Taking measurement errors into account requires multiplication of equation (18.2.3.3) with an appropriate probability distribution (usually a conditional Gaussian distribution with standard deviation σ_o) of the observed structure-factor amplitudes ($|\mathbf{F}_o|$) around the 'true' structure-factor amplitudes ($|\mathbf{F}|$),

$$P_{\text{meas}}(|\mathbf{F}_o|;\, |\mathbf{F}|). \tag{18.2.3.5}$$

Prior knowledge of the phases of the structure factors can be incorporated by multiplying equation (18.2.3.3) with a phase probability distribution

$$P_{\text{phase}}(\varphi) \tag{18.2.3.6}$$

and rewriting equation (18.2.3.3) in terms of the structure-factor moduli and amplitudes of $\mathbf{F} = |\mathbf{F}| \exp(i\varphi)$.

The unknown variables $|\mathbf{F}|$ and φ in equations (18.2.3.3)–(18.2.3.5) have to be eliminated by integration in order to obtain the conditional probability distribution of the observed structure-factor amplitudes, given a partial model with errors, the amplitude measurement errors and prior phase information:

$$P_a(|\mathbf{F}_o|;\, \mathbf{F}_c) = (1/\pi\varepsilon\sigma_\Delta^2)\int \mathrm{d}\varphi\, \mathrm{d}|\mathbf{F}|\, |\mathbf{F}| P_{\text{meas}}(|\mathbf{F}_o|;\, |\mathbf{F}|) \times P_{\text{phase}}(\varphi) \exp\left\{-[|\mathbf{F}| \exp(i\varphi) - D\mathbf{F}_o]^2/\varepsilon\sigma_\Delta^2\right\}. \tag{18.2.3.7}$$

The likelihood, $\mathcal{L}$, of the model is defined as the joint probability distribution of the structure factors of all reflections in the working set. Assuming independent and uncorrelated structure factors, $\mathcal{L}$ is simply the product of the distributions in equation (18.2.3.7) for all reflections. Instead of maximizing the likelihood, it is more common to minimize the negative logarithm of the likelihood,

$$E_{\text{X-ray}} = \mathcal{L} = -\sum_{hkl \in \text{working set}} \log[P_a(|\mathbf{F}_o|;\, \mathbf{F}_c)]. \tag{18.2.3.8}$$

Empirical estimates of σ_Δ [and D through equation (18.2.3.4)] can be obtained by minimizing $\mathcal{L}$ for a particular atomic model. It is generally assumed that σ_Δ and D show relatively little variation among neighbouring reflections. Accepting this assumption, σ_Δ and D can be estimated by considering narrow resolution shells of reflections and assuming that the two parameters are constant in these shells. Minimization of $\mathcal{L}$ can then be performed as a function of these constant shell parameters while keeping the atomic model fixed (Read, 1986, 1997). Alternatively, one can assume a two-term Gaussian model for σ_Δ (Murshudov *et al.*, 1997) and minimize $\mathcal{L}$ as a function of the Gaussian parameters. Note that individual atomic B factors are taken into account by the calculated model structure factors ($\mathbf{F}_c$).

This empirical approach to estimate σ_Δ and D requires occasional recomputation of these values as the model improves. Refinement methods that improve the model structure factors, $\mathbf{F}_c$, will therefore have a beneficial effect on σ_Δ and D. Better estimates of these values will then enhance the next refinement cycle. Thus, powerful optimization methods and maximum-likelihood targets are expected to interact in a synergistic fashion (*cf.* Fig. 18.2.5.1). Structure-factor averaging of multi-start refinement models can provide another layer of improvement by producing a better description of $\mathbf{F}_c$ if the model shows significant variability due to errors or intrinsic flexibility (see below).

