International Tables for Crystallography (2006). Vol. F, Section 14.2.1, pp. 299–303.

14.2. MAD and MIR

BY J. L. SMITH, W. A. HENDRICKSON, T. C. TERWILLIGER AND J. BERENDZEN

14.2.1. Multiwavelength anomalous diffraction

(J. L. SMITH AND W. A. HENDRICKSON)

Anomalous-scattering effects measured at several X-ray wavelengths can provide a direct solution to the crystallographic phase problem. For many years this was appreciated as a hypothetical possibility (Okaya & Pepinsky, 1956), but, until tunable synchrotron radiation became available, experimental investigation with the weakly diffracting crystals of biological macromolecules was limited to one heroic experiment (Hoppe & Jakubowski, 1975). Multiwavelength anomalous diffraction (MAD) became a dominant phasing method in macromolecular crystallography with the advent of reliable, brilliant synchrotron-radiation sources, the adoption of cryopreservation techniques for crystals of macromolecules, and the development of general anomalous-scatterer labels for proteins and nucleic acids.

Anomalous scattering, first recognized as a source of phase information by Bijvoet (1949), has been employed since the early days of macromolecular crystallography (Blow, 1958). It has been used to locate positions of anomalous scatterers (Rossmann, 1961), to supplement phase information from isomorphous replacement (North, 1965; Matthews, 1966*a*) and to identify the enantiomorph of the heavy-atom partial structure in multiple isomorphous replacement (MIR) phasing (Matthews, 1966*b*). Anomalous scattering at a single wavelength was the sole source of phase information in the structure determination of crambin (Hendrickson & Teeter, 1981), an important precursor to development of MAD. MAD differs from these other applications in using anomalous scattering at several wavelengths for complete phase determination without approximations or simplifying assumptions.

14.2.1.1. *Anomalous scattering factors*

The scattering of X-rays by an isolated atom is described by the atomic scattering factor, f^0, based on the assumption that the electrons in the atom oscillate as free electrons in response to X-ray stimulation. The magnitude of f^0 is normalized to the scattering by a single electron. Thus the 'normal' scattering factor f^0 is a real number, equal to the Fourier transform of the electron-density distribution of the atom. At zero scattering angle ($s = \sin\theta/\lambda = 0$), f^0 equals Z, the atomic number. f^0 falls off rapidly with increasing scattering angle due to weak scattering by the diffuse parts of the electron-density distribution. In reality, electrons in an atom do not oscillate freely because they are bound in atomic orbitals. Deviation from the free-electron model of atomic scattering is known as anomalous scattering. Using a classical mechanical model (James, 1948), an atom scatters as a set of damped oscillators with resonant frequencies matched to the absorption frequencies of the electronic shells. The total atomic scattering factor, f, is thus a complex number. f is denoted as a sum of 'normal' and 'anomalous' components, where the anomalous components are corrections to the free-electron model:

$$f = f^0 + f' + if''. \qquad (14.2.1.1)$$

f' and f'' are expressed in electron units, as is f^0. The real component of anomalous scattering, f', is in phase with the normal scattering, f^0, whilst the imaginary component, f'', is out of phase by $\pi/2$.

The imaginary component of anomalous scattering, f'', is proportional to the atomic absorption coefficient of the atom, μ_a, at X-ray energy E:

$$f''(E) = (mc/4\pi e^2\hbar)E\mu_a(E), \qquad (14.2.1.2)$$

where m is the electronic mass, c is the speed of light, e is the electronic charge and h $(= 2\pi\hbar)$ is Planck's constant. Thus, f'' can be determined experimentally by measurement of the atomic absorption coefficient. The relationship between f'' and f' is known as the Kramers–Kronig dispersion relation (James, 1948; Als-Nielsen & McMorrow, 2001):

$$f'(E) = \left(\frac{2}{\pi}\right)P\int_0^\infty \frac{E'f''(E')}{E^2 - E'^2}\,\mathrm{d}E', \qquad (14.2.1.3)$$

where P represents the Cauchy principal value of the integral such that integration over E' is performed from 0 to $(E - \varepsilon)$ and from $(E + \varepsilon)$ to ∞, and then the limit $\varepsilon \to 0$ is taken. The principal value of the integral can be evaluated numerically from limited spectral data that have been scaled to theoretical f'' scattering factors (or μ_a absorption coefficients) at points remote from the absorption edge.

