If SR is not likely to be employed, then a higher resolution may be aimed for, requiring more time, and again dependent on the pressure on local resources.

Whatever the resource, it is good to define a strategy that will provide high completeness of the unique amplitudes at the highest resolution, with the realization that there may be some conflict between these two requirements owing to radiation damage.

9.1.13.5. A series of mutant or complex structures

The detailed geometry of the molecule is already known and the rather general effects of ligand binding or mutation can be initially identified at a relatively modest resolution and completeness. As with heavy-atom screening, it is often advisable to check that the desired complex or structural modification has been achieved by first recording data at low resolution.

However, if the analysis then proves to be of real chemical interest, with a need for accurate definition of structural features, the data should be subsequently extended in resolution and quality. As with the identification of isomorphous derivatives, this approach has benefited greatly from cryogenic vitrification, where the sample can be screened at low resolution and then preserved for subsequent use.

9.1.13.6. Atomic resolution applications

As for MAD data, the needs for atomic resolution data are extreme, but rather different in nature. Atomic resolution refinement is addressed in Chapter 18.4. Suffice it to say that by atomic resolution it is meant that meaningful experimental data extend close to 1 Å resolution. There are two principal reasons for recording such data. Firstly, they allow the refinement of a full anisotropic atomic model, leading to a more complete description of subtle structural features. Secondly, direct methods of phasing are dependent upon the principle of atomicity.

The problems to be faced include:

(1) The high contrast in intensities between the low- and high-angle reflections. This may be much larger than the dynamic range of the detector. If exposure times are long enough to give good counting statistics at high resolution, then the low-resolution spots will be saturated. The solution is to use more than one pass with different effective exposure times.

(2) The overall exposure time is often considerable and substantial radiation damage may finally result. The completeness of the low-resolution data is crucial, and it is strongly recommended to collect the low-resolution pass first as the time taken for this is relatively small.

(3) The close spacing between adjacent spots within the lunes on the detector, dependent on the cell dimensions. The only aid is to use fine collimation.

(4) The overlap of adjacent lunes at high diffraction angle, especially if a long cell axis lies along the beam direction. Using an alternative mount of the crystal is the simplest solution. Otherwise, the rotation range per image must be reduced, increasing the number of exposures. This was a problem with slow read-out detectors, but is largely alleviated with CCDs.

(5) For direct-methods applications, a liberal judgement of resolution limit should be adopted. Even a small percentage of meaningful reflections in the outer shells can assist the phasing. These weak shells can be rejected or given appropriate low weights in the refinement. The strong, low-resolution terms are vital for direct methods.

9.1.14. The importance of low-resolution data

The low-resolution terms define the overall shape of the object irradiated in the diffraction experiment. Omission of the low-resolution reflections, especially those with high amplitude, considerably degrades the contrast between the major features of the object and its surroundings. For a macromolecule, this means that the contrast between it and the envelope of the disordered aqueous solvent is diminished and, furthermore, the continuity of structural features along the polymeric chain may be lost. Refinement and analysis of macromolecules at all resolutions, be they high or low, involves the inspection of electron-density syntheses. These can be interpreted visually, on a graphics station, or automatically with a variety of software. In all of these, at all resolutions, the importance of the low-resolution terms is crucial. A special problem is in the interpretation of the partially ordered solvent interface. The biological activity of most enzymes and ligand-binding proteins is located precisely at this interface, and for a true structural understanding of how they function this region should be optimally defined. This is seriously impaired by the absence of the strong, low-resolution terms. The problems become more severe as the upper resolution limit of the analysis becomes poorer. Thus, at 1 Å resolution the omission of the 7 Å data shell will have less effect compared with a 3 Å analysis – but remember that ideally, no low-resolution data should be omitted!

In some phasing procedures, the presence of complete, especially high-intensity, low-resolution, data is even more crucial. The big, low-resolution amplitudes dominate the Patterson function, and methods based on the Patterson function are therefore especially sensitive. This encompasses one of the major techniques of phase determination for macromolecules: molecular replacement. Direct methods of phase determination utilize normalized structure factors and predominantly exploit those of high amplitude. The relations between the phases of those reflections with high amplitudes, such as the classical triple-product relationship, are strongest and most abundant for reflections with low Miller indices, hence at low resolution.

