automated analysis (*e.g.* Terwilliger *et al.*, 1987; Chang & Lewis, 1994; Vagin & Teplyakov, 1998). In other cases, direct-methods approaches have been used to find heavy-atom sites (Sheldrick, 1990; Miller *et al.*, 1994). Potential heavy-atom solutions found in any of these approaches are often just a starting point for structure solution, with additional sites found by difference Fourier or other approaches.

The analysis of the quality of potential heavy-atom solutions is also very similar in the MIR and MAD methods. In both cases a partial structure is used to calculate native phases for the entire structure, and the electron density that results is examined to see if the expected features of the macromolecule are found. Additionally, the agreement of the heavy-atom model with the difference Patterson function and the figure of merit of phasing are commonly used to evaluate the quality of a solution. In many cases, an analysis of heavy-atom sites by sequential deletion of individual sites or derivatives is often an important criterion of quality as well (Dickerson *et al.*, 1961).

14.2.2.3. *Decision making and structure solution*

The process of structure solution can be thought of largely as a decision-making process. In the early stages of solution, a crystallographer must choose which of several potential trial solutions may be worth pursuing. At a later stage, the crystallographer must choose which peaks in a heavy-atom difference Fourier are to be included in the heavy-atom model, and which hand of the solution is correct. At a final stage, the crystallographer must decide whether the solution process is complete and which of the possible heavy-atom models is the best. The most important feature of the *Solve* software is the use of a consistent scoring algorithm as the basis for making all these decisions.

14.2.2.4. *The need for rapid refinement and phasing during automated structure solution*

In order to make automated structure solution practical, it was necessary to be able to evaluate heavy-atom solutions very rapidly. This is because the automated approach used by *Solve* requires analysis of many heavy-atom solutions (typically 300–1000). For each heavy-atom solution examined, the heavy-atom sites have to be refined and phases calculated. In implementing automated structure solution, it was important to recognize the need for a trade-off between the most accurate heavy-atom refinement and phasing at all stages of structure solution and the time required to carry it out. The balance chosen for *Solve* was to use the most accurate available methods for final phase calculations, and to use approximate but much faster methods for all refinements and phase calculations. The refinement method chosen on this basis was origin-removed Patterson refinement (Terwilliger & Eisenberg, 1983), which treats each derivative in an MIR data set independently and which is very fast because it does not require phase calculation. The phasing approach used for MIR data thoughout *Solve* is Bayesian correlated phasing (Terwilliger & Berendzen, 1996; Terwilliger & Eisenberg, 1987), which takes into account the correlation of non-isomorphism among derivatives without substantially slowing down phase calculations.

For MAD data, Bayesian calculations of phase probabilities are very slow (*e.g.* Terwilliger & Berendzen, 1997; de La Fortelle & Bricogne, 1997). Consequently, we have used an alternative procedure for all MAD phase calculations except those done at the very final stage. This alternative is to convert the MAD data set into a form that is similar to one obtained in the single isomorphous replacement with anomalous scattering (SIRAS) method. In this way, a single data set with isomorphous and anomalous differences is obtained that can be used in heavy-atom refinement by the origin-removed Patterson refinement method and in phasing by conventional SIRAS phasing (Terwilliger & Eisenberg, 1987).

14.2.2.5. *Conversion of MAD data to a pseudo-SIRAS form*

The conversion of MAD data to a pseudo-SIRAS form that has almost the same information content requires two important assumptions. The first assumption is that the structure factor corresponding to anomalously scattering atoms in a structure varies in magnitude but not in phase at various X-ray wavelengths. This assumption will hold when there is one dominant type of anomalously scattering atom. The second is that the structure factor corresponding to anomalously scattering atoms is small compared to the structure factor from all other atoms. As long as these two assumptions hold, the information in a MAD experiment is largely contained in just three quantities: a structure factor (F_o) corresponding to the scattering from non-anomalously scattering atoms, a dispersive or isomorphous difference at a standard wavelength λ_o ($\Delta^{\rm ISO}_{\lambda_o}$), and an anomalous difference ($\Delta^{\rm ANO}_{\lambda_o}$) at the same standard wavelength (Terwilliger, 1994*b*). It is easy to see that these three quantities could be treated just like a SIRAS data set with the 'native' structure factor F_P replaced by F_o, the derivative structure factor F_{PH} replaced by $F_o + \Delta^{\rm ISO}_{\lambda_o}$, and the anomalous difference replaced by $\Delta^{\rm ANO}_{\lambda_o}$ (Terwilliger, 1994*b*). This is the approach taken by *Solve*. In this section, it is briefly shown how these three quantities can be estimated from MAD data.

For a particular reflection and a particular wavelength λ_j, we can write the total normal (*i.e.*, non-anomalous) scattering from a structure ($\mathbf{F}_{{\rm tot},\,\lambda_j}$) as the sum of two components. One is the scattering from all non-anomalously scattering atoms ($\mathbf{F}_o$). This scattering is wavelength-independent. The second is the normal scattering from anomalously scattering atoms ($\mathbf{F}_{H_{\lambda_j}}$) at wavelength λ_j. This term includes wavelength-dependent dispersive shifts in atomic scattering due to the f' term in the scattering factor, but not the anomalous part due to the f'' term. The magnitude of the total scattering factor can then be written in the form

$$F_{{\rm tot},\,\lambda_j} = |\mathbf{F}_o + \mathbf{F}_{H_{\lambda_j}}|. \qquad (14.2.2.1)$$

Here $\mathbf{F}_o$ and $\mathbf{F}_{{\rm tot},\,\lambda_j}$ can be thought of corresponding, respectively, to the native structure factor, F_P, and the derivative structure factor, F_{PH}, as used in the method of isomorphous replacement (Blundell & Johnson, 1976). If the scattering from anomalously scattering atoms is small compared to that from all other atoms, equation (14.2.2.1) can be rewritten in the approximate form

$$F_{{\rm tot},\,\lambda_j} \simeq F_o + F_{H_{\lambda_j}} \cos(\alpha), \qquad (14.2.2.2)$$

where α is the phase difference between the structure factors corresponding to non-anomalously and anomalously scattering atoms in the unit cell, $\mathbf{F}_o$ and $\mathbf{F}_{H_{\lambda_j}}$, respectively, at this X-ray wavelength.

The data in a MAD experiment consist of observations of structure-factor amplitudes for Bijvoet pairs, $F^+_{\lambda_j}$ and $F^-_{\lambda_j}$, for several X-ray wavelengths λ_j. These can be rewritten in terms of an average structure-factor amplitude $\overline{F}_{\lambda_j}$ and an anomalous difference $\Delta^{\rm ANO}_{\lambda_j}$ (*cf.* Blundell & Johnson, 1976). We would like to convert these into estimates of the amplitude of the structure factor corresponding to the non-anomalously scattering atoms alone, the amplitude of the structure factor corresponding to the entire structure at a standard wavelength, and the anomalous difference at the standard wavelength.

The normal scattering due to anomalously scattering atoms ($\mathbf{F}_{H_{\lambda_j}}$) changes in magnitude but not direction as a function of X-ray wavelength. We can therefore write (Terwilliger, 1994*b*)

references