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protocol may waste computer time, since it is more efficient to run multiple trials than one long refinement protocol (unpublished results).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.

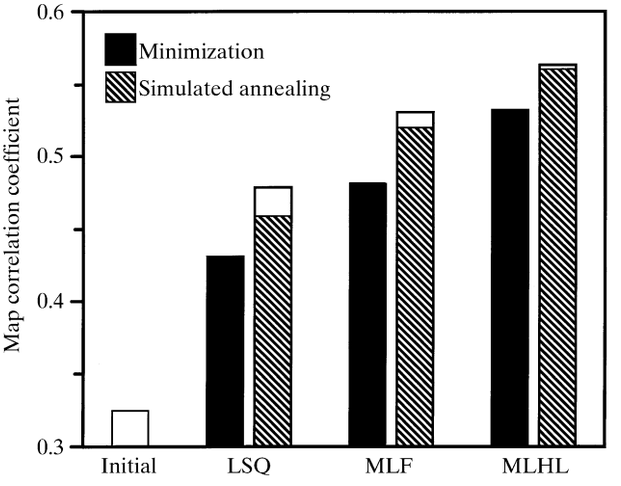

Fig. 18.2.5.1. Simulated annealing produces better models than extensive conjugate-gradient minimization. Map correlation coefficients were computed before and after refinement against the native penicillopepsin diffraction data (Hsu *et al.*, 1977) for the polyalanine model derived from Rhizopuspepsin (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 $\beta = 115.4°$. 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.

Cross validation is essential in the calculation of the maximum-likelihood target (Kleywegt & Brünger, 1996; Pannu & Read, 1996; Adams *et al.*, 1997). Maximum-likelihood refinement without cross validation gives much poorer results, as indicated by higher free R values, $R_{\rm free} - R$ differences and phase errors (Adams *et al.*, 1997). It should be noted that the final normal R value is in general increased, compared to refinements with the least-squares target, when using the cross-validated maximum-likelihood formulation. This is a consequence of the reduction of overfitting by this method.

18.2.6. Multi-start refinement and structure-factor averaging

Multiple simulated-annealing refinements starting from the same model, termed 'multi-start' refinement, will generally produce somewhat different structures. Even well refined structures will show some variation consistent with the estimated coordinate error of the model (*cf.* results for 1.8 Å resolution in Fig. 18.2.2.1). More importantly, the poorer the model, the more variation is observed (Brünger, 1988). Some of the models resulting from multi-start refinement may be better than others, for example, as judged by the free R value. Thus, if computer time is available, multi-start refinement has several advantages. A more optimal single model than that produced by a single simulated-annealing calculation can usually be obtained. Furthermore, each separate model coming from a multi-start refinement fits the data slightly differently. This could be the result of intrinsic flexibility in the molecule (see below) or the result of model-building error. Regions in the starting model that contain significant errors often show increased variability after multi-start refinement, and a visual inspection of the ensemble of models produced can be helpful in identifying these incorrectly modelled regions.

To better identify the correct conformation, structure factors from each of the models can be averaged (Rice *et al.*, 1998). This averaging tends to reduce the effect of local errors (noise) that are presumably different in each member of the family. The average structure factor can produce phases that contain less model bias than phases computed from a single model. It should also produce better estimates of σ_Δ and D for maximum-likelihood targets and for σ_A-weighted electron-density maps, because $\mathbf{F}_c$ is used in the computation of these parameters [equation (18.2.3.7)]. Because it is inherently a noise-reducing technique, multi-start refinement followed by structure-factor averaging should be most useful in situations where there is significant noise, namely when the data-to-parameter ratio is low (*e.g.* if only moderate-resolution diffraction data are available).