In order to achieve an improvement over the least-squares residual [equation (18.2.3.2)], cross validation was found to be essential (Pannu & Read, 1996; Adams *et al.*, 1997) for the estimation of model incompleteness and errors (σ_Δ and D). Since the test set typically contains only 10% of the diffraction data, these cross-validated quantities can show significant statistical fluctuations as a function of resolution. In order to reduce these fluctuations, Read (1997) devised a smoothing method by applying restraints to σ_A values between neighbouring resolution shells where

$$\sigma_A = \left[1 - (\sigma_\Delta/\langle|\mathbf{F}_o|^2\rangle)\right]^{1/2}. \tag{18.2.3.9}$$

Pannu & Read (1996) have developed an efficient Gaussian approximation of equation (18.2.3.7) in cases of no prior phase information, termed the 'MLF' target function. In the limit of a perfect model (*i.e.* $\sigma_\Delta = 0$ and $D = 1$), MLF reduces to the traditional least-squares residual [equation (18.2.3.2)] with $1/\sigma_o^2$ weighting. In the case of prior phase information, the integration over the phase angles has been carried out numerically in equation (18.2.3.7), termed the 'MLHL' target (Pannu *et al.*, 1998). A maximum-likelihood function which expresses equation (18.2.3.7) in terms of observed intensities has also been developed, termed 'MLI' (Pannu & Read, 1996).

18.2.3.2. *A priori chemical information*

The parameters for the covalent terms in E_{chem} [equation (18.2.3.1)] can be derived from the average geometry and (r.m.s.) deviations observed in a small-molecule database. Extensive statistical analyses were undertaken for the chemical moieties of proteins (Engh & Huber, 1991) and polynucleotides (Parkinson *et al.*, 1996) using the Cambridge Structural Database (Allen *et al.*, 1983). Analysis of the ever-increasing number of atomic resolution macromolecular crystal structures will no doubt cause some modifications of these parameters in the future.

It is common to use a purely repulsive quartic function ($E_{\text{repulsive}}$) for the non-bonded interactions that are included in E_{chem} (Hendrickson, 1985):

$$E_{\text{repulsive}} = \sum_{ij}[(cR_{ij}^{\min})^n - R_{ij}^n]^m, \tag{18.2.3.10}$$

where R_{ij} is the distance between two atoms i and j, $R_{ij}^{\min}$ is the van der Waals radius for a particular atom pair ij, $c \leq 1$ is a constant that is sometimes used to reduce the radii, and $n = 2$, $m = 2$ or $n = 1$, $m = 4$. van der Waals attraction and electrostatic interactions are usually not included in crystallographic refinement. These simplifications are valid since the diffraction data contain information that is able to produce atomic conformations consistent with actual non-bonded interactions. In fact, atomic resolution crystal structures can be used to derive parameters for electrostatic charge distributions (Pearlman & Kim, 1990).

18.2.4. Searching conformational space

Annealing denotes a physical process wherein a solid is heated until all particles randomly arrange themselves in a liquid phase and is then cooled slowly so that all particles arrange themselves in the lowest energy state. By formally defining the target, E [equation (18.2.3.1)], to be the equivalent of the potential energy of the system, one can simulate such an annealing process (Kirkpatrick *et al.*, 1983). There is no guarantee that simulated annealing will find the global minimum (Laarhoven & Aarts, 1987). However, compared to conjugate-gradient minimization, where search directions must follow the gradient, simulated annealing achieves more optimal solutions by allowing motion against the gradient (Kirkpatrick *et al.*, 1983). The likelihood of *uphill* motion is determined by a control parameter referred to as *temperature*. The higher the temperature, the more likely it is that simulated annealing will overcome barriers (Fig. 18.2.4.1). It should be noted that the simulated-annealing temperature normally has no physical meaning and merely determines the likelihood of overcoming barriers of the target function in equation (18.2.3.1).

