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.

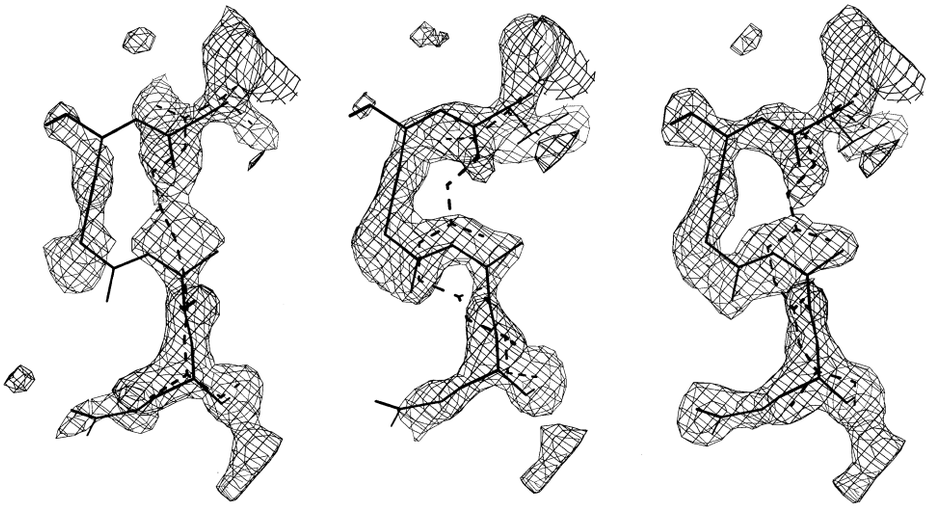

Fig. 18.2.5.2. Maximum-likelihood targets significantly decrease model bias in simulated-annealing refinement. σ_A-weighted electron-density maps contoured at 1.25σ for models from simulated-annealing refinement with different targets are shown. Residues 233 to 237 are shown for the published penicillopepsin 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