where $w_i$ is the weight of observation $y_i$ and $f_i(\mathbf{x})$ is the calculated value of observation $i$ given the parameters $\mathbf{x}$. In essence, least-squares refinement poses the problem as 'Given these data, what are the parameters of the model that give the minimum variance of the observations?'. The $L_2$ norm is strongly affected by the largest deviations, which is not a desirable property in the early stages of refinement where the model may be seriously incomplete. In the early stages, it may be better to refine against the $L_1$ norm,

$$L_1 = \sum_i w_i |[y_i - f_i(\mathbf{x})]|,$$

the sum of the absolute value of the residuals. At present, this technique is not used in macromolecular crystallography.

The observable quantity in crystallography is the diffracted intensity of radiation. Fourier inversion of the model gives us a complex structure factor. The phase information is normally lost in the formulation of $f_i(\mathbf{x})$. This is a root cause of some of the problems of least-squares refinement from poor starting models. Many of the problems of least-squares refinement can be addressed by changing the measure of agreement from least squares to maximum likelihood, which evaluates to the likelihood of the observations given the model. In this formulation, the problem is posed as 'Given this model, what is the probability that the given set of data would be observed?'. The model is adjusted to maximize the probability of the given observations. This procedure is subtly different from least squares in that it is reasonably straightforward to account for incomplete models and errors in the model in computing the probability of the observations. Maximum-likelihood refinement is particularly useful for incomplete models because it produces residuals that are less biased by the current model than those produced by least squares. Maximum likelihood also provides a rigorous formulation for all forms of error in both the model and the observations, and allows incorporation of additional forms of prior knowledge (like additional phase information) into the probability distributions.

The likelihood of a model given a set of observations is the product of the probabilities of all of the observations given the model. If $P_a(\mathbf{F}_i; \mathbf{F}_{i,c})$ is the conditional probability distribution of the structure factor $\mathbf{F}_i$ given the model structure factor $\mathbf{F}_{i,c}$, then the likelihood of the model is

$$L = \prod_i P_a(\mathbf{F}_i; \mathbf{F}_{i,c}).$$

This is usually transformed into a more tractable form by taking the logarithm,

$$\log L = \sum_i \log P_a(\mathbf{F}_i; \mathbf{F}_{i,c}).$$

Since the logarithm increases monotonically with its argument, the two versions of the equation have maxima at the same values of the parameters of the model. This formulation is described in more detail in Chapter 18.2, in *IT* C Section 8.2.1 and by Bricogne (1997), Pannu & Read (1996), and Murshudov *et al.* (1997).

### 18.1.5. Optimization

Once the choice of criteria for agreement has been made, the next step is to adjust the parameters of the model to minimize the disagreement (or maximize the agreement) between the model and the data. The literature on optimization in numerical analysis and operations research, discussed in *IT* C Chapters 8.1–8.5, is very rich. The methods can be characterized by their use of gradient information (no gradients, first derivatives, or second derivatives), by their search strategy (none, downhill, random, annealed, or a combination of these), and by various performance measures on different classes of problems. These will be discussed more fully in Section 18.1.8.

### 18.1.6. Data

Resolution, accuracy, completeness and weighting of data all have an impact on the refinement process. Small-molecule crystals usually, but not always, diffract to well beyond atomic resolution. Macromolecular crystals do not generally diffract to atomic resolution. Macromolecular structures are by definition large, which in turn means that the unit cells are large and the number of diffracting unit cells per crystal is small when compared to small-molecule crystals of similar size. Fortunately, the situation can be partially offset with the use of the much more intense radiation generated by synchrotrons (Part 8) and by improved data-collection methods (Parts 7–11). Synchrotron-radiation sources designed to produce intense beams of X-rays for the study of materials are becoming much more readily available. As a consequence, both higher resolution and statistically better data can be obtained. Improvements in area-detector technology, protein purification, cryocrystallography and data-integration software beneficially influence the refinement process.

Refinement of crystal structures is a statistical process. There is no substitute for adequate amounts of accurate, correctly weighted data. Lower accuracy can be accommodated by increased amounts of data and correct weighting. Unfortunately, determining the correct weighting for macromolecular diffraction data is difficult. Maximum-likelihood methods are more robust than least-squares methods against improperly weighted data.

It has been clearly demonstrated that the best procedure for refining small molecules is to include all of the observations as integrated intensities, properly weighted, without preliminary symmetry averaging. Inclusion of weak data and refinement on diffracted intensity does not change the results very much, but has a strong effect on the precision of the parameter estimates derived from the refinement.