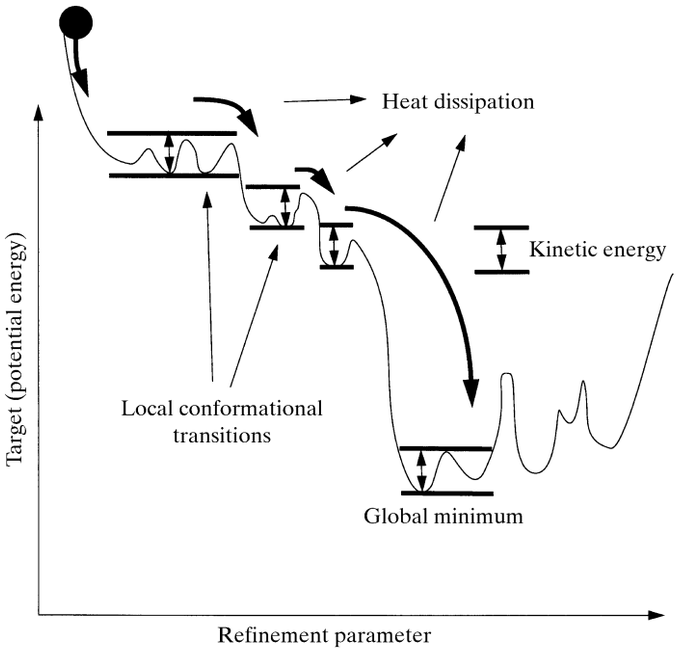

Fig. 18.2.4.1. Illustration of simulated annealing for minimization of a one-dimensional function. The kinetic energy of the system (a 'ball' rolling on the one-dimensional surface) allows local conformational transitions with barriers smaller than the kinetic energy. If a larger drop in energy is encountered, the excess kinetic energy is dissipated. It is thus unlikely that the system can climb out of the global minimum once it has reached it.

The simulated-annealing algorithm requires a mechanism to create a Boltzmann distribution at a given temperature, T, and an annealing schedule, that is, a sequence of temperatures $T_1 \geq T_2 \geq \ldots \geq T_l$ at which the Boltzmann distribution is computed. Implementations differ in the way they generate a transition, or move, from one set of parameters to another that is consistent with the Boltzmann distribution at a given temperature. The two most widely used methods are Metropolis Monte Carlo (Metropolis *et al.*, 1953) and molecular dynamics (Verlet, 1967) simulations. For X-ray crystallographic refinement, molecular dynamics has proven extremely successful (Brünger *et al.*, 1987) because it limits the search to physically reasonable 'moves'.

18.2.4.1. *Molecular dynamics*

A suitably chosen set of atomic parameters can be viewed as generalized coordinates that are propagated in time by the classical equations of motion (Goldstein, 1980). If the generalized coordinates represent the x, y, z positions of the atoms of a molecule, the classical equations of motion reduce to the familiar Newton's second law:

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = -\nabla_i E. \qquad (18.2.4.1)$$

The quantities m_i and $\mathbf{r}_i$ are, respectively, the mass and coordinates of atom i, and E is given by equation (18.2.3.1). The solution of the partial differential equations (18.2.4.1) can be achieved numerically using finite-difference methods (Verlet, 1967; Abramowitz & Stegun, 1968). This approach is referred to as molecular dynamics.

Initial velocities for the integration of equation (18.2.4.1) are usually assigned randomly from a Maxwell distribution at the appropriate temperature. Assignment of different initial velocities will generally produce a somewhat different structure after simulated annealing. By performing several refinements with different initial velocities, one can therefore improve the chances of success of simulated-annealing refinement. Furthermore, this improved sampling can be used to study discrete disorder and conformational variability, especially when using torsion-angle molecular dynamics (see below).

Although Cartesian (*i.e.* flexible bond lengths and bond angles) molecular dynamics places restraints on bond lengths and bond angles [through $E_{\rm chem}$, equation (18.2.3.1)], one might want to implement these restrictions as constraints, *i.e.*, fixed bond lengths and bond angles (Diamond, 1971). This is supported by the observation that the deviations from ideal bond lengths and bond angles are usually small in macromolecular X-ray crystal structures. Indeed, fixed-length constraints have been applied to crystallographic refinement by least-squares minimization (Diamond, 1971). It is only recently, however, that efficient and robust algorithms have become available for molecular dynamics in torsion-angle space (Bae & Haug, 1987, 1988; Jain *et al.*, 1993; Rice & Brünger, 1994). We chose an approach that retains the Cartesian-coordinate formulation of the target function and its derivatives with respect to atomic coordinates, so that the calculation remains relatively straightforward and can be applied to any macromolecule or their complexes (Rice & Brünger, 1994). In this formulation, the expression for the acceleration becomes a function of positions and velocities. Iterative equations of motion for constrained dynamics in this formulation can be derived and solved by finite-difference methods (Abramowitz & Stegun, 1968). This method is numerically very robust and has a significantly increased radius of convergence in crystallographic refinement compared to Cartesian molecular dynamics (Rice & Brünger, 1994).

