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.

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 all the interatomic distances, ADPs and occupancies are 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 [see, for example, Nayal & Di Cera (1996)]. 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 by computation of the anomalous difference synthesis (Dauter & Dauter, 1999). 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 recent review of this field is 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. Such studies have been applied to peptides and related molecules (Souhassou *et al.*, 1992; Jelsch *et al.*, 1998).

The application of atomic resolution analysis to proteins has allowed the first steps towards 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. 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 $Å^{-3}$.

18.4.6. Quality assessment of the model

The refinement of proteins at resolution lower than atomic depends upon the use of restraints on the geometry and ADPs. Most target libraries for refinement and validation of structures (*e.g.* Engh & Huber, 1991) are derived from either the Cambridge Structural Database (Allen *et al.*, 1979) or from protein structures in the Protein Data Bank (PDB; Bernstein *et al.*, 1977). The availability of atomic resolution structures provides more objective data for the construction of target libraries. Stereochemical parameters, such as conformational angles φ, ψ, should ideally not be restrained, as they allow independent validation of the model. Analysis of eight structures determined at atomic resolution (EU 3-D Validation Network, 1998) indicates that they follow the expected rules of chemistry more closely than those of lower-resolution analyses in the PDB, confirming that atomic resolution indeed provides more precise coordinates.

18.4.7. Relation to biological chemistry

A question arises as to what biological issues are addressed by analysis of macromolecular structures at atomic resolution. For any protein, the overall structure of its fold, and hence its homology with other proteins, can already be provided by analyses at low to medium resolution. However, proteins are the active entities of cells and carry out recognition of other macromolecules, ligand binding and catalytic roles that depend upon subtle details of chemistry, for which accurate positioning of the atoms is required. Even at atomic resolution, the accuracy of structural definition is less than what would ideally be required for the changes observed during a chemical reaction. At lower resolutions, structure–function relations require yet further extrapolation of the experimental data.

To understand the function of many macromolecules, such as enzymes, it is not sufficient to determine the structure of a single state. Alongside the native structure, those of various complexes will also be required. The differences between the states provide additional information on the functionality. For an understanding of the chemistry involved, atomic resolution has tremendous advantages in terms of accuracy, as reliable judgments can be based on the experimental data alone.

Advantages of atomic resolution include the following:

(1) The positions of all atoms that possess defined conformations are more accurately defined. This means that all bond lengths and angles in the structure have lower standard uncertainties (EU 3-D Validation Network, 1998). For regions of the molecule where the conformation is representative this is of purely quantitative significance, but where the stereochemistry deviates from the expected value this accuracy takes on a special significance, which poses questions to the theoretical chemist. Such deviations from standard geometry often play an important role in biological function.

(2) The better the ADP definition, notably its anisotropy, the greater the insight into the static or thermal flexibility of individual regions of the molecule. Macromolecules are crucially dependent upon flexibility for properties, such as induced fit in substrate or ligand recognition, allosteric responses or responses to the biological environment. More detailed definition of the position and mobility of flexible regions may be assisted by atomic resolution analysis.

(3) A few amino-acid side chains play an active role in catalysis (those that do include histidine, aspartic and glutamic acids and serine) throughout protonation–deprotonation events, and hydrogen

references