9. MONOCHROMATIC DATA COLLECTION

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 freezing, 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 largely dependent upon the principle of atomicity.

The problems likely 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 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 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 is again a problem with slow read-out detectors.

(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 it high or low, involves the inspection of electron-density syntheses. These can be interpreted visually, on a graphics station, or interpreted 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_{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_{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 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 frozen 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. 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.

Synchrotron beamlines become brighter, detectors faster and data-processing software ever more sophisticated. Existing software has advanced to the stage where many decisions regarding the geometric restraints on data completeness and minimalist data collection are automatically proposed to the user. Decisions regarding the qualitative completeness, with respect to the optimum resolution limit, exposure time and redundancy, are more nebulous concepts and are not yet addressed in an automated manner. This must be the area of major advance in the next years.

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.

Amemiya, Y. & Miyahara, J. (1988). *Imaging plates illuminate many fields. Nature (London)*, **336**, 89–90.

Arndt, U. W., Duncumb, P., Long, J. V. P., Pina, L. & Inneman, A. (1998). *Focusing mirrors for use with microfocus X-ray tubes. J. Appl. Cryst.* **31**, 733–741.

Arndt, U. W., Long, J. V. P. & Duncumb, P. (1998). *A microfocus X-ray tube used with focusing collimators. J. Appl. Cryst.* **31**, 936–944.

Arndt, U. W. & Willis, B. T. M. (1966). *Single crystal diffractometry.* Cambridge University Press.

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.

Buerger, M. J. (1964). *The precession method.* New York: Wiley.

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

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., Terry, H., Witzel, H. & Wilson, K. S. (1990). *Refinement of glucose isomerase from Streptomyces albus at 1.65 Å with data from an imaging plate. Acta Cryst.* B**46**, 833–841.

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

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

Gruner, S. M. & Ealick, S. E. (1995). *Charge coupled device X-ray detectors for macromolecular crystallography. Structure*, **3**, 13–15.

Hamlin, R. (1985). *Multiwire area X-ray diffractometers. Methods Enzymol.* **114**, 416–452.

Helliwell, J. R. (1992). *Macromolecular crystallography with synchrotron radiation.* Cambridge University Press.

Henderson, R. (1990). *Cryo protection of protein crystals against radiation damage in electron and X-ray diffraction. Proc. R. Soc. London Ser. B*, **241**, 6–8.

Hendrickson, W. A. (1991). *Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science*, **254**, 51–58.

Howell, P. L. & Smith, G. D. (1992). *Identification of heavy-atom derivatives by normal probability methods. J. Appl. Cryst.* **25**, 81–86.

International Tables for Crystallography (1995). Vol. A. *Space-group symmetry*, edited by Th. Hahn. Dordrecht: Kluwer Academic Publishers.

International Tables for Crystallography (1999). Vol. C. *Mathematical, physical and chemical tables*, edited by A. J. C. Wilson & E. Prince. Dordrecht: Kluwer Academic Publishers.

Kabsch, W. (1988). *Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. J. Appl. Cryst.* **21**, 916–924.

Matthews, B. W. (1968). *Solvent content in protein crystals. J. Mol. Biol.* **33**, 491–497.

Read, R. J. (1999). *Detecting outliers in non-redundant diffraction data. Acta Cryst.* D**55**, 1759–1764.

Rodgers, D. W. (1997). *Practical cryocrystallography. Methods Enzymol.* **276**, 183–203.

Rosenbaum, G., Holmes, K. C. & Witz, J. (1971). *Synchrotron radiation as a source for X-ray diffraction. Nature (London)*, **230**, 434–437.

Sakabe, N. (1991). *X-ray diffraction data collection system for modern protein crystallography with a Weissenberg camera and an imaging plate using synchrotron radiation. Nucl. Instrum. Methods A*, **303**, 448–463.

Sheldrick, G. M. (1990). *Phase annealing in SHELX-90: direct methods for larger structures. Acta Cryst.* A**46**, 467–473.

Smith, J. L. (1991). *Determination of three-dimensional structure by multiwavelength anomalous diffraction. Curr. Opin. Struct. Biol.* **1**, 1002–1011.

Turkenburg, J., Brady, L., Bailey, S., Ashton, A., Broadhurst, P. & Brown, D. (1999). Editors. *Data collection and processing. Proceedings of the CCP4 study weekend. Acta Cryst.* D**55**, 1631–1772.

Wakatsuki, S., Belrhali, H., Mitchell, E. P., Burmeister, W. P., McSweeney, S. M., Kahn, R., Bourgeois, D., Yao, M., Tomizaki, T. & Theveneau, P. (1998). *ID14 'Quadriga', a beamline for protein crystallography at the ESRF. J. Synchrotron Rad.* **5**, 215–221.

Weiss, M. S. & Hilgenfeld, R. (1997). *On the use of the merging R factor as a quality indicator for X-ray data. J. Appl. Cryst.* **30**, 203–205.

195

references