Anomalous scattering is present for all atomic types at all X-ray energies. However, the magnitudes of f' and f'' are negligible at X-ray energies far removed from the resonant frequencies of the atom. This includes all light atoms (H, C, N, O) of biological macromolecules at all X-ray energies commonly used for crystallography. f' and f'' are rather insensitive to scattering angle, unlike f^0, because the electronic resonant frequencies pertain to inner electron shells, which have radii much smaller than the X-ray wavelengths used for anomalous-scattering experiments. The magnitudes of f' and f'' are greatest at X-ray energies very near resonant frequencies, and are also highly energy-dependent (Fig. 14.2.1.1). This property of anomalous scattering is exploited in MAD.

Three means are available for evaluating anomalous scattering factors, f' and f''. Calculations from first principles on isolated elemental atoms are accurate for energies remote from resonant frequencies (Cromer & Liberman, 1970*a*,*b*). However, these calculated values do not apply to the energies most critical in a MAD experiment. f' and f'' can also be estimated by fitting to diffraction data measured at different energies (Templeton *et al.*, 1982). Finally, f'' can be obtained from X-ray absorption spectra by the equation above, and f' from f'' by the Kramers–Kronig transform [equation (14.2.1.3); Hendrickson *et al.*, 1988; Smith, 1998]. Both the precise position of a resonant frequency and the values of f' and f'' near resonance generally depend on transitions to unoccupied molecular orbitals, and are quite sensitive to the electronic environment surrounding the atom. Complexities in the X-ray absorption edge, particularly so-called 'white lines', can enhance the anomalous scattering considerably (Fig. 14.2.1.1). Thus, experimental measurements are needed to select wavelengths for optimal signals, and the values of f' and f'' should be determined either from an absorption spectrum or by refinement against the diffraction data.

X-ray spectra near absorption edges of anomalous scatterers depend on the orientation of the local chemical environment in the X-ray beam, which is polarized for synchrotron radiation. The anisotropy of anomalous scattering may affect both the edge position and the magnitude of absorption. In such cases, f' and f'' for individual atoms are also dependent on orientation. Orientational averaging due to multiple anomalous scatterer sites or crystallographic symmetry may prevent macroscopic detection of polarization effects in crystals. A formalism to describe anisotropic anomalous scattering in which f' and f'' are tensors has been

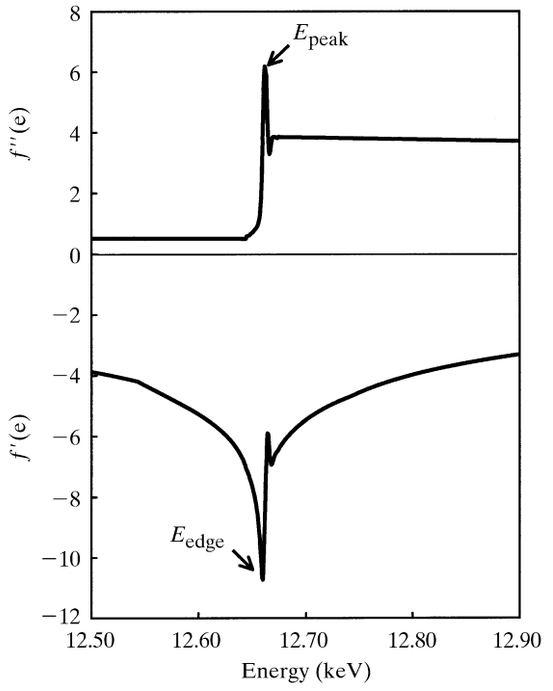

Fig. 14.2.1.1. Anomalous scattering factors for Se in a protein labelled with selenomethionine. The spectra are a hybrid of experimental values derived from an absorption spectrum of a SeMet protein for energies near the Se *K* absorption edge and calculated values for energies remote from the edge. The Se *K* edge occurs at 12 660 eV, corresponding to a wavelength of 0.9800 Å. Anomalous-scattering effects are enhanced by a white line just above the edge. The position of the absorption edge (E_{edge}) is the inflection point of the μ_a (and f'') spectrum, and E_{peak} is the energy of peak absorption just above the edge. These energies correspond to the wavelengths λ_{edge} and λ_{peak} in a MAD experiment because the magnitudes of f' and f'' are greatest at E_{edge} and E_{peak}, respectively.

developed by Templeton & Templeton (1988). Fanchon & Hendrickson (1990) have developed a technique to refine f' and f'' tensors against MAD data. Although anomalous scattering labels such as the commonly used selenomethionine and related compounds are strongly anisotropic (Templeton & Templeton, 1988; Hendrickson *et al*., 1990), anisotropy is generally ignored in MAD for biological macromolecules.

