the individual models. A weight, W_{wARP}, is assigned to each structure factor on the basis of the variance of the two-dimensional distribution of the individual structure factors around the mean. The mean value of W_{wARP} over all reflections and the R factor after averaging can be used to judge the progress of the averaging procedure.

25.2.5.2.4. *Ab initio solution of metalloproteins*

If the coordinates of one or a few heavy atoms are known, initial phases can be calculated. The problem of solving the structure of such a metalloprotein from the sites of the metal alone can be considered in the same framework as for heavy-atom-replacement solutions. Maps calculated from the phases of heavy atoms alone often have the best defined features within a defined radius of the heavy atom(s). Thus protocols that do not place all atoms at the start but instead perform a slow building while extending the model in a growing sphere around the heavy atom are preferred. When such a model is essentially complete, it can be used for automated tracing and completion of the model.

25.2.5.2.5. *Solvent building*

In this application, the protein (or nucleic acid) model is not rebuilt during refinement, and only the solvent structure is continuously updated, allowing the construction of a solvent model without iterative manual map inspection.

25.2.5.3. *Applicability and requirements*

Density-based atom selection for the whole structure is only possible if the X-ray data extend to a resolution where atomic positions can be estimated from the Fourier syntheses with sufficient accuracy for them to refine to the correct position. If the structural model is of reasonable quality, at 2.5 Å or better, at least a part of the solvent structure or a small missing or badly placed part of the protein can be located. This provides indirect improvement of the whole structure. For automated model rebuilding, or for refining poor molecular-replacement solutions, higher resolution is essential. The general requirement is that the number of X-ray reflections should be at least six to eight times higher than the number of atoms in the model, which roughly corresponds to a resolution of 2.3 Å for a crystal with 50% solvent. However, the method can work at lower resolution or fail with a higher one, depending less on the quality of the initial phases and more on the internal quality of the data and on the inherent disorder of the molecule.

The X-ray data should be complete. If strong low-resolution data (*e.g.* 4 to 10 Å) are systematically missing, *e.g.* due to detector saturation, the electron density even for good models is often discontinuous. Because *ARP* involves updating on the basis of density maps, such discontinuity will lead to incorrect interpretation of the density and slow convergence or even uninterpretable output.

25.2.5.4. *An example*

The structure of chitinase A from *Serratia marcescens* (Perrakis *et al.*, 1994) was initially solved by multiple isomorphous replacement with anomalous signal (MIRAS), with only a single derivative contributing to resolution higher than 5.0 Å. The MIRAS map (2.5 Å) was solvent-flattened. Model building was not straightforward and much time was spent in tracing the protein chain.

As an experiment, the solvent-flattened map was used to initiate building of free-atom models, using least-squares minimization against the native 2.3 Å data combined with *ARP*. This resulted in crystallographic R factors ranging between 20.1 and 22.4%. Each *ARP* model gave phases marginally worse than those available by solvent flattening alone, due to the limited resolution of the native data. However, the *wARP* averaging procedure resulted in a reduction of 11.2° in the weighted mean phase error. The map correlation coefficient between the final map and the *wARP* map was 81.2%, better by 12.8% compared with the solvent-flattened map.

The *wARP* model with the lowest R factor was used to initiate model building. In the initial tracing, 75 residues were identified, belonging to more than 20 different main-chain fragments. After autobuilding, ten cycles of restrained *ARP* were run according to the standard protocol. One *REFMAC* cycle of conjugate-gradient minimization was executed to optimize a maximum-likelihood residual and bulk solvent scaling. σ_A-weighted maps were calculated and *ARP* was used to update the model. All atoms (main-chain, side-chain and free atoms) were allowed to be removed and new atoms were added where appropriate. After ten iterations, a new building cycle was invoked. After every 'big' cycle, a more complete model was obtained. This 'big' cycle was iterated 20 times. Finally, 515 residues were traced in nine chains, all of which were docked unambiguously into the sequence. This is the lowest-resolution application to date. 2.3 Å was the real resolution limit of the data measured from these crystals; however, the high solvent content (61%) provided on average seven observations per atom and an almost complete trace was easily accomplished.

25.2.6. *PROCHECK*: validation of protein-structure coordinates (R. A. Laskowski, M. W. MacArthur and J. M. Thornton)

25.2.6.1. *Introduction*

As in all scientific measurements, the parameters that result from a macromolecular structure determination by X-ray crystallography (*e.g.* atomic coordinates and B factors) will have associated uncertainties. These arise not only from systematic and random errors in the experimental data but also in the interpretation of those data. Currently, the uncertainties cannot easily be estimated for macromolecular structures due to the computer- and memory-intensive nature of the calculations required (Tickle *et al.*, 1998). Thus, more indirect methods are necessary to assess the reliability of different parts of the model, as well as the reliability of the model as a whole. Among these methods are those which rely on checking only the stereochemical and geometrical properties of the model itself, without reference to the experimental data (MacArthur *et al.*, 1994; Laskowski *et al.*, 1998). Here we describe *PROCHECK* (Laskowski *et al.*, 1993), which is one of these structure-validation methods.

The *PROCHECK* program computes a number of stereochemical parameters for the given protein model and compares them with 'ideal' values obtained from a database of well refined high-resolution protein structures in the Protein Data Bank (PDB; Bernstein *et al.*, 1977). The results of these checks are output in easy-to-understand coloured plots in PostScript format (Adobe Systems Inc., 1985). Significant deviations from the derived standards of normality are highlighted as being 'unusual'.

The program's primary use is during the refinement of a protein structure; the highlighted regions can direct the crystallographer to parts of the structure that may have problems and which may need attention. It should be noted that outliers may just be outliers; they are not necessarily errors. Unusual features may have a reasonable explanation, such as distortions due to ligand binding in the protein's active site. However, if there are many oddities throughout the model, this could signify that there is something wrong with it as a whole. Conversely, if a model has good stereochemistry, this alone is not proof that it is a good model of the protein structure.

references