18.2.7. Ensemble models

In cases of conformational variability or discrete disorder, there is not one single correct solution to the global minimization of equation (18.2.3.1). Rather, the X-ray diffraction data represent a spatial and temporal average over all conformations that are assumed by the molecule. Ensembles of structures, which are simultaneously refined against the observed data, may thus be a more appropriate description of the diffraction data. This has been used for some time when alternate conformations are modelled locally. Alternate conformations can be generalized to global conformations (Gros *et al.*, 1990; Kuriyan *et al.*, 1991; Burling & Brünger, 1994), *i.e.*, the model is duplicated n-fold, the calculated structure factors corresponding to each copy of the model are summed, and this composite structure factor is refined against the observed X-ray diffraction data. Each member of the family is chemically 'invisible' to all other members. The optimal number, n, can be determined by cross validation (Burling & Brünger, 1994; Burling *et al.*, 1996).

An advantage of a multi-conformer model is that it directly incorporates many possible types of disorder and motion (global disorder, local side-chain disorder, local wagging and rocking motions). Furthermore, it can be used to detect automatically the most variable regions of the molecule by inspecting the atomic r.m.s. difference around the mean as a function of residue number. Thermal factors of single-conformer models may sometimes be misleading because they underestimate the degree of motion or disorder (Kuriyan *et al.*, 1986), and, thus, the multiple-conformer model can be a more faithful representation of the diffraction data. A disadvantage of the multi-conformer model is that it introduces many more parameters in the refinement.

Although there are some similarities between averaging structure factors of individually refined structures and performing multi-conformer refinement, there are also fundamental differences. For example, multi-start averaging seeks to improve the calculated electron-density map by averaging out the noise present in the individual models, each of which is still a good representation of the diffraction data. This method is most useful at the early stages of refinement when the model still contains errors. In contrast, multi-conformer refinement seeks to create an ensemble of structures at the final stages of refinement which, taken together, best represent the data. It should be noted that each individual conformer of the ensemble does not necessarily remain a good description of the diffraction data, since the whole ensemble is refined against the data. Clearly, multi-conformer refinement requires a high observable-to-parameter ratio.

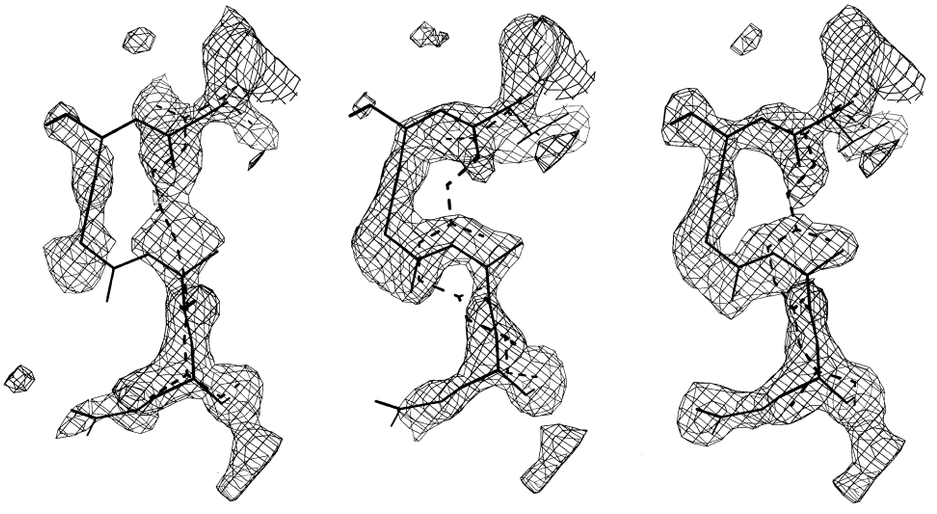

Fig. 18.2.5.2. Maximum-likelihood targets significantly decrease model bias in simulated-annealing refinement. σ_A-weighted electron-density maps contoured at 1.25σ for models from simulated-annealing refinement with different targets are shown. Residues 233 to 237 are shown for the published penicillopepsin crystal structure (Hsu *et al.*, 1977) as solid lines, and for the model with the lowest free R value from five independent refinements as dashed lines.

references