14.2.1.2. *A phase equation for MAD*

The impact of anomalous scattering on diffraction measurements can be evaluated by substituting the scattering-factor expression [equation (14.2.1.1)] into the structure-factor equation.

$$ {}^{\lambda}F_{\rm obs}(\mathbf{h}) = \sum_{i=1}^{\text{No. of atoms}} (f^0 + f'_\lambda + if''_\lambda)_i \exp[-B_i s^2(\mathbf{h})] \exp(2\pi i \mathbf{h}\cdot\mathbf{x}_i). \quad (14.2.1.4)$$

It is convenient to separate terms in the structure-factor expression according to wavelength dependence (Karle, 1980). A wavelength-independent structure factor ${}^0F_T$ with phase φ_T is defined to represent the total normal scattering from all atoms (Hendrickson, 1985) as

$$ \begin{aligned} {}^0F_T(\mathbf{h}) &= \sum_{i=1}^{\text{No. of atoms}} f_i^0 \exp\left[-B_i s^2(\mathbf{h})\right] \exp(2\pi i \mathbf{h}\cdot\mathbf{x}_i) \\ &= |{}^0F_T| \exp(i\varphi_T). \end{aligned} \quad (14.2.1.5)$$

The anomalous scattering is confined to wavelength-dependent 'anomalous' structure factors ${}^{\lambda}\mathbf{F}'$ and ${}^{\lambda}\mathbf{F}''$, representing the real and imaginary components of anomalous scattering for all atoms. In general, the anomalous structure factors are considered for only the subset of atoms with detectable anomalous scattering, leading to the anomalous structure factors ${}^{\lambda}\mathbf{F}'_A$ and ${}^{\lambda}\mathbf{F}''_A$:

$$ {}^{\lambda}F'_A(\mathbf{h}) = \sum_{j=1}^{N_{\rm ano}} f'_{\lambda j} \exp\left[-B_j s^2(\mathbf{h})\right] \exp(2\pi i \mathbf{h}\cdot\mathbf{x}_j) \quad (14.2.1.6)$$

$$ {}^{\lambda}F''_A(\mathbf{h}) = \sum_{j=1}^{N_{\rm ano}} if''_{\lambda j} \exp\left[-B_j s^2(\mathbf{h})\right] \exp(2\pi i \mathbf{h}\cdot\mathbf{x}_j), \quad (14.2.1.7)$$

where $N_{\rm ano}$ is the number of anomalous scatterers. The anomalous structure factors ${}^{\lambda}F'_A$ and ${}^{\lambda}F''_A$ can also be expressed in terms of the normal structure-factor components, ${}^0F_{A_k}$ with phase φ_{A_k}, for the subsets of atoms that comprise each kind of anomalous scatterer (Karle, 1980),

$$ {}^{\lambda}F'_A = \sum_{k=1}^{\text{No. of kinds}} (f'_\lambda/f^0)_k \, {}^0F_{A_k} \quad (14.2.1.8)$$

$$ {}^{\lambda}F''_A = \sum_{k=1}^{\text{No. of kinds}} (if''_\lambda/f^0)_k \, {}^0F_{A_k}. \quad (14.2.1.9)$$

Thus, the wavelength-dependent experimental structure factor ${}^{\lambda}F_{\rm obs}$ can be represented using normal structure factors only:

$$ {}^{\lambda}F_{\rm obs} = {}^0F_T + \sum_{k=1}^{\text{No. of kinds}} \left(\frac{f'_\lambda}{f^0} + i\frac{f''_\lambda}{f^0}\right)_k {}^0F_{A_k} \quad (14.2.1.10)$$

This factorization is convenient because all wavelength dependence is confined to the anomalous scattering factors f' and f'', which are independent of atomic positions, occupancies and thermal parameters. In addition, an electron-density map for the total structure should be based on normal scattering by all atoms, represented by the structure factor ${}^0F_T$ with phase φ_T. As described in Section 14.2.1.1, the normal scattering factor f^0 is strongly dependent on scattering angle, whereas f' and f'' are nearly invariant with s.

A very useful observational equation is obtained by applying the law of cosines to ${}^0F_T$, ${}^{\lambda}F'_A$ and ${}^{\lambda}F''_A$ (Karle, 1980; Hendrickson, 1985). Anomalous structure factors are treated separately for each kind of anomalous scatterer. The number of terms in the resulting expression is $(q+1)^2$ for q kinds of anomalous scatterers. For the commonest case of one kind of anomalous scatterer

$$ \begin{aligned} |{}^{\lambda}F_{\rm obs}(\pm\mathbf{h})|^2 = {} & |{}^0F_T|^2 + a_\lambda |{}^0F_A|^2 \\ & + b_\lambda |{}^0F_T| \, |{}^0F_A| \cos(\varphi_T - \varphi_A) \\ & \pm c_\lambda |{}^0F_T| \, |{}^0F_A| \sin(\varphi_T - \varphi_A), \end{aligned} \quad (14.2.1.11)$$

