

18. REFINEMENT

but at the traditional resolution limits of macromolecular structure analyses their positions can only be inferred rather than defined from the experimental data. Indeed, even at a resolution of 2.5 Å, hydrogen atoms should be included in the refined model, as their exclusion biases the position of the heavier atoms, but with their 'riding' positions fixed by those of the parent atoms.

As for small structures, independent refinement of hydrogen-atom positions is not always warranted, even by atomic resolution data, and hydrogen atoms are rather attached as riding rigidly on the positions of the parent atoms. Nevertheless, atomic resolution data allow the experimental confirmation of the positions of many of the hydrogen atoms in the electron-density maps, as they do account for one-sixth of the diffracting power of a carbon atom. Inspection of the maps can in principle allow the identification of (1) the presence or absence of hydrogen atoms on key residues, such as histidine, aspartate and glutamate or on ligands, and (2) the correct location of hydrogen atoms where more than one position is possible, such as in the hydroxyl groups of serine, threonine or tyrosine.

The correct placement of hydrogen atoms riding on their parent atoms involves computation of the appropriate position after each cycle of refinement. This is done automatically by programs such as *SHELXL*, *REFMAC* or *PHENIX.REFINE*. For rigid groups such as the NH amide, aromatic rings, $-\text{CH}_2-$ or $=\text{CH}-$, the position is accurately defined by the bonding scheme. For groups such as methyl CH_3 or OH, the position is not absolutely defined, and the software is required to make judgmental decisions. For example, *SHELXL* offers the opportunity to inspect the maximum density on a circular Fourier synthesis for optimal positioning. The bond length is fixed according to results from a small-molecule database. The location of hydrogen atoms on polar atoms can be assisted by software that analyses the local hydrogen-bonding networks; this involves maximization of the hydrogen-bonding potential of the relevant groups. Sheldrick advocates assigning ADP values to riding hydrogen atoms of 1.2 times that of the parent atom, or 1.5 times in the case of methyl and similar entities.

18.4.5.2. Anisotropic atomic displacement parameters

Refinement of an isotropic model involves four independent parameters per atom, three positional and one isotropic ADP. In contrast, an anisotropic model requires nine parameters per atom, with the anisotropic atomic displacement described by an ellipsoid represented by six parameters. At 1 Å resolution, the data certainly justify an anisotropic atomic model. Extension of the model from isotropic to anisotropic should generally result in a reduction in the *R* factor of the order of 5–6% and a comparable drop in R_{free} . As a consequence of the diminution of the observable-to-parameter ratio, the *R* factor at all resolutions will drop by a similar amount; however, R_{free} will not. Experience shows that at 2 Å or less there is no drop in R_{free} , and an anisotropic model is totally unsupported by the data. At intermediate resolutions, the result depends on the data quality and completeness. At lower resolution, to account for anisotropy of the atoms, the overall motion of molecules or domains can be refined using translation/libration/screw (TLS) parameters (Schomaker & Trueblood, 1968).

Until the end of the 1990s, anisotropic ADPs had only been handled by programs originally developed for small-molecule analysis, which use conventional algebraic computa-

tions of the calculated structure-factor amplitudes, *SHELXL* being a prime example. A limitation of this approach is the substantial computation time required. The use of fast-Fourier-transform algorithms for the structure-factor calculation leads to a significant saving in time (Murshudov *et al.*, 1999). Anisotropic modelling of the individual ADPs is essential if the thermal vibration is to be analysed in terms of coordinated motion of the whole molecule or of domains (Schomaker & Trueblood, 1968). Painter & Merritt (2006) have provided a means of analysing refined models to suggest appropriate TLS groupings.

18.4.5.3. Alternative conformations

Proteins are not rigid units with a single allowed conformation. *In vivo* they spontaneously fold from a linear sequence of amino acids to provide a three-dimensional phenotype that may exhibit substantial flexibility, which can play a central role in biological function, for example in the induced fit of an enzyme by a substrate or in allosteric conformational changes. Flexibility is reflected in the nature of the protein crystals, in particular the presence of regions of disordered solvent between neighbouring macromolecules in the lattice (see Section 18.4.5.6).

The structure tends to be highly ordered at the core of the protein, or more properly, at the core of the individual domains. Atoms in these regions in the most ordered protein crystals have ADP values comparable to those of small molecules, reflecting the fact that they are, in essence, closely packed by surrounding protein. In general, as one moves towards the surface of the protein, the situation becomes increasingly fluid. Side chains and even limited stretches of the main chain may show two (or multiple) conformations. These may be significant for the biological function of the protein.

The ability to model the alternative conformations is highly resolution dependent. At atomic resolution, the occupancy of two alternative but well defined conformations can be refined to an accuracy of about 5%, thus second conformations can be seen, provided that their occupancy is about 10% or higher. The limited number of proteins for which atomic resolution structures are available suggest that up to 20% of the 'ordered residues' show multiple conformations. This confers even further complexity on the description of the protein model. A constraint can be imposed on residues with multiple conformations: namely that the sum of all the alternatives must be 1.0. Protein regions (whether they are side- or main-chain regions) with alternative conformations and partial occupancy can form clusters in the unit cell with complementary occupancy. This often coincides with alternative sets of solvent sites, which should also be refined with complementary occupancies.

The atoms in two alternative conformations occupy independent and discrete sites in the lattice, about which each vibrates. However, if the spacing between two sites is small and the vibration of each is large, then it becomes impossible to differentiate a single site with high anisotropy from two separate sites. There is no absolute rule for such cases: programs such as *SHELXL* place an upper limit on the anisotropy and then suggest splitting the atom over two sites. Some regions can show even higher levels of disorder, with no electron density being visible for their constituent atoms. Such fully disordered regions do not contribute to the diffraction at high resolution, and the definition of their location will not be improved with atomic resolution data.