell *et al*., 2006).

Solution	Relative cost	Time	Success
More neutrons			
Increase source intensity	Very expensive	Long term	Certain
Reduce distance to source	Expensive	Medium term	Certain
Increase exposure time	Inexpensive but reduces throughput	Short term	Certain
Improved detection/optics			
New/improved detector technology	Expensive	Medium term	Good
Focusing optics	Moderately expensive	Medium term	Good
Improve signal-to-noise			
Deuteration	Relatively inexpensive	Short term	Very good
Diffracting volume			
Grow larger crystals	Inexpensive	Short-term	Good

1.3 Crystal Volume and Quality

A high-quality crystal can be defined as one that provides structural information that is of sufficient detail (resolution) to answer the structural question being posed. A crystal of sufficient volume and internal order is required to measure complete, observable diffraction data at high resolution.

Examining this definition, the first requirement, crystal sample volume, is an easily quantified metric. For X-ray single-crystal studies structural information can currently be extracted from crystals as small as 20 μm in diameter (Hedman *et al*., 1985; Pechkova and Nicolini, 2004a; Pechkova and Nicolini, 2004b). In the extreme case structural information can be extracted from microcrystallites or powder. For neutron diffraction the volume requirement is now falling below 1 mm^3 with perdeuteration techniques of the protein itself (Hazemann *et al*., 2005) and now new neutron spallation sources coming on-line. Many crystal-growth methods and optimization techniques exist to address improving crystal volume.

While crystal volume can be determined visually, the requirement for order within the crystal can only be determined by diffraction methods. What determines order within a crystal? Ideally, well-ordered crystal macromolecules would be conformationally identical, down to the position of every sidechain on each amino acid residue. If a single sidechain on a sequentially identical amino acid residue holds two stationary positions we have what is termed a static disorder. An example of this situation is when two structural states are seen in the structural model derived from the electron-density map, rather than one 'unique' state. The split-occupancy values in favourable cases are refined against the Bragg reflection intensities. If one can see

such 'split-occupancy' details in a crystal structure, e.g. of an amino acid sidechain in two states, then it can be viewed as a remarkable state of order each with their fractional occupancy (constrained to add to 100%). Furthermore, if one cannot see evidence of the structure at all in regions where it is clearly present then the term disordered makes the most descriptive sense. There is in such a way of describing it a logical progression here of fully ordered, partially ordered and finally disordered. However, the tradition has grown up of referring to cases of less than 100% order as 'static disorder', i.e. a crystal is termed imperfect even if, say, 40% of it is ordered one way and 60% in another. Another type of disorder is occupational, i.e. two or more elements occupying the same site. Further, aperiodic structures are those that are fully ordered in a higher dimension (e.g., a modulated structure or quasi-crystal).

The lattice scale internal crystal order can be divided into two distinct properties, short-range order and long-range order (Snell *et al.*, 2003). Disruptions to the internal crystal order lead to a 'bad' crystal with a number of associated symptoms;

- The crystal may only diffract to low resolution.
- The diffraction spots themselves have texture features and/or are basically split.
- There may be few or no spots even if the crystal looks optically well defined.
- There is a significant anisotropic local order variation within the crystal revealed by detailed examination of the diffraction data.
- There is a superlattice of closely spaced diffraction spots.
- Crystal twinning can be bad but is now increasingly amenable to detwinning diffraction data analysis computational procedures. Alternatively more crystal growth 'remediation methods' are now known.

To understand these symptoms we will look in detail at both short- and long-range order then the tools and techniques that can be brought to bear for a systematic exploration of order in terms of crystal quality.

