structure by applying the appropriate scale to the solvent contribution:

$$F_{\rm total} = F_{\rm prot} + k_m \exp(-B_m s^2/4) F_{\rm mask}, \qquad (18.4.5.2)$$

where k_m and B_m are the scale and temperature factor for the solvent contribution, and $F_{\rm total}$, $F_{\rm prot}$ and $F_{\rm mask}$ are the complex structure factors corresponding to the combined contribution from the protein and solvent region. The scale and B values are usually refined iteratively to find the best match to the observed structure factors, and are implemented in *CNS*, *REFMAC* and *PHENIX.REFINE*.

Nevertheless, there remain severe problems in the modelling of the interface. The border between the two regions is not abrupt, as there is a smooth and continuous change from the region with fully occupied, discrete sites to one which is truly fluid, but this passes through a volume with an increasing level of dynamic disorder and associated partial occupancy. Modelling of this region poses major problems, as described above, and the definition of disordered sites with low occupancy remains difficult even at atomic resolution. At which stage the occupancy and associated ADP can be defined with confidence is not yet an objective decision. In addition, refinement and modelling at this level of detail is very time consuming in terms of human intervention.

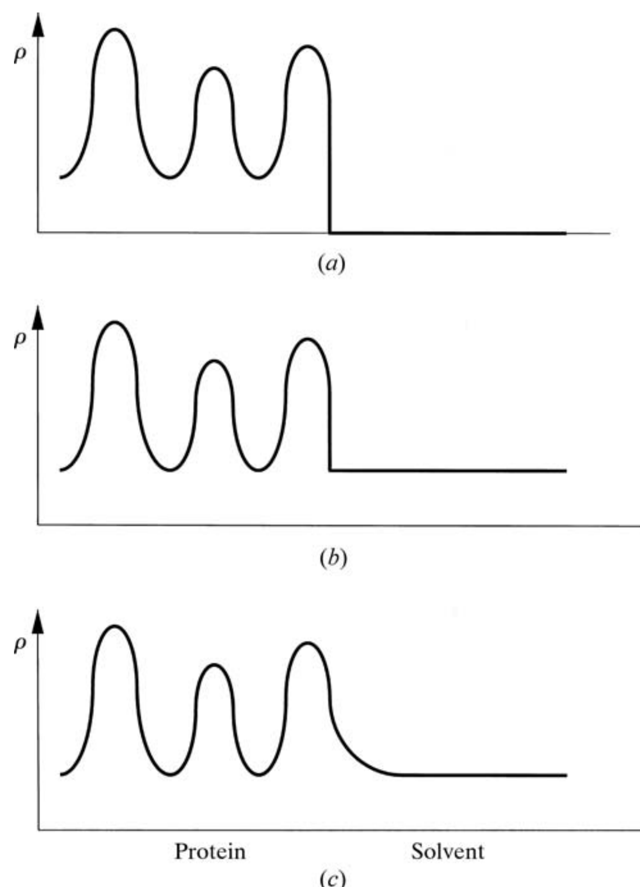

Figure 18.4.5.2
Schematic representation of the bulk-solvent models described in the text. (*a*) No bulk-solvent correction, *i.e.* solvent density set to zero. (*b*) Constant level of solvent outside the macromolecule and ordered water envelope. Here, sharp edge effects remain. (*c*) The model as in (*b*), but smoothed at the edge of a macromolecule, equivalent to the application of a B value to the solvent model.

18.4.5.7. Metal ions and other ligands in the solvent

In general, proteins are crystallized from aqueous solutions which contain various additives, such as anions or cations (especially metals), organic solvents, including those used as cryoprotectants, and other ligands. Some of these may bind in specific or indeed non-specific sites in the ordered solvent shell, in addition to any functional binding sites of the protein. To identify such entities at limited resolution is often impossible, as the range of expected ADPs is large and there is very poor discrimination in the appearance of such sites and of water in the electron density. Atomic resolution assists in resolving ambiguities, as the interatomic distances, ADPs and occupancies are all better defined.

For metal ions, two additional criteria can be invoked. Firstly, the coordination geometry, with well defined bond lengths and angles, provides an indication of the identity of the ion, as different metals have different preferred ligand environments (Harding, 1999, 2006). The bond-valence approach is also applicable (Müller *et al.*, 2003). In addition, the value of the refined ADP and/or occupancy is helpful. Secondly, the anomalous signal in the data should reveal the presence of metal and some other non-water sites in the solvent through computation of the anomalous difference synthesis (Dauter & Dauter, 1999). This emphasizes the need to retain the anomalous signal during the collection and reduction of native data. While these approaches can be applied at lower resolution, they both become much more powerful at atomic resolution.

The presence of bound organic ligands has become especially relevant since the advent of cryogenic freezing. Compounds such as ethylene glycol and glycerol possess a number of functional hydrogen-bonding groups that can attach to sites on the protein in a defined way. Indeed, these may often bind in the active sites of enzymes such as glycosyl hydrolases, where they mimic the hydroxyl groups of the sugar substrate. It is most important to identify such moieties properly, particularly if substrate studies are to be planned successfully.

18.4.5.8. Deformation density

X-ray structures are generally modelled using the spherical-atom approximation for the scattering, which ignores the deviation from sphericity of the outer bonding and lone-pair electrons. Extensive studies over a long period have confirmed that the so-called deformation density, representing deviation from this spherical model, can be determined experimentally using data to very high resolution, usually from 0.8 to 0.5 Å. An excellent review of this field has been provided by Coppens (1997). The observed deviations can be compared with those expected from the available theories of chemical bonding and the densities derived therefrom.

The application of atomic resolution analysis to proteins allowed the observation of the deformation density in macromolecules (Lamzin *et al.*, 1999). Data for two proteins were analysed: crambin (molecular weight 6 kDa) at 0.67 Å resolution and a subtilisin (molecular weight 30 kDa) at 0.9 Å resolution. Significant and interpretable deformation density could not be observed for the individual residues. However, on averaging the density over 40 peptide units for crambin and more than 250 for the subtilisin, the deformation density within the peptide unit was clearly visible and could be related to the expected bonding features in these units. This shows the real power of atomic resolution crystallography, which can reveal features containing no more than 0.2 e Å^{-3}.

Deformation density studies are now being applied to many polypeptides (Jelsch *et al.*, 1998; Koritsanszky *et al.*, 2002; Pichon-

Copyright © 2012 International Union of Crystallography references