The importance of the low-resolution reflections in terms of geometric and qualitative context cannot be overemphasized.

9.1.15. Data quality over the whole resolution range

It is not possible to judge data quality from a single global parameter, especially $R_{\rm merge}$, not even from the overall $I/\sigma(I)$ ratio. Such a parameter may totally neglect problems such as the omission of all low-resolution terms due to detector saturation. A set of key parameters including $I/\sigma(I)$, $R_{\rm merge}$, percentage completeness, redundancy of measurements and number of overloaded high-intensity measurements must be tabulated in a series of resolution shells. This information should be assessed during data collection to guide the experimenter in the optimization of the choice of such parameters as exposure time, attainable resolution and required redundancy. As stated in Section 9.1.13, the requirements will vary with the application.

The effect of sample decay also requires such tables. The X-ray intensities decay more rapidly at high angle than at low, and consideration of this effect requires knowledge of the relative B values that need to be applied to the individual images during data scaling. An often subjective decision will need to be made regarding at what stage the decay is sufficiently high that further

images should be ignored. The effects of damage are likely to be systematic rather than just random, and cannot be totally compensated for by scaling. This remains true even for cryogenically vitrified crystals, especially with ultra-bright synchrotron sources.

Following an earlier recommendation by the IUCr Commission on Biological Molecules (Baker *et al.*, 1996), this tabulated information, as a function of resolution, should be deposited with the data and the final model coordinates in the Protein Data Bank. Only then is it possible to have a true record of the experiment and for users of the database to judge the correctness and information content of a structural analysis.

9.1.16. Strategies for automated data acquisition

Progress in crystal-handling hardware has resulted in the development of sample changers both for synchrotron beamlines and the home laboratory. A sequence of samples contained in a Dewar can be mounted on the goniostat, centred in the beam and exposed to X-rays without user intervention. The gain in efficiency arises from the fact that manual intervention is no longer needed between samples, and at SR sites access to the hutch is avoided.

While this provides the potential for greatly increased throughput, it still requires intelligent decision-making software for evaluation of crystals and for optimal strategies of data collection to be achieved with minimal (ideally zero) user input. The steps required in such a system are:

(1) Automated mounting and dismounting of samples from a Dewar.
(2) Automated centring of the sample (or at least loop) in the beam.
(3) Recording and interpretation of two images preferably 90° apart in crystal orientation.
(4) Repetition of steps (1)–(3) for a number of samples of the same protein and ranking of these samples in terms of diffraction quality.
(5) Selection of the best sample and definition of the optimum strategy for data collection, taking account of information provided by the user with regard to the minimum resolution *etc*.
(6) Collection and integration of complete and ideally redundant data.

Robotic sample changers with at least some elements of the above are now operational at many sites [for example, see Leslie *et al.* (2002); McPhillips *et al.* (2002) and Cipriani *et al.* (2006), and also Chapter 9.2]. Considerable advances are expected in the near future, allowing routine automated screening of samples at major synchrotrons. In addition, remote access to synchrotons, by submission of crystals in Dewars for collection by SR staff (the so-called Fedex procedures) or for direct control by the user from their home laboratory through the Internet, is now possible at a number of synchrotron facilities. All of these moves towards automation require electronic databases for the tracking and transfer of samples and their associated data and parameters.

9.1.17. Final remarks

Optimal strategies for data collection are dependent on a number of factors. The alternative data-collection facilities to which access is potentially available, how long it takes to gain access and the overall time allocated all place restraints on the planning of the experiment. In view of this, it is not possible to provide absolute rules for optimal strategies.

Even after the source and overall time have been allocated or planned, the strategy is still the result of a compromise between several competing requirements. Some are general, others depend on the characteristics of a particular crystal or detector. As seen in the previous section, it is not possible to define protocols relevant for all applications. Rather, it is important to consider the relative importance of the parameters that can be varied to the problem in question and make the appropriate decisions.