1.3.1 Short-range order (intermolecular)

Good short-range order in a crystal is a primary needs-driver for the crystallographer where high-resolution structural information is the required feature. An atom will contribute coherently to the intensity of a reflection only if its disorder relative to symmetry-related atoms is small. Figure 1.2 dissects the various short-range intermolecular ***disorders*** that can occur on the molecular scale within a crystal. Atoms can be displaced by thermal vibrations; they can have multiple or partial occupancies; their position may be uncertain, especially in the case of waters and there may be variations in the main chain or sidechains and in the intermolecular packing. There may also be vacancies or interstitial substitutions, which clearly also interrupt crystal periodicity.

Atomic displacement or temperature effects

Temperature effects reduce the integrated reflection intensity as a function of resolution. This temperature-dependent atomic displacement parameter, *ADP*,

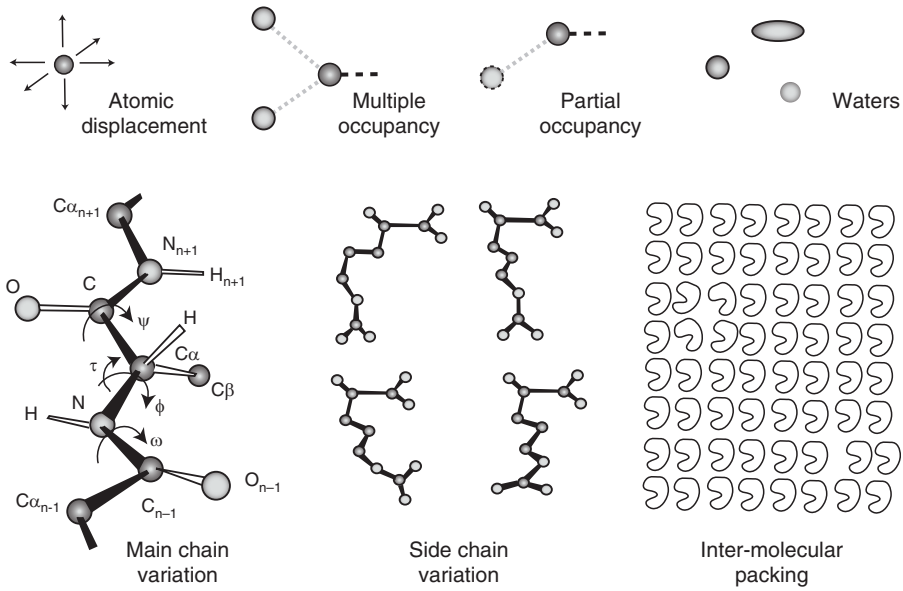

Fig. 1.2. Contributors to the reduction of short-range order within a macromolecular crystal.

often referred to as the B_{factor}, is related to the mean square atomic displacement by:

$$B_{\text{factor}} = ADP_i = 8\pi^2 < u_i^2 > \tag{1.3}$$

where $< u_i^2 >$ is the mean square atomic displacement of the atomic vibration for atom i. The overall temperature factor for the crystal, averaged over the whole structure, can be calculated from a Wilson plot (1942) where E_{obs} the observed intensity of the reflection, is plotted against $(\sin\theta/\lambda)^2$. The $<B_{\text{factor}}>$ is extracted from the slope of this plot as:

$$E_{\text{obs}}(h,k,l) = E_{(h,k,l)}\mathrm{e}^{(-2<B_{\text{factor}}>\sin^2\theta/\lambda^2)} \tag{1.4}$$

The Wilson plot also provides a scale factor, where it crosses the vertical axis, allowing intensities to be put on an absolute scale.