where $a_\lambda = (f''^2_\lambda + f'^2_\lambda)/(f^0)^2$, $b_\lambda = 2f'_\lambda/f^0$ and $c_\lambda = 2f''_\lambda/f^0$. $|{}^{\lambda}F_{\rm obs}(\pm\mathbf{h})|$ refers to the Bijvoet mate reflections $+\mathbf{h}$ and $-\mathbf{h}$. The MAD observational equation illustrates the orthogonal contributions to phasing made by the real (f') and imaginary (f'') components of anomalous scattering. 'Dispersive' phase information derives from differences between $|F_{\rm obs}|$ at wavelengths having different values of f' and contributes to $\cos(\varphi_T - \varphi_A)$. 'Bijvoet' phase information derives from the Friedel difference $|F_{\rm obs}(+\mathbf{h})| - |F_{\rm obs}(-\mathbf{h})|$ at a wavelength with substantial f'' and contributes to $\sin(\varphi_T - \varphi_A)$. Phase information is enhanced by selection of wavelengths for data measurement that maximize the magnitudes of both f' and f''. Apart from ignoring the very weakest anomalous-scattering effects, the MAD observational equation (14.2.1.11) involves no approximations.

14.2.1.3. *Diffraction ratios for estimating the MAD phasing signal*

The first consideration in design of a MAD experiment is the choice of anomalous scatterer(s) with consideration of the magnitude of the phasing signal. Estimation of the total scattering by the macromolecule and the potential phasing signal generated by the anomalous scatterer(s) under consideration is informative.

The magnitude of the MAD phasing signal is estimated as the ratio of the expected dispersive or Bijvoet difference to the expected total scattering of the macromolecule. This is based on calculation of the expected root-mean-square structure amplitude $(\mathrm{rms}|F|)$ (Wilson, 1942).

$$\mathrm{rms}|F| = \langle |F|^2 \rangle^{1/2} = \left(\sum f_i^2\right)^{1/2} = N^{1/2} f \qquad (14.2.1.12)$$

for N identical atoms. The expected total scattering of the macromolecule is estimated at $s = 0$ using an average non-hydrogen atom. Based on atomic frequencies in biological macromolecules, the average values of f^0 are 6.70 e for proteins, 7.20 e for DNA and 7.26 e for RNA. The average number of non-hydrogen atoms and molecular mass per residue are 7.7 atoms and 110 Da for proteins, 21.8 atoms and 292 Da for DNA, and 22.4 atoms and 304 Da for RNA. These averages result in the following expressions for estimated total scattering of biological macromolecules:

$$\begin{aligned}
\mathrm{rms}|{}^0F_T|_{\mathrm{protein}} &\approx 6.70(\text{No. of atoms})^{1/2} \approx (346 \times \text{No. of amino acids})^{1/2} \\
&\approx (3.14 \times \text{molecular mass})^{1/2} \\
\mathrm{rms}|{}^0F_T|_{\mathrm{DNA}} &\approx 7.20(\text{No. of atoms})^{1/2} \approx (1128 \times \text{No. of nucleotides})^{1/2} \\
&\approx (3.87 \times \text{molecular mass})^{1/2} \\
\mathrm{rms}|{}^0F_T|_{\mathrm{RNA}} &\approx 7.26(\text{No. of atoms})^{1/2} \approx (1183 \times \text{No. of nucleotides})^{1/2} \\
&\approx (3.89 \times \text{molecular mass})^{1/2}. \qquad (14.2.1.13)
\end{aligned}$$

Note: the estimated total scattering of a protein is coincidentally $\approx (\pi \times \text{molecular mass})^{1/2}$.

The diffraction ratios relevant to a MAD experiment with N anomalous-scatterer sites are

$$\frac{\mathrm{rms}\|{}^{\lambda 1}F_{\mathrm{obs}}| - |{}^{\lambda 2}F_{\mathrm{obs}}\|}{\mathrm{rms}|{}^0F_T|} \approx (N/2)^{1/2} \frac{|f'_{\lambda 1} - f'_{\lambda 2}|}{\mathrm{rms}|{}^0F_T|} \qquad (14.2.1.14)$$

for the dispersive signal and

$$\frac{\mathrm{rms}\|{}^{\lambda}F^{+}_{\mathrm{obs}}| - |{}^{\lambda}F^{-}_{\mathrm{obs}}\|}{\mathrm{rms}|{}^0F_T|} \approx (N/2)^{1/2} \frac{2f''_{\lambda}}{\mathrm{rms}|{}^0F_T|} \qquad (14.2.1.15)$$

for the Bijvoet signal. The diffraction ratios, analogous to similar relations for isomorphous replacement (Crick & Magdoff, 1956), are equivalent to the expected fractional changes in intensity due to anomalous scattering, and, as such, can be compared directly to the R_{sym} estimate of error in the experimental data for evaluation of the phasing signal. Of course, the phasing signal may be diminished by partial occupancy or thermal motion, as for normal scattering.