The long-standing debate as to whether refinement should be against structure-factor amplitudes or diffracted intensity has been resolved for small-molecule crystallography. Refinement against intensity is preferred because it is closer to the experimentally observed quantity, and the statistical weighting of the data is superior to that obtained for structure-factor amplitudes. If the model is correct and the data are reasonably good, the primary distinction between the two approaches is in the standard uncertainties of the derived parameters, which are usually somewhat better if the refinement is against diffracted intensity.

### 18.1.7. Models

Atomic resolution models are generally straightforward. A reasonably well phased diffraction pattern at atomic resolution shows the location of each atom. The primary problem (which can be substantial) is deciding how to model any disorder that may be present. Structural chemistry is derived from the model. Macromolecular models generally have most of the structural chemistry built in as part of the model. This approach is required as a direct consequence of having too little data at too limited a resolution to determine the positions of all of the atoms without using this additional information.

There are two procedures for building structural chemistry into a model. The first is to use known molecular geometry to reduce the number of variables. For example, if the distance between two atoms is held constant, the locus of possible positions for the second atom is the surface of a sphere centred on the first atom. This means that the position of the second atom can be specified given the

position of the first atom and two variables to locate the point on the sphere – a total of five variables instead of six. Every non-redundant constraint reduces the number of degrees of freedom in the model by one. If the second atom in this example were replaced by a group of atoms with known geometry (*e.g.* a phenyl group containing six atoms), the number of positional parameters could be reduced from 21 to eight. Constrained refinement is discussed extensively in *IT* C Chapter 8.3.

The second procedure is to treat the additional information as additional observations. A bond length is assumed to be an observation, based on other crystal structures, which has a mean value and a variance. This observation is added to the data instead of being used to reduce the number of parameters in the model.

The two approaches have different consequences on the ratio of observations to parameters. If we have $N_o$ observations, $N_p$ parameters and $N_r$ non-redundant geometric features to add to the problem, we have either $C = N_o/(N_p - N_r)$ and $(\mathrm{d}C/\mathrm{d}N_r) = C/(N_p - N_r)$ or $C = (N_o + N_r)/N_p$ and $(\mathrm{d}C/\mathrm{d}N_r) = 1/N_p$, where $C$ is the ratio of observations to parameters. The former are parameter *constraints* and the latter are parameter *restraints*. Constraints are more effective at increasing the ratio of observations to parameters, but since these features are built into the model, it is difficult to evaluate how appropriate they actually are for the problem at hand. Restraints provide an automatic evaluation of the appropriateness of the assumed geometry to the current data, because the deviations from the assumed values can be tested for statistical significance.

The most common constraints and restraints applied to macromolecular crystal structures are those which preserve or reinforce the molecular geometry of the amino acid or nucleotide residues (Chapter 18.3). Expected values for the geometry of these structural fragments are available from the small-molecule crystallographic literature and databases. A further step, which reduces the parameter count substantially, is to treat parts of the molecule as a set of linked rigid groups. This is particularly appropriate for aromatic fragments such as the side chains of phenylalanine, tyrosine, tryptophan and histidine, but can also be appropriate for small groups like valine and threonine. The extreme form of this approach is torsion-angle dynamics (Rice & Brünger, 1994), in which the only variables are torsion angles about bonds, and the position and orientation of the whole molecule. This description of the model works well with the right kind of optimization procedure.

Positional restraints can be parameterized in a variety of ways. For example, the geometry of three atoms can be treated as the three distances involved or as two distances and the angle between them. Several of the more popular restrained refinement programs treat the parameters for bond distances, bond angles and planarity as distances with a set of standard deviations. Others treat them as bond distances, bond angles and torsion angles weighted by the energy terms derived from experimental conditions. Different methods of parameterization and weighing have different effects on the refinement process, but to date these differences are not well characterized. The primary effects should be on the approach to convergence, as all of these formulations are normally satisfied by correct structures.

Additional criteria can be added to the model besides simple geometry. Preservation of bond lengths is usually done by adding terms

$$\sum_{\substack{\text{bonded}\\ \text{atoms}}} \left(1/\sigma_{ij}^2\right)\left(d_{ij} - d_{ij}^o\right)^2$$

to the objective function, where $d_{ij}$ is the distance between atoms $i$ and $j$, $d_{ij}^o$ is the ideal bond length, and $\sigma_{ij}$ is the weight applied to the bond. This is formally equivalent to treating bond stretching as a spring. Additional energy parameters can be added, such as electrostatic energy terms. Whatever vision of reality is applied to the objective function becomes part of the model.