18.2.4.2. *Temperature control*

Simulated annealing requires the control of the temperature during molecular dynamics. The current temperature of the simulation ($T_{\rm curr}$) is computed from the kinetic energy

$$E_{\rm kin} = \sum_i^n \tfrac{1}{2} m_i \left(\frac{\partial r_i}{\partial t}\right)^2 \qquad (18.2.4.2)$$

of the molecular-dynamics simulation,

$$T_{\rm curr} = 2E_{\rm kin}/3nk_B. \qquad (18.2.4.3)$$

Here, n is the number of atoms, m_i is the mass of the atom and k_B is Boltzmann's constant. One commonly used approach to control the temperature of the simulation consists of coupling the equations of motion to a heat bath through a 'friction' term (Berendsen *et al.*, 1984). Another approach is to rescale periodically the velocities in order to match $T_{\rm curr}$ with the target temperature.

18.2.4.3. *Annealing schedules*

The simulated-annealing temperature needs to be high enough to allow conformational transitions, but not so high that the model moves too far away from the correct structure. The optimal temperature for a given starting structure is a matter of trial and error. Starting temperatures that work for the average case have been determined for a variety of simulated-annealing protocols (Brünger, 1988; Adams *et al.*, 1997). However, it might be worth trying a different temperature if a particularly difficult refinement problem is encountered. In particular, significantly higher temperatures are attainable using torsion-angle molecular dynamics. Note that each simulated-annealing refinement is subject to 'chance' by using a random-number generator to generate the initial velocities. Thus, multiple simulated annealing runs can be carried out in order to increase the success rate of the refinement. The best structure(s) (as determined by the free R value) among a set of refinements using

different initial velocities and/or temperatures can be taken for further refinement or structure-factor averaging (see below).

The annealing schedule can, in principle, be any function of the simulation step (or 'time' domain). The two most commonly used protocols are linear slow-cooling or constant-temperature followed by quenching. A slight advantage is obtained with slow cooling (Brünger *et al.*, 1990). The duration of the annealing schedule is another parameter. Too short a protocol does not allow sufficient sampling of conformational space. Too long a protocol may waste computer time, since it is more efficient to run multiple trials than one long refinement protocol (unpublished results).

18.2.4.4. *An intuitive explanation of simulated annealing*

The goal of any optimization problem is to find the global minimum of a target function. In the case of crystallographic refinement, one searches for the conformation or conformations of the molecule that best fit the diffraction data and that simultaneously maintain reasonable covalent and non-covalent interactions. Simulated-annealing refinement has a much larger radius of convergence than conjugate-gradient minimization (see below). It must, therefore, be able to find a lower minimum of the target E [equation (18.2.3.1)] than the local minimum found by simply moving along the negative gradient of E.

It is most easy to visualize this property of simulated annealing in the case of a one-dimensional problem, where the goal is to find the global minimum of a function with multiple minima (Fig. 18.2.4.1). An intuitive way to understand a molecular-dynamics simulation is to envisage a ball rolling on this one-dimensional surface. When the ball is far from the global minimum, it gains a certain momentum which allows it to cross barriers of the target function [equation (18.2.4.3)]. Slow-cooling temperature control ensures that the ball will eventually reach the global minimum rather than just bouncing across the surface. The initial temperature must be large enough to overcome smaller barriers, but low enough to ensure that the system will not escape the global minimum if it manages to arrive there.

While temperature itself is a global parameter of the system, temperature fluctuations arise principally from local conformational transitions, for example, from an amino-acid side chain falling into the correct orientation. These local changes tend to lower the value of the target E, thus increasing the kinetic energy, and hence the temperature, of the system. Once the temperature control has removed this excess kinetic energy through 'heat dissipation', the reverse transition is very unlikely, since it would require a localized increase in kinetic energy where the conformational change occurred in the first place (Fig. 18.2.4.1). Temperature control maintains a sufficient amount of kinetic energy to allow local conformational corrections, but does not supply enough to allow escape from the global minimum. This explains the observation that, on average, the agreement with the diffraction data will improve, rather than worsen, with simulated annealing.