14.2.1.4. *Experimental considerations*

The design and execution of a MAD experiment are distinguished from monochromatic experiments in macromolecular crystallography primarily by the stringent criteria for wavelength selection.

The largest MAD phasing signal is obtained at energies with the most extreme values of f' and f'', which correspond to the sharpest features of the absorption edge (Fig. 14.2.1.1). The energy of peak absorption just above the edge (E_{peak}) corresponds to the wavelength of maximum f'' and optimal Bijvoet signal $(\lambda_{\mathrm{peak}})$. Typically, the orthogonal dispersive signal is optimized by recording one data set at the wavelength corresponding to the inflection point of the absorption edge (minimum f', λ_{edge}), and one or more data sets at remote wavelengths having f' with smaller magnitudes $(\lambda_{\mathrm{remote}})$. The choice of the remote wavelength(s) is experiment dependent. If only one remote wavelength is used, it is typically on the high-energy side of the absorption edge due to the larger Bijvoet signal. The remote wavelength(s) may also be chosen to avoid complications from other edges or to obtain data at a wavelength optimal for model refinement. In the case of anomalous scatterers that exhibit sharp 'white line' features, the dispersive signal may be optimized between the minimum of f' at the ascending edge $(\lambda_{\mathrm{edge}})$ and the local maximum of f' at the descending inflection point $(\lambda_{\mathrm{descent}})$.

The features of an X-ray absorption edge are in many cases very sharp, with the energies of the inflection point and peak absorption separated by as little as 2 eV. Therefore, it is critical to determine λ_{edge} and λ_{peak} experimentally by recording the absorption edge from the labelled macromolecule at the time of a MAD experiment. Even when the position of the edge is well known, small unanticipated chemical changes in the sample or calibration errors in the X-ray beam can reduce the MAD signal very significantly.

The MAD phasing signal is derived from intensity differences that may be similar in magnitude to measurement errors. Thus a general philosophy in the design of a MAD experiment is to equalize systematic errors among the measurements whose differences will contribute to each phase determination. This is achieved for each unique reflection by recording Bijvoet measurements at all wavelengths from the same asymmetric portion of diffraction space at nearly the same time. If crystal decay necessitates use of multiple crystals in a MAD experiment, blocks of Bijvoet data should be recorded identically at all of the selected wavelengths from each crystal contributing to the data set. Bijvoet mates can be recorded simultaneously by alignment of the crystal with a mirror plane of diffraction symmetry perpendicular to the rotation axis, or Friedel images can be recorded in an 'inverse beam' experiment. Inverse-beam geometry is a hypothetical method for measurement of Friedel data using both the forward and reverse directions of the incident X-ray beam. In a real experiment, diffraction images and their Friedel equivalents are recorded at crystal positions related by 180° rotation about any axis perpendicular to the incident beam, usually the data-collection axis. The inverse-beam experiment requires neither crystal symmetry nor crystal alignment, and is well suited to crystals mounted in random orientations.

The multiwavelength measurements for each unique reflection will be identically redundant and have nearly equal systematic errors if identical blocks of Bijvoet data are collected, as described above. When such a data-collection strategy is followed, the resulting MAD data set will include all multiwavelength Bijvoet measurements for all regions of the reciprocal lattice that are covered in the experiment.

Cryopreservation of crystals is of enormous benefit to MAD. Systematic error due to radiation damage is eliminated or greatly diminished. Systematic differences between crystals are eliminated in cases where a complete MAD data set is measured from a single frozen crystal. Intensities of weak reflections are estimated more accurately because less material contributes to diffuse background scatter in the mounts used for frozen crystals than for unfrozen crystals.

Measurement errors are of major importance in all areas of macromolecular crystallography, but are the limiting factor in phase determination by MAD. MAD data should be of high quality by the usual measures (R_{sym}, redundancy, completeness), especially in experiments where the phasing signal is weak. Good counting statistics are of paramount importance. Experimental error,

estimated by $R_{\rm sym}$, increases with increasing scattering angle because of the strong fall-off of f^0 with s. In a carefully designed experiment, the effect of increasing $R_{\rm sym}$ with s is mitigated somewhat by equalizing systematic errors and by averaging highly redundant data. Disappearance of the phasing signal into $R_{\rm sym}$ noise is the major reason that useful MAD phases are not obtained to the diffraction limit of crystals, even though anomalous scattering does not diminish with increasing s.