The atomic displacement factors ($B$ factors) present a different set of problems from the coordinates. The behaviour of these parameters is strongly affected by coordinate errors, and in fact large atomic displacement parameters are frequently used to determine which parts of a structure are likely to contain errors. The $B$ factors are strongly related to the rate at which the diffraction pattern diminishes with resolution and thus cannot be accurately determined unless the diffraction pattern has been measured over a sufficiently wide range of resolution to determine this rate. As a practical matter, it is not feasible to refine individual atomic displacement parameters at resolution less than about 2 Å, and they frequently present problems even in atomic resolution small-molecule structures. In high-resolution small-molecule structures, $B$ factors are frequently represented as anisotropic ellipsoids described by six parameters per atom. In spite of the larger displacements found in macromolecules relative to small molecules, it is rarely possible to support the number of parameters required to refine a structure with independent anisotropic displacement factors. Nevertheless, the $B$ factors of the atoms are essential parts of the crystallographic model. Several methods for reducing the number of independent $B$ factors have been developed. The simplest is group $B$ factors, in which one parameter is refined for all atoms in a particular group of atoms. Another method is to apply a simple model to the change in displacement parameter within a group of atoms. In this treatment, a $B$ factor is refined for one atom, say the $C_\alpha$ atom of an amino-acid residue, and the remainder of the atoms in the residue are assigned displacement parameters that depend on their distance from the $C_\alpha$ atom (Konnert & Hendrickson, 1980). A third method is to enforce similarity of displacement parameters based on the correlation coefficients between pairs of displacement parameters in highly refined high-resolution structures (Tronrud, 1996).

Small-molecule refinement programs also apply restraints to the displacement parameters. The SIMU command of *SHELX* restrains the axes of the anisotropic displacement parameters of bonded atoms to be similar. This approach has been applied to a number of very high resolution macromolecular refinements.

Large $B$ factors do not represent large thermal motions of the atom but rather a distribution of positions occupied by the atom over time or in different unit cells of the crystal. The line between describing atoms with large $B$ factors as distributed about a single point or several points (disordered atoms) is sometimes blurred. At some point, the disorder can become resolved into alternative positions or the atoms disappear from the observable electron density. There are two kinds of disorder that can be easily modelled if data are available to sufficient resolution:

(1) *Static disorder* describes the situation in which portions of the structure have a small number of possible alternative conformations. The atoms in any given unit cell are in only one of the possible conformations, but different cells may have different conformations. Since the diffraction experiment averages the structure over all unit cells in the X-ray beam, the observations correspond to an average structure in which each conformation is weighted according to the fraction of the unit cells containing that conformation. The normal bond-length and angle restraints apply to each conformation, and the fractional occupancy of all conformations should sum to 1.0.

(2) *Dynamic disorder* describes the situation in which portions of the structure are not in fixed positions. This form of disorder is frequently encountered in amino-acid side chains on the molecular surface. The electrons are spread over a sufficiently large volume that the average electron density is very low and the atoms are essentially invisible to X-rays. In such cases, the best model is to simply omit the atoms from the diffraction calculation. They are

commonly placed in the model in plausible positions according to molecular geometry, but this can be misleading to people using the coordinate set. If the atoms are included in the model, the atomic displacement parameters generally become very large, and this may be an acceptable flag for dynamic disorder. The hazard with this procedure is that including these atoms in the model provides additional parameters to conceal any error signal in the data that might relate to problems elsewhere in the model.

At high resolution, it is sometimes possible to model the correlated motion of atoms in rigid groups by a single tensor that describes translation, libration and screw. This is rarely done for macromolecules at present, but may be an extremely accurate way to model the behaviour of the molecules. The recent development of efficient anisotropic refinement methods for macromolecules by Murshudov *et al.* (1999) will undoubtedly produce a great deal more information about the modelling of dynamic disorder and anisotropy in macromolecular structures.

Macromolecular crystals contain between 30 and 70% solvent, mostly amorphous. The diffraction is not accurately modelled unless this solvent is included (Tronrud, 1997). The bulk solvent is generally modelled as a continuum of electron density with a high atomic displacement parameter. The high displacement parameter blurs the edges, so that the contribution of the bulk solvent to the scattering is primarily at low resolution. Nevertheless, it is important to include this in the model for two reasons. First, unless the bulk solvent is modelled, the low-resolution structure factors cannot be used in the refinement. This has the unfortunate effect of rendering the refinement of *all* of the atomic displacement parameters ill-determined. Second, omission or inaccurate phasing of the low-resolution reflections tends to produce long-wavelength variations in the electron-density maps, rendering them more difficult to interpret. In some regions, the maps can become overconnected, and in others they can become fragmented.

## 18.1.8. Optimization methods

Optimization methods for small molecules are straightforward, but macromolecules present special problems due to their sheer size. The large number of parameters vastly increases the volume of the parameter space that must be searched for feasible solutions and also increases the storage requirements for the optimization process. The combination of a large number of parameters and a large number of observations means that the computations at each cycle of the optimization process are expensive.