The atomic-displacement parameter is a measure of how much dynamic disorder of an atom is within the crystal lattice on the molecular level. X-rays interact with the electron cloud, which depends on the atomic position. Each atom vibrates around an equilibrium position so the nuclear position and electron charge cloud both move. The ADP is an indication of both this motion around the equilibrium of the *i*th atom position **and** the deviation in position of that atom from molecule to molecule making up the crystal lattice. As can be seen from eqn (1.4), reducing the $<B_{\text{factor}}>$ has a

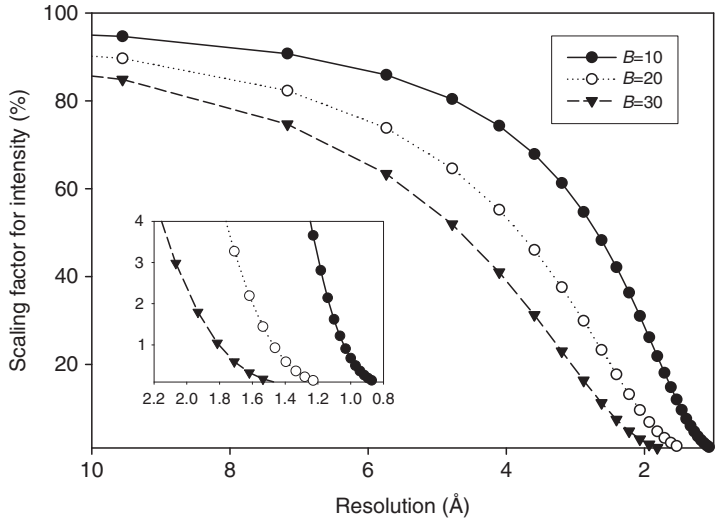

Fig. 1.3. Graph of scaling factor for intensity with resolution (eqn (1.5)) showing that a decrease in the $<B_{\text{factor}}>$ causes a significant increase in the resolution. This is a simplified example neglecting crystal path length absorption differences and Bragg reflection signal-to-noise smearing at higher resolutions.

dramatic effect on increasing the signal-to-noise of the data. A small ADP_i is indicative of a combination of small dynamic motion of the ith atom and good intermolecular order within the crystal. In Figure 1.3 the scaling factor $e^{(-2<B_{\text{factor}}>\sin^2\theta/\lambda^2)}$ is plotted against resolution for varying values of B. Increasing $<B_{\text{factor}}>$ results in a more rapid fall off of intensity as a function of resolution.

For a macromolecule, in practice, there is considerable shape to this plot and a curve rather than a straight line is seen at resolutions around 3 to 4 Å. This is due to the nature of macromolecular secondary structures, alpha helix and/or beta sheet, and their regularity, namely periodicity of the helix on the one hand and inter-strand separation on the other, causing the molecular transform to peak at these distances in reciprocal space. At higher resolution a straight line results and an accurate assessment of the $<B$ value$>$ becomes possible. The weakness of this parameter as the sole judge of optimal crystal-growth conditions is that it is only an average indicator summed over the innate flexibilities of the protein and the periodicity in a given crystal sample.

The relative Wilson plot has been used to evaluate the variation of crystal to crystal quality for crystals of the same protein either grown by different methods or where there is known to be variation within the same batch. This calculation has the advantage that the helix and sheet molecular effects then cancel out.

If we compare the protein molecule in a protein crystal situation in practice, described above, with a small-molecule crystal, then the distinctions regarding the $<B_{\text{factor}}>$ are as follows. A small-molecule crystal has, firstly, a generally much

lower set of individual atomic ADP_i values and a generally better adherence to periodicity. Such molecules much less frequently have extensive mobile parts such as the loops of a protein and contain very little solvent of crystallization, unlike protein crystals. It used to be the case that whilst it was routine to cool a small-molecule crystal this was not done for a protein crystal. The latter has only become common practice since around the late 1980s, championed in a major way by the ribosome protein structure work of Ada Yonath with the cryogenic developments particularly of Hope *et al*. (1989), although conceived of in the 1950s (e.g. see King (1958)). Typically, a small-molecule crystal at room temperature has an overall B_{factor} of 6 and at cryogenic temperature ($\sim$100 K) of 3 whereas a protein crystal at room temperature has an overall $<B_{\text{factor}}>$ in the range of 25 to 50 and at 100 K cryogenic temperature in the range of 8 to 25.

