guarantee, that the derivative is worth pursuing. Normal probability plots can be helpful in this respect (Howell & Smith, 1992).

(4) Given a positive result from point (3), complete data may be recorded on the same or an equivalent crystal. Again, it may be useful to record data to low resolution in the first instance. 4 Å resolution is again quite sufficient to solve the structure of a heavy-atom constellation using direct or Patterson methods, allowing the more complete characterization of the potential derivative.

(5) If the compound proves to be a useful derivative, data can then be recorded to higher resolution for the computation of phase information. It may not be appropriate to record data to the highest resolution as for the native protein. In this context, the strength of the data is of primary importance, and relatively weak data at high resolution may be less relevant.

Some practical points are highly relevant here. The ability to store and reuse frozen crystals means that potential derivatives can first be screened at the lowest possible resolution, and the crystal preserved and used later only if the derivative proves to provide useful phase information. The final resolution for data collection will then depend on the degree of isomorphism. The wavelength, if tunable, should be set to a value just below the absorption edge in order to maximize the anomalous signal. The redundancy can also play an important role, as it is useful to have a large number of independent measurements so that outliers in the native or derivative data can be excluded, as these can cause major problems in either the Patterson or direct-methods approaches for locating the heavy atom (Part 12).

9.1.13.2. *Anomalous scattering, MAD and SAD*

The requirements for collecting data with an intrinsically weak anomalous signal are several. As with the isomorphous measurements in the previous section, the highest possible resolution may not be the primary consideration. Here the emphasis lies in data quality, as the measurement of very small differences in macromolecular amplitudes, which are already in themselves relatively weak, is required. Important considerations include the following.

(1) Optimization of the wavelength, particularly for MAD experiments.

(2) Ensuring that the anomalous data are complete in terms of all possible Bijvoet pairs. This is not always addressed by the currently available data-processing software.

(3) High redundancy of measurements significantly enhances the quality of the signal, as this provides effective averaging of errors and allows the rejection of statistical outliers. The latter is especially important for direct-methods solution of the anomalous-scattering constellation.

For MAD experiments (Hendrickson, 1991; Smith, 1991), which can only be carried out at SR sites, the optimum number of wavelengths at which data should be recorded remains unclear. The minimum is one (SAD) and the conventional wisdom is that four are optimal. Given finite beam time, the trade-off is between measuring with limited redundancy at several wavelengths as against higher redundancy at a smaller number of wavelengths. The jury is still out on this one.

Single-wavelength anomalous dispersion (SAD) represents the limiting case. All data are recorded at one wavelength, reducing the requirement for fine monochromatization and for fine tunability and stability. Now quality, especially in the form of redundancy, is the dominating factor since all phasing is based purely on a single anomalous difference for each reflection.

9.1.13.3. *Molecular replacement*

For the initial data required for molecular replacement (MR), high resolution is not essential. Firstly, the method depends on homologous models that are usually only an imperfect representation of the structure under investigation and hence high-resolution data cannot be accurately modelled, and will only introduce noise into the analysis. Secondly, the rotation function, the first step in MR, is based on the representation of the Patterson function in terms of spherical harmonics, which is limited in its accuracy.

In contrast, it is essential for MR applications that the most intense low-resolution terms are measured. The lack of such reflections strongly affects the rotation- and translation-function computations, as the functions are based on Patterson syntheses involving the square of the structure-factor amplitudes, and are dominated by the largest terms. Elimination of the strongest few per cent of the low-resolution data may well prevent a successful solution by MR.

However, for refinement of structures solved by MR, it is essential that data be recorded to a resolution sufficient to allow escape from the phase bias introduced by the model.

9.1.13.4. *Definitive data on relevant biological structures*

Here it is intended to include all structures that benefit from the highest accuracy in their atomic coordinates to shed light on the details of their biological function. These may include substrate or inhibitor complexes and mutants if the analysis requires the full potential of X-ray crystallography. Many of these will not diffract to atomic resolution; nevertheless, all steps in a detailed crystal structure analysis are made simpler as the resolution and quality of the data are increased. This includes the solution of the phase problem, interpretation of the electron-density maps and the refinement of the model.

The most appropriate strategy for data collection involves decisions based on a complex and mutually dependent set of parameters including:

(1) Crystal quality and availability. If only one crystal is available, the choices are limited. If many are available, then some experimentation is recommended to select a high-quality sample.

(2) Cryogenic freezing. This has become *de rigueur* for the modern protein crystallographer. In many cases it allows collection of data from a single crystal. If appropriate cryogenic freezing conditions cannot be established, making it necessary to record room-temperature data, this can affect strategy-making dramatically, in that several crystals might well be required to achieve the target resolution and completeness.

(3) X-ray source and detector. The availability of these again places restrictions on the experiments which are tractable. An SR source will always provide better data, but has logistical problems of availability and access. For some problems, SR becomes *sine qua non* and a rotating anode is just insufficient. These include the use of MAD techniques, very small crystals, large and complex structures with large unit cells such as viruses, and where atomic resolution data are needed.

(4) Overall data-collection time allocated. This has an obvious overlap with point (3). In particular, if SR is to be used later, then the resolution limit on the home source may be modest. 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 is some conflict between these two requirements.

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

references