Optimization methods can be roughly classified according to the order of derivative information used in the algorithm. Methods that use no derivatives find an optimum through a search strategy; examples are Monte Carlo methods and some forms of simulated annealing. First-order methods compute gradients, and hence can always move in a direction that should reduce the objective function. Second-order methods compute curvature, which allows them to predict not only which direction will reduce the objective function, but how that direction will change as the optimization proceeds. The zero-order methods are generally very slow in high-dimensional spaces because the volume that must be searched becomes huge. First-order methods can be fast and compact, but cannot determine whether or not the solution is a true minimum. Second-order methods can detect null subspaces and singularities in the solution, but the computational cost grows as the cube of the number of parameters (or worse), and the storage requirements grow as the square of the number of parameters – undesirable properties where the number of parameters is of the order of $10^4$.

Historically, the most successful optimization methods for macromolecular structures have been first-order methods. This is beginning to change as multi-gigabyte memories are becoming more common on computers and processor speeds are in the gigahertz range. At this time, there are no widely used refinement programs that run effectively on multiprocessor systems, although there are no theoretical barriers to writing such a program.

### 18.1.8.1. *Solving the refinement equations*

Methods for solving the refinement equations are described in *IT* C Chapters 8.1 to 8.5 and in many texts. Prince (1994) provides an excellent starting point. There are two commonly used approaches to finding the set of parameters that minimizes equation (18.1.4.1). The first is to treat each observation separately and rewrite each term of (18.1.4.1) as

$$w_i[y_i - f_i(\mathbf{x})] = w_i \sum_{j=1}^{N} \left(\frac{\partial f_i(\mathbf{x})}{\partial x_j}\right)(x_j^0 - x_j), \qquad (18.1.8.1)$$

where the summation is over the $N$ parameters of the model. This is simply the first-order expansion of $f_i(\mathbf{x})$ and expresses the hypothesis that the calculated values should match the observed values. The system of simultaneous *observational equations* can be solved for the parameter shifts provided that there are at least as many observations as there are parameters to be determined. When the number of observational equations exceeds the number of parameters, the least-squares solution is that which minimizes (18.1.4.1). This is the method generally used for refining small-molecule crystal structures, and increasingly for macromolecular structures at atomic resolution.

### 18.1.8.2. *Normal equations*

In matrix form, the observational equations are written as

$$\mathbf{A}\Delta = \mathbf{r},$$

where $\mathbf{A}$ is the $M$ by $N$ matrix of derivatives, $\Delta$ is the parameter shifts and $\mathbf{r}$ is the vector of residuals given on the left-hand sides of equation (18.1.8.1). The *normal equations* are formed by multiplying both sides of the equation by $\mathbf{A}^T$. This produces an $N$ by $N$ square system, the solution to which is the desired least-squares solution for the parameter shifts.

$$\mathbf{A}^T\mathbf{A}\Delta = \mathbf{A}^T\mathbf{r} \text{ or } \mathbf{M}\Delta = \mathbf{b},$$

$$m_{ij} = \sum_{k=1}^{M} w_k \left(\frac{\partial f_k(\mathbf{x})}{\partial x_i}\right)\left(\frac{\partial f_k(\mathbf{x})}{\partial x_j}\right),$$

$$b_i = \sum_{k=1}^{M} w_k[y_k - f_k(\mathbf{x})]\left(\frac{\partial f_k(\mathbf{x})}{\partial x_i}\right).$$

Similar equations are obtained by expanding (18.1.4.1) as a second-order Taylor series about the minimum $\mathbf{x}_0$ and differentiating.

$$\Phi(\mathbf{x}-\mathbf{x}_0) \approx \Phi(\mathbf{x}_0) + \left\langle \left(\frac{\partial\Phi}{\partial x_i}\right)_{\mathbf{x}_0} \middle| (\mathbf{x}-\mathbf{x}_0) \right\rangle + \frac{1}{2}\left\langle (\mathbf{x}-\mathbf{x}_0) \middle| \left(\frac{\partial^2\Phi}{\partial x_i \partial x_j}\right)_{\mathbf{x}_0} \middle| (\mathbf{x}-\mathbf{x}_0) \right\rangle,$$

$$\left|\left(\frac{\partial\Phi}{\partial\mathbf{x}}\right)\right\rangle \approx \left|\left(\frac{\partial^2\Phi}{\partial x_i \partial x_j}\right)_{\mathbf{x}_0}\right| (\mathbf{x}-\mathbf{x}_0)\right\rangle.$$

The second-order approximation is equivalent to assuming that the matrix of second derivatives does not change and hence can be computed at $\mathbf{x}$ instead of at $\mathbf{x}_0$.



references