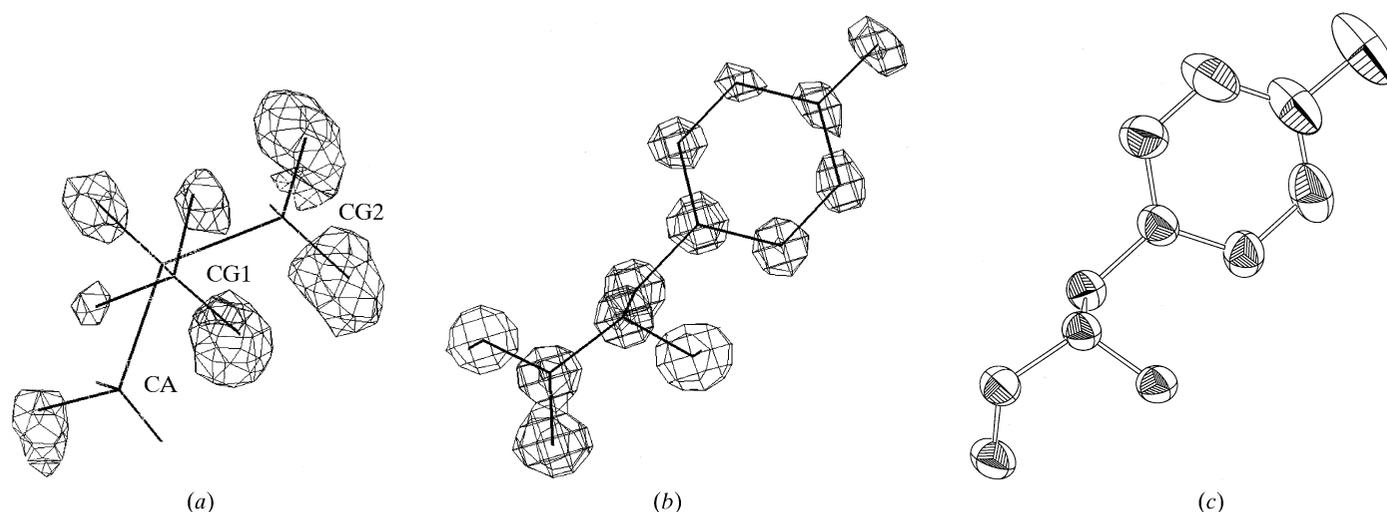


## 18.4. REFINEMENT AT ATOMIC RESOLUTION

**Figure 18.4.5.1**

(a), (b) Representative electron-density maps for the refinement of *Clostridium acidurici* ferredoxin at 0.94 Å resolution (Dauter, Wilson *et al.*, 1997). (a) The density for hydrogen atoms (at  $3\sigma$ ) omitted from the structure-factor calculation for Val42. (b) The  $(2F_o - F_c)$  density for Tyr30, contoured at  $3\sigma$ . (c) The thermal ellipsoids corresponding to (b), drawn at the 33% probability level using ORTEPII (Johnson, 1976). There is a clear correlation between the density in (b) and the ellipsoids in (c), showing increased displacement towards the end of the side chain, particularly in the plane of the phenyl ring.

prove difficult to refine the occupancy satisfactorily with statistical certainty.

#### 18.4.4.4. Validation of extra parameters during the refinement process

The introduction of additional parameters into the model always results in a reduction in the least-squares or maximum-likelihood residual – in crystallographic terms, the  $R$  factor. However, the statistical significance of this reduction is not always clear, since this simultaneously reduces the observation-to-parameter ratio. It is therefore important to validate the significance of the introduction of further parameters into the model on a statistical basis.

Brünger (1992a) introduced the concept of statistical cross validation to evaluate the significance of introducing extra features into the atomic model. For this, a small and randomly distributed subset of the experimental observations is excluded from the refinement procedure, and the residual against this subset of reflections is termed  $R_{\text{free}}$ . It is generally sufficient to include about 1000 reflections in the  $R_{\text{free}}$  subset; further increase in this number provides little, if any, statistical advantage but diminishes the power of the minimization procedure. For atomic resolution structures, cross validation is important in establishing whether the introduction of an additional type of feature to the model (with its associated increase in parameters) is justified. There are two limitations to this. Firstly, if  $R_{\text{free}}$  shows zero or minimal decrease compared to that in the  $R$  factor, the significance remains unclear. Secondly, the introduction of individual features, for example the partial occupancy of five water molecules, can provide only a very small change in  $R_{\text{free}}$ , which will be impossible to substantiate. To recapitulate, at atomic resolution the prime use of cross validation is in establishing protocols with regard to extended sets of parameter types. The sets thus defined will depend on the quality of the data.

In the final analysis, validation of individual features depends on the electron density, and Fourier maps must be judiciously inspected. Nevertheless, this remains a somewhat subjective approach and is in practice intractable for extensive sets of parameters, such as the occupancies and ADPs of all solvent sites.

For the latter, automated procedures, which are being developed at present, are an absolute necessity, but they may not be optimal in the final stages of structure analysis, and visual inspection of the model and density is often needed.

The problems of limited data and reparameterization of the model remain. At high resolution, reparameterization means having the same number of atoms, but changing the number of parameters to increase their statistical significance, for example switching from an anisotropic to an isotropic atomic model or *vice versa*. In contrast, when reparameterization is applied at low resolution, this usually involves constraints, *i.e.*, a reduction in the number of independent atoms, but this is not an ideal procedure, as real chemical entities of the model are sacrificed. Reducing the number of independent atoms will inevitably result in disagreement between the experiment and model, which in turn will affect the precision of other parameters. It would be more appropriate to reduce the number of parameters without sacrificing the number of atoms, for example by describing the model in torsion-angle space. Water poses a particular problem, as at low as well as at high resolution the water molecules cannot all be described as discrete atoms. Algorithms are needed to describe them as a continuous model with only a few parameters. In the simplest model, the solvent can be described as a constant electron density.

### 18.4.5. Features in the refined model

All features of the refined model are more accurately defined if the data extend to higher resolution (Fig. 18.4.5.1). In this section, those features that are especially enhanced in an atomic resolution analysis are described. Introduction of an additional feature to the model should be assessed by the use of cross- or self-validation tools: only then can the significance of the parameters added to the model be substantiated.

#### 18.4.5.1. Hydrogen atoms

Hydrogen atoms possess only a single electron and therefore have low electron density and are relatively poorly defined in X-ray studies. They play central roles in the function of proteins,