The optimal number of data-collection wavelengths for successful phase determination by MAD has been debated. In most cases, it is necessary to measure data at $\lambda_{\rm edge}$, $\lambda_{\rm peak}$ and a $\lambda_{\rm remote}$ in order to take advantage of the most extreme values of f' and f''. If f' values at $\lambda_{\rm edge}$ and $\lambda_{\rm peak}$ are different enough to produce a detectable dispersive signal, then phases can be obtained from three measurements: $|F^+|$ and $|F^-|$ at $\lambda_{\rm peak}$, and either $|F^+|$ or $|F^-|$ at $\lambda_{\rm edge}$. However, redundancy is one of the best ways to minimize the effects of measurement error in macromolecular crystallography. Redundant Bijvoet signals can be obtained at $\lambda_{\rm peak}$ and at any $\lambda_{\rm remote}$ above the absorption edge if both $|F^+|$ and $|F^-|$ are measured at each wavelength. Likewise, the dispersive signal between measurements at $\lambda_{\rm edge}$ and $\lambda_{\rm remote}$ is also redundant if both $|F^+|$ and $|F^-|$ measurements are taken at each wavelength. More highly redundant four- or five-wavelength MAD experiments may be advantageous, although greater redundancy should not be gained at the cost of good counting statistics. Brilliant synchrotron sources and high-speed detectors make rapid measurement of complete multi-wavelength data sets possible, but the practical feasibility is often compromised by radiation damage. Phase information from the Bijvoet signal at a single wavelength (preferably $\lambda_{\rm peak}$) can also be used as a basis for structure determination. The phase probability distribution from single-wavelength anomalous scattering is in general bimodal, and must be resolved with additional phase information. This could be the partial structure of anomalous scatterers, as in the classic crambin experiment (Hendrickson & Teeter, 1981), or the real-space constraints, such as solvent flattening or redundancy averaging, that are applied in common schemes for phase refinement by density modification.

14.2.1.5. *Data handling*

Two general approaches to data handling for MAD have been employed.

An extreme interpretation of the scheme for equalizing systematic errors is known as 'phase first, merge later' (Hendrickson, 1985; Hendrickson & Ogata, 1997). The idea is that systematic errors may be amplified by merging data, and that this may obscure a weak phasing signal. In this approach, the individual observations constituting a multiwavelength Bijvoet set, as determined by the data-collection strategy, are grouped together and scaled. There may be redundant multiwavelength sets of observations, but these are merged only after individual phase evaluations have been made. Error estimates from the phasing, or the agreement of redundant phase determinations, can be incorporated into weights for averaging, or can be used to reject outliers. Complicated, experiment-dependent book-keeping is required to assemble exactly the correct observations into each unmerged set of multiwavelength measurements. However, the 'phase first, merge later' approach may be advantageous for MAD data sets from multiple crystals, or when minor disasters disrupt the experiment and thwart the data-collection strategy.

A second approach, known as 'merge first, phase later', is to scale and merge data at each wavelength, keeping Bijvoet pairs separate, and then to scale data at all wavelengths to one another (Ramakrishnan & Biou, 1997). The idea is that the multiwavelength Bijvoet measurements are identically redundant for each unique reflection if the MAD data were measured according to the strategy outlined in Section 14.2.1.4. Thus, merging the redundancies should reduce systematic errors in the amplitude differences used for phasing. The 'merge first, phase later' approach is computationally simpler than the 'phase first, merge later' approach because it is experiment independent. However, unanticipated experimental disasters may be more difficult to overcome in the 'merge first, phase later' approach to data handling.

Of course, if the MAD signal is strong relative to the experimental error, either approach to data handling should be successful. Data scaling in both approaches may be done most easily and reliably by scaling all data against a standard data set, such as the unique data from one wavelength with Bijvoet mates averaged. In general a dogmatic approach to data handling is best avoided in favour of whichever computational technique or combination of techniques is most suited to the problem at hand. Factors such as the strength of the MAD signal, data-collection strategy, number of crystals contributing to the data set, crystal quality and experimental disasters should be taken into account.