Variation in atomic positions or occupancy

The structure determined from X-ray diffraction data is then a time- and spatial-averaged structure of the complete number of individual macromolecules making up the crystal. This can be a large number. For example, if we consider a cubic crystal with dimensions of 100 μm made up of macromolecules with a unit cell of 100 Å on edge then there will be 10^{12} of those molecules in the crystal. An atomic displacement parameter takes into account the variation about a single atomic position. A number of residues may have alternate positions or multiple conformations. In some cases, e.g. in the interior of the macromolecule or at crystal lattice contacts, these conformations are clearly defined. In other cases, e.g. long flexible residues or complete loops on the outside of the protein, the atomic positions are not clearly defined. Multiple sidechain conformations can be determined in high-resolution analyses from about 1.5 Å onwards and the fraction of atoms in this category seems to be often more when the crystal structure has been determined at cryogenic rather than room temperature.

There are particularly long and therefore flexible amino acid sidechains such as lysine. The atoms in these sidechains have high atomic values of their ADP_i. Site-directed mutagenesis to replace such sidechains has become an approach to improve overall crystal quality and has enjoyed success (Derewenda, 2004a; Derewenda, 2004b) and is discussed in Chapter 6.

The bound-water structure on the surface of a protein can also reach around 10% of the X-ray crystal-structure determined atom positions in the cryo case and around 3 to 5% in the room-temperature case. The average $ADP_i s$ for this population of atoms is usually considerably more than the well-ordered protein atoms, e.g. twice as high but is a strong function of whether the bound water molecule is found in the first water layer or the second (a few waters can also be found in a third layer at cryo-temperature).

In the case of neutron protein crystallography where fully deuterated protein has been isolated and the bound water is D_2O the proportion of atoms that diffract as well as carbon atoms has risen considerably. This has a strong effect on increasing the

diffraction signal. Moreover, the large incoherent background contribution of hydrogen in the interaction with a neutron beam is considerably reduced by replacement with deuterium in the protein and in the D_2O. Since deuterium scatters as well as carbon with neutrons then determination of the protonation states of ionizable amino acid sidechains is viable even at 2Å (or up to 2.3 Å or so) resolution. The use of X-rays, on the other hand, is vulnerable to the effects of specific atom ADP_i values, thus losing the relatively weak hydrogen scattering signal, even at atomic resolution, which is why neutrons can provide the critical complementary information even with a more modest diffraction resolution.

1.3.2 Long-range order (domain structure)

Long-range order is a whole-crystal length-scale effect. Good long-range order results in high signal-to-noise in the reflection profiles. The intensity of the reflection is directly proportional to illuminated volume, see eqn (1.1), but the peak width is also somewhat narrowed due to the larger volume, i.e. for an infinite perfect crystal the peak would be a Dirac delta function.

Figure 1.4 illustrates perturbations to long-range order and how long-range disorder contributes to broadening the resulting diffraction reflection profile. The figure makes use of the mosaic model of crystals proposed by Darwin (1922) and approximates the crystal to an array of perfectly ordered volumes (domains) slightly misaligned with respect to each other (the boundaries between these domains are ignored and no model for them is proposed). In addition to having small random misalignments, the domains can be of varying volume and to a much smaller degree the unit cells in the crystal can vary in dimension. Each of these phenomena has a distinct effect on the crystal (Boggon *et al.*, 2000; Nave, 1998). In the case shown in Figure 1.4(a) all the domains are well aligned so their contributions to the reflection overlap. Misalignment of the domains broadens the reflection profile reducing the signal-to-noise. If the volume of the domains becomes small, the reflections will become broadened from Fourier truncation effects (the transition from diffraction grating to a few slits is the analogous situation in optical diffraction and interference theory). The effect is known as domain-size broadening. A lattice parameter variation, Figure 1.4(c) causes a reflection to have a range of slightly different Bragg angles, also resulting in a smearing out of the reflection.