18.2.5. Examples

Many examples have shown that simulated-annealing refinement starting from initial models (obtained by standard crystallographic techniques) produces significantly better final models compared to those produced by least-squares or conjugate-gradient minimization (Brünger *et al.*, 1987; Brünger, 1988; Fujinaga *et al.*, 1989; Kuriyan *et al.*, 1989; Rice & Brünger, 1994; Adams *et al.*, 1997). In another realistic test case (Adams *et al.*, 1999), a series of models for the aspartic proteinase penicillopepsin were generated from homologous structures present in the Protein Data Bank. The sequence identity among these structures ranged from 100% to 25%, thus providing a set of models with increasing coordinate error compared to the refined structure of penicillopepsin. These models, after truncation of all residues to alanine, were all used as search models in molecular replacement against the native penicillopepsin diffraction data. In all cases, the correct placement of the model in the penicillopepsin unit cell was found.

Both conjugate-gradient minimization and simulated annealing were carried out in order to compare the performance of the $E^{\rm LSQ}$ least-squares residual [equation (18.2.3.2)], MLF (the maximum-likelihood target using amplitudes) and MLHL (the maximum-likelihood target using amplitudes and experimental phase information). In the latter case, phases from single isomorphous replacement (SIR) were used. A very large number of conjugate-gradient cycles were carried out in order to make the computational requirements equivalent for both minimization and simulated annealing. The conjugate-gradient minimizations were converged, *i.e.* there was no change when further cycles were carried out.

For a given target function, simulated annealing always outperformed minimization (Fig. 18.2.5.1). For a given starting model, the maximum-likelihood targets outperformed the least-squares-residual target for both minimization and simulated annealing, producing models with lower phase errors and higher map correlation coefficients when compared with the published penicillopepsin crystal structure (Fig. 18.2.5.1). This improvement is illustrated in σ_A-weighted electron-density maps obtained from the resulting models (Fig. 18.2.5.2). The incorporation of experimental phase information further improved the refinement significantly despite the ambiguity in the SIR phase probability distributions. Thus, the most efficient refinement will make use of simulated annealing and phase information in the MLHL maximum-likelihood target function.

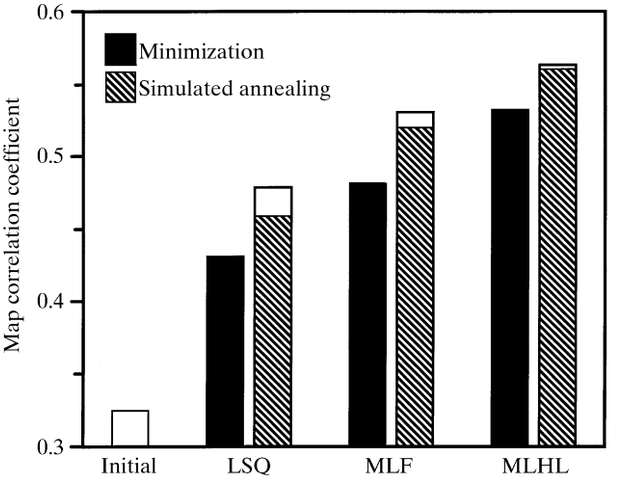

Fig. 18.2.5.1. Simulated annealing produces better models than extensive conjugate-gradient minimization. Map correlation coefficients were computed before and after refinement against the native penicillopepsin diffraction data (Hsu *et al.*, 1977) for the polyalanine model derived from Rhizopuspepsin (Suguna *et al.*, 1987, PDB code 2APR). Correlation coefficients are between σ_A-weighted maps calculated from each model and from the published penicillopepsin structure. The observed penicillopepsin diffraction data were in space group $C2$ with cell dimensions $a = 97.37$, $b = 46.64$, $c = 65.47$ Å and $\beta = 115.4^\circ$. All refinements were carried out using diffraction data from the lowest-resolution limit of 22.0 Å up to 2.0 Å. The MLHL refinements used single isomorphous phases from a $K_3UO_2F_5$ derivative of the penicillopepsin crystal structure, which covered a resolution range of 22.0 Å to 2.8 Å. Simulated-annealing refinements were repeated five times with different initial velocities. The numerical averages of the map correlation coefficients for the five refinements are shown as hashed bars. The best map correlation coefficients from simulated annealing are shown as white bars.

references