14.2.1.6. *Approaches to MAD phasing*

There are two general approaches to MAD phasing. In the explicit approach, the MAD observational equation is solved directly (Hendrickson *et al.*, 1988; Hendrickson & Ogata, 1997). In the pseudo-MIR approach, MAD phasing is treated as a special case of multiple isomorphous replacement (Burling *et al.*, 1996; Terwilliger, 1997; Ramakrishnan & Biou, 1997). Both approaches have been quite successful, and each has advantages and disadvantages. For complete phase determination by either method, the partial structure of the anomalous scatterers must be determined. The explicit and pseudo-MIR approaches differ in when the partial structure is determined and in how it is refined.

The explicit approach provides the quantities $|^0F_T|$, $|^0F_A|$ and $(\varphi_T - \varphi_A)$ by direct fit of the $|^\lambda F_{\rm obs}|$ to the MAD observational equation (14.2.1.11). No anomalous-scatterer partial structure model is required in this first step of phasing. Estimates of the anomalous scattering factors at the wavelengths of data collection are required. These estimates can be refined (Weis *et al.*, 1991), so they need not be highly accurate. Redundancies are merged to produce a unique data set at the level of the derived quantities $|^0F_T|$, $|^0F_A|$, $(\varphi_T - \varphi_A)$ and their error estimates. The anomalous-scatterer partial structure is determined from the derived estimates of $|^0F_A|$ and refined against these amplitudes. In the second step of phasing, φ_T is derived from the phase difference $(\varphi_T - \varphi_A)$ and weights are calculated for a Fourier synthesis from $|^0F_T|$ and φ_T. Phase probability distributions (*ABCD* coefficients; Hendrickson & Lattman, 1970) derived from the MAD observational equation (14.2.1.11) can be used directly in the explicit approach (Pähler *et al.*, 1990). A probabilistic treatment based on maximum likelihood theory has also been developed (de La Fortelle & Bricogne, 1997). There are two advantages to the explicit approach. First, it is amenable to the 'phase first, merge later' scheme of data handling because refinement of the anomalous-scatterer partial structure is entirely separate from phase calculation. The second principal advantage of the explicit approach is the calculation of an experimentally derived estimate of the normal structure amplitude $|^0F_A|$ for the anomalous scatterer. This is the quantity with which the partial structure of anomalous scatterers is most directly solved and refined. However, extraction of reliable $|^0F_A|$ estimates from data with low signal-to-noise can be difficult. Bayesian methods of $|^0F_A|$ estimation (Terwilliger, 1994*a*; Krahn *et al.*, 1999) have been shown to be more robust than least-squares methods.

In the pseudo-MIR approach, data at one wavelength are designated as 'native' data, which include anomalous scattering, and data at the other wavelengths as 'derivative' data. This approach has the advantage that nothing need be known about the

anomalous scattering factors at any time during phasing. These quantities are incorporated into heavy-atom atomic 'occupancies' and refined along with other parameters. Of course, the partial structure of anomalous scatterers must be known, and its refinement is concurrent with phasing. This may be a principal advantage of the pseudo-MIR approach, because the anomalous-scatterer parameter refinement may be more reliable when incorporated into phasing than when done against $|^0F_A|$ estimates. Greater weight is given to the data set selected as 'native' in refinement of the 'heavy-atom' parameters in some implementations of the pseudo-MIR approach, although others treat data at all wavelengths equivalently (Terwilliger & Berendzen, 1997). The amplitudes $|^0F_A|$ are not a by-product of the pseudo-MIR approach.

14.2.1.7. *Determination of the anomalous-scatterer partial structure*

Determination of the partial structure of anomalous scatterers is a prerequisite for MAD-phased electron density, regardless of the phasing technique. As described above, the optimal quantities for solving and refining the partial structure of anomalous scatterers are the normal structure amplitudes $|^0F_A|$. Frequently $|^0F_A|$ values are not extracted from the MAD measurements, and the largest Bijvoet or dispersive differences are used instead. This involves the approximation of representing structure amplitudes ($|^0F_A|$) as the subset of larger differences ($||F^+| - |F^-||$ or $||F_{\lambda^1}| - |F_{\lambda^2}||$). The approximation is accurate for only a small fraction of reflections because there is little correlation between φ_A and φ_T. However, it suffices for a suitably strong signal and a suitably small number of sites. Patterson methods are quite successful in locating anomalous scatterers when the number of sites is small. However, the aim of MAD is to solve the macromolecule structure from one MAD data set using any number of anomalous scatterer sites. For larger numbers of sites, statistical direct methods may be employed.