Long-range disorder in the crystal gives rise to localized effects in reciprocal space (Boggon *et al.*, 2000; Nave, 1998). Improved long-range order in a crystal reduces the mosaicity, which results in an increase in the signal-to-noise of the reflections. This is illustrated in Figure 1.5, where Gaussian profiles of the same integrated intensity but different full width at half-maximum values are shown. The increase in signal is directly proportional to the decrease in mosaicity.

Mosaicity is a global term covering contributions from all components shown in Figure 1.4. For practical purposes, it is defined as a function of the angular extent of the diffraction profile. In Chapter 11, the measurement of mosaicity and the

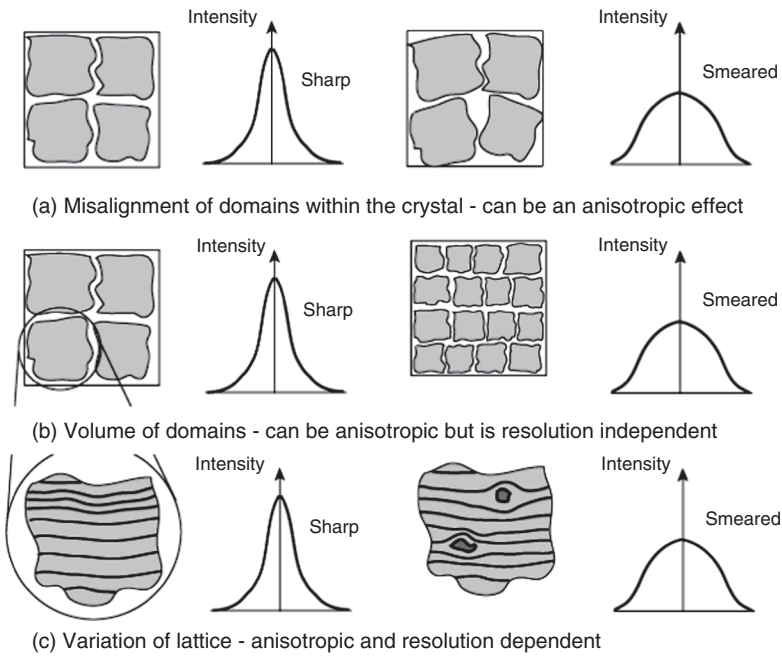

Fig. 1.4. Long-range disorder and the resulting effect on the diffraction profile.

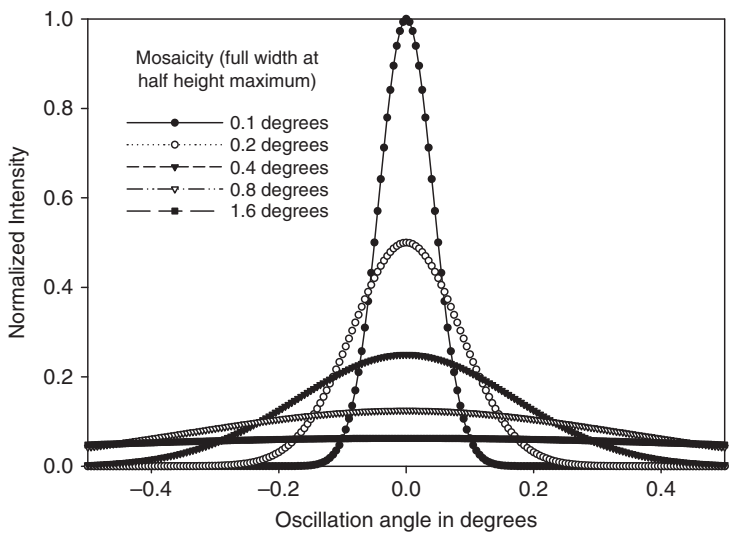

Fig. 1.5. Graph showing a Gaussian reflection profile for identical integrated intensities but differing full width at half-height maxima to illustrate how an increase in mosaicity causes a corresponding decrease in the reflection peak intensity signal.

deconvolution of it from experimental parameters and into the component parts is discussed.