Thus, data collection may have become easier from a technical point of view, but several crucial scientific decisions still have to be made by the experimenter. It is always beneficial to sacrifice some beam time and interpret the initial diffraction images, so as to avoid mistakes which may have an adverse effect on data quality and the whole of the subsequent structural analysis.

References

Amemiya, Y. (1995). *Imaging plates for use with synchrotron radiation. J. Synchrotron Rad.* **2**, 13–21.

Arndt, U. W. & Wonacott, A. J. (1977). Editors. *The Rotation Method in Crystallography*. Amsterdam: North Holland.

Baker, E. N., Blundell, T. L., Vijayan, M., Dodson, E., Dodson, G., Gilliland, G. L. & Sussman, J. L. (1996). *Deposition of macromolecular data. Acta Cryst.* D**52**, 609.

Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967). *On the conformation of the hen egg-white lysozyme molecule. Proc. R. Soc. London Ser. B*, **167**, 365–377.

Bloomer, A. C. & Arndt, U. W. (1999). *Experiences and expectations of a novel X-ray microsource with focusing mirror. I. Acta Cryst.* D**55**, 1672–1680.

Bourenkov, G. P. & Popov, A. N. (2006). *A quantitative approach to data-collection strategies*. *Acta Cryst.* D**62**, 58–64.

Broennimann, Ch., Eikenberry, E. F., Henrich, B., Horisberger, R., Huelsen, G., Pohl, E., Schmitt, B., Schulze-Briese, C., Suzuki, M., Tomizaki, T., Toyokawa, H. & Wagner, A. (2006). *The PILATUS 1M detector*. *J. Synchrotron Rad.* **13**, 120–130.

Carter, C. W. Jr & Sweet, R. M. (1997). Editors. *Methods in Enzymology*, Vol. 276, pp. 183–358. San Diego: Academic Press.

Cipriani, F., Felisaz, F., Launer, L., Aksoy, J.-S., Caserotto, H., Cusack, S., Dallery, M., di-Chiaro, F., Guijarro, M., Huet, J., Larsen, S., Lentini, M., McCarthy, J., McSweeney, S., Ravelli, R., Renier, M., Taffut, C., Thompson, A., Leonard, G. A. & Walsh, M. A. (2006). *Automation of sample mounting for macromolecular crystallography*. *Acta Cryst.* D**62**, 1251–1259.

Collaborative Computational Project, Number 4 (1994). *The CCP4 suite: programs for protein crystallography. Acta Cryst.* D**50**, 760–763.

Cruickshank, D. W. J. (1999*a*). *Remarks about protein structure precision. Acta Cryst.* D**55**, 583–601.

Cruickshank, D. W. J. (1999*b*). *Remarks about protein structure precision. Erratum. Acta Cryst.* D**55**, 1108.

Dauter, Z. (1999). *Data-collection strategies. Acta Cryst.* D**55**, 1703–1717.

Dauter, Z. (2005). *Efficient use of synchrotron radiation for macromolecular diffraction data collection*. *Prog. Biophys. Mol. Biol.* **89**, 153–172.

Diederichs, K. & Karplus, P. A. (1997). *Improved R-factor for diffraction data analysis in macromolecular crystallography. Nat. Struct. Biol.* **4**, 269–275.

Diederichs, K., McSweeney, S. & Ravelli, R. B. G. (2003). *Zero-dose extrapolation as part of macromolecular synchrotron data reduction*. *Acta Cryst*. D**59**, 903–909.

Evans, G. & Walsh, M. (2005). Editors. *Data Collection and Analysis. Proceedings of the CCP4 Study Weekend*. *Acta Cryst.* D**62**, 1–124.

Garman, E. F. & Owen, R. L. (2006). *Cryocooling and radiation damage in macromolecular crystallography*. *Acta Cryst.* D**62**, 32–47.

Garman, E. F. & Schneider, T. R. (1997). *Macromolecular cryocrystallography. J. Appl. Cryst.* **30**, 211–237.