The correct enantiomorph for the anomalous-scatterer partial structure also must be determined (φ_A *versus* $-\varphi_A$) in order to obtain an electron-density image of the macromolecule. However, it cannot be determined directly from MAD data. The correct enantiomorph is chosen by comparison of electron-density maps based on both enantiomorphs of the partial structure. Unlike the situation for pure MIR, the density based on the incorrect enantiomorph of the anomalous-scatterer partial structure is not the mirror image of that based on the correct enantiomorph and contains no image of the macromolecule. The correct map is distinguished by features such as a clear solvent boundary, positive correlation of redundant densities and a macromolecule-like density histogram. If the anomalous-scattering centres form a centric array, then the two enantiomorphs are identical and both maps are correct.

14.2.1.8. *General anomalous-scatterer labels for biological macromolecules*

MAD requires a suitable anomalous scatterer, of which none are generally present in naturally occurring proteins or nucleic acids. However, selenomethionine (SeMet) substituted for the natural amino acid methionine (Met) is a general anomalous-scattering label for proteins (Hendrickson, 1985), and is the anomalous scatterer most frequently used in MAD. The *K* edge of Se is the most accessible for MAD experiments ($\lambda = 0.98$ Å).

The SeMet label is especially general and convenient because it is introduced by biological substitution of SeMet for methionine. This is achieved by blocking methionine biosynthesis and substituting SeMet for Met in the growth medium of the cells in which the protein is produced. Production of SeMet protein in bacteria is generally straightforward (Hendrickson *et al.*, 1990; Doublié, 1997) and has also been accomplished in eukaryotic cells (Lustbader *et al.*, 1995; Bellizzi *et al.*, 1999).

Methionine is a particularly attractive target for anomalous-scatterer labelling because the side chain is usually buried in the hydrophobic core of globular proteins where it is relatively better ordered than are surface side chains. The labelling experiment provides direct evidence for isostructuralism of Met and SeMet proteins. All proteins in the biological expression system have SeMet substituted for Met at levels approaching 100%. The cells are viable, therefore the proteins are functional and isostructural with their unlabelled counterparts to the extent required by function.

The natural abundance of methionine in soluble proteins is approximately one in fifty amino acids, providing a typical MAD phasing signal of 4–6% of $|F|$ [equations (14.2.1.14) and (14.2.1.15)]. Typical extreme values for the anomalous scattering factors are $f'_{\min} \sim -10$ e and $f''_{\max} \sim 6$ e (Fig. 14.2.1.1). SeMet is more sensitive to oxidation than is Met, and care must be taken to maintain a homogeneous oxidation state. Generally, the reduced state is maintained by addition of disulfide reducing agents to SeMet protein and crystals. However, the oxidized forms of Se have sharper *K*-edge features and f' and f'' values of greater magnitude than does the reduced form (Smith & Thompson, 1998). This property has been exploited to enhance anomalous signals by intentional oxidation of SeMet protein (Sharff *et al.*, 2000). SeMet is also a useful isomorphous-replacement label with a signal of $\sim$10% of $|F|$. Prior knowledge of the sites of labelling is extremely useful during initial fitting of a protein sequence to electron density. Also, noncrystallographic symmetry operators can usually be defined more reliably from Se positions in SeMet protein than by heavy-atom positions in MIR due to the uniformity and completeness of labelling (Tesmer *et al.*, 1996).

An analogous general label is available for nucleic acids in the form of brominated bases, particularly 5-bromouridine, which is isostructural with thymidine. The *K* edge of Br corresponds to a wavelength of 0.92 Å, which is quite favourable for data collection.

14.2.2. Automated MAD and MIR structure solution

(T. C. TERWILLIGER AND J. BERENDZEN)

14.2.2.1. *Introduction*

In favourable cases, structure solution by X-ray crystallography using the MAD or MIR methods can be a straightforward, though often lengthy, process. The recently developed *Solve* software (Terwilliger & Berendzen, 1999*b*) is designed to fully automate this class of structure solution. The overall approach is to link together all the analysis steps that a crystallographer would normally carry out into a seamless procedure, and in the process to convert each decision-making step into an optimization problem.

In the case of both MAD and MIR data, a key element of the procedure is the scoring and ranking of possible solutions. This scoring procedure makes it possible to treat structure solution as an optimization procedure, rather than a decision-making one. In the case of MAD data, a second key element of the procedure is the conversion of MAD data to a pseudo-SIRAS form (Terwilliger, 1994*b*) that allows much more rapid analysis than one involving the full MAD data set.

14.2.2.2. *MAD and MIR structure solution*

The MAD and MIR approaches to structure solution are conceptually very similar and share several important steps. Two of these are the identification of possible locations of heavy or anomalously scattering atoms and an analysis of the quality of each of these potential heavy-atom solutions. In each method, trial partial structures for these heavy or anomalously scattering atoms are often obtained by inspection of difference Patterson functions or by semi-