1.3.3 The combination of short- and long-range order

Disorder within a crystal is a combination of short- and long-range disorder. In Figure 1.6, the scaling factor on intensity is plotted for a crystal having an overall $<B_{\text{factor}}>$ of 10 but mosaicity varying from 0.1 to 0.8 degrees.

Figure 1.6 shows a plot of the scaling factor for intensity as a function of resolution. The peak reflection height as a function of mosaicity has been used to calculate initial intensity for the plot. This provides a worst-case scenario since integration techniques are profile rather than peak based. However, it shows that in principle mosaicity has little effect on resolution. In practice, in crystal-structure analysis, the percentage of data with $I>2\sigma(I)$ is very important and is improved as Figure 1.6 shows nicely. How to optimize data collection to take advantage of reduced mosaicity is discussed later, but basically in practice account has to be taken of the background noise in a diffraction pattern. It is in this area that sharpening of mosaicity can have a positive impact on the quality of diffraction data measurement. Alternatively, of course the electronic gating of a detector to exclude noise is another strategy for achieving a signal-to-noise measurement optimization. In summary, reducing mosaicity in this case does not significantly extend the theoretical resolution but does dramatically increase the peak intensity of the signal. This can

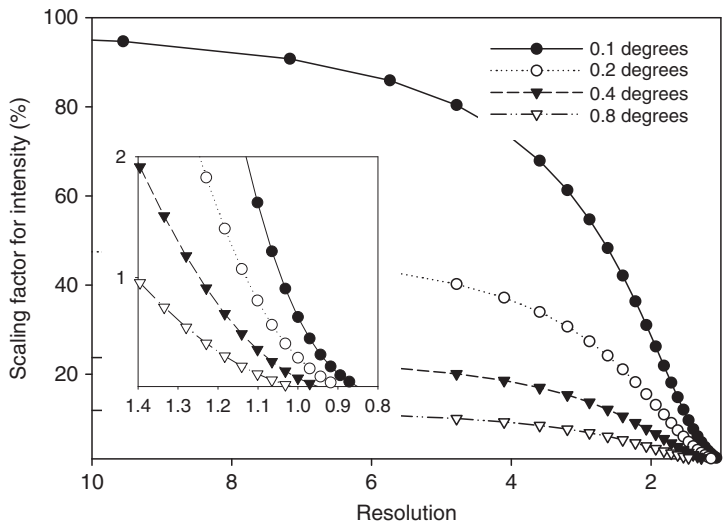

Fig. 1.6. Theoretical plot of scaling factor for intensity signal versus resolution for a crystal with mosaicity of 0.1 to 0.8 degrees. The plot is normalized against the 0.1 degree mosaicity crystal.

impact the resolution likely to be achieved in practice as measured by, e.g. the percentage of data with $I > 2\sigma(I)$.

1.4 Chapter Summary

To obtain structural data from crystallography crystals that diffract are needed. To obtain good data, those crystals have to be as perfect as possible. In the chapters ahead we explore crystallization going into theory and practical aspects. We discuss screening and optimization and how some of the high-throughput technologies developed for structural genomics efforts can be used to provide fundamental data on the science of crystallization. We discuss alternative strategies and salvage pathways if a crystal is not obtained or if the crystal that does result does not yield good enough structural data to resolve the question of interest. We shift from this to diffraction and what diffraction can reveal about the crystal quality and some of the problems associated with it. We address ways to improve the quality of diffraction data at this stage then look at rapidly developing areas where smaller crystals can be utilized, e.g. microfocus capabilities down to powder techniques. Finally, we end with a short note on complementary techniques and what we perceive to be on the near-term future horizon.