Pesme *et al.*, 2004; Afonine *et al.*, 2007; Zarychta *et al.*, 2007). It has been observed that details of deformation densities are most clearly revealed when only the highest resolution (so-called high order) terms are included in the refinement and the Fourier maps (Coppens, 1997). This is reported to result from the deconvolution of the effects of the anisotropic ADPs, which can to some extent take up the fine features corresponding to the deformation density (bonding electrons and lone pairs). After proper modelling of the deformation density features, the overall model is refined against all the data.

18.4.6. Quality assessment of the model

The refinement of proteins at resolutions lower than atomic depends upon the use of restraints on the geometry and ADPs. The almost exclusively used library of target geometric restraints for refinement and validation of protein structures (Chapter 18.3) is derived from structures of amino acids and peptides in the Cambridge Crystallographic Data Centre's small-molecule crystal structure database (Allen *et al.*, 1979). Stereochemical parameters, such as conformational angles φ, ψ, should ideally not be restrained, as they allow independent validation of the model. As stated in Section 18.4.4.2, these restraints are required even at atomic resolution to maintain the chemical integrity of flexible regions, although their impact will be limited on ordered regions.

Owing to the excess of accurate X-ray observations over parameters at atomic resolution, extensive validation of individual structures should be less challenging than for those at lower resolution. It is hard to achieve an R factor around 10% with an incorrect model. However, considerable attention needs to be given to detail and great care taken to avoid over-interpretation, especially of the flexible regions.

An analysis of eight structures determined at atomic resolution some years ago (EU 3-D Validation Network, 1998) indicated 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. A subsequent analysis of ten structures at ultra-high resolution, 0.8 Å or better (Jaskolski *et al.*, 2007) confirmed these conclusions but identified a few possible adjustments to some targets. Following this analysis, Karplus *et al.* (2008) proposed that protein stereochemistry is context dependent, *e.g.* it differs in detail between α-helices and β-strands, and that this should be reflected in future target libraries. Thus, the availability of atomic resolution structures will provide a more objective basis for the construction of such libraries.

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 of the norm, 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, all through protonation–deprotonation events, and hydrogen atoms are crucial to their function. Hydrogen atoms are usually treated as riding on their parent atoms and should be included in the model, even at medium resolution. Unfortunately, those hydrogen atoms that are of interest can only rarely be treated as rigidly bonded at a predictable position. However, atomic resolution allows many hydrogen atoms to be clearly identified in the refined electron density. In addition, the presence or absence of hydrogen may be inferred by accurate estimation of the bond lengths between atoms, *e.g.* within the carboxylate groups.

(4) The relative orientation of reacting moieties is crucial to enzyme catalysis. If chemical hypotheses of mechanism are to be subjected to appropriate Popperian scrutiny (Popper, 1959), then precise definition of atomic coordinates in native and complex structures is necessary.

(5) Enzyme catalysis provides a reduction of the activation energy of the reaction, which can be achieved by distortion of the conformation of the substrate bound to the enzyme (the so-called Michaelis complex) towards the transition state or by the stabilization of the latter by the enzyme. For both, the study of complexes of inhibitors or substrate analogues at a sufficient resolution to clarify the fine detail of the structures is required.

(6) Adaptation of the enzyme to the substrate is postulated by the induced-fit theory of catalysis. The level of adjustment can be very small, and energy calculations again require that this be precisely defined.

(7) In metalloproteins, the ligand field, and hence geometry and bond lengths, around the metal ion are essential indicators of any variation in valence electrons between different states. For example, bond lengths between oxidized and reduced states of metal ions vary by the order of 0.1 Å or less, and

clear distinction between alternative oxidation states requires an accuracy only provided by atomic resolution.

Almost all atomic resolution analyses require data recorded from cryogenically frozen crystals. This does pose some problems of biological relevance, as proteins *in vivo* have adapted to operate at ambient cellular temperatures. The structure that is required is that of the protein and surrounding solvent at the corresponding temperature. The trade-off is that cryogenic structures may be better defined, but that they are only so because of the increased order of the protein and solvent at low temperature. This has to be weighed against the lack of fine detail in a medium-resolution analysis at room temperature.

A question often raised with regard to the worth of atomic resolution data concerns the effort required in refining a protein at such resolution. To define all details, such as alternative conformations, hydrogen-atom positions and solvent, is certainly time consuming, especially if an anisotropic model is adopted. However, the advantages outweigh the disadvantages, as even if a full anisotropic model is not refined to exhaustion, nevertheless all density maps will be clearer if the resolution is better, resulting in an improved definition of the features of interest.

18.4.8. Practical strategies

Pioneering work was carried out by Teeter and colleagues on crambin, using data recorded on this small and highly stable protein using a conventional diffractometer (Teeter *et al.*, 1993). Some of the earliest atomic resolution structures using data from a synchrotron source with an imaging plate detector included rubredoxin at 1.0 Å (Dauter *et al.*, 1992), ribonuclease Sa at 1.1 Å (Sevcik *et al.*, 1996) and triclinic lysozyme at 0.9 Å resolution (Walsh *et al.*, 1998), at room temperature for the rubredoxin and ribonuclease, and 100 K for the lysozyme. The strategy used involved the application of conventional restrained least-squares or maximum-likelihood techniques in the early stages of refinement, followed by switching over to *SHELXL* to introduce a full anisotropic model and riding hydrogen atoms.

Subsequent structures have almost exclusively used cryogenically vitrified crystals to minimize the effects of radiation damage on high-intensity SR beamlines. A review of the field (Dauter, 2003) reported over 100 structures at better than 1.2 Å spacing, and in January 2011 there were around 1450 such coordinate sets in the PDB.

Initially, only *SHELXL* was able to refine macromolecular structures with anisotropic ADPs, riding hydrogen atoms and appropriate stereochemical restraints. Subsequently, such options have been implemented in programs such as *REFMAC* and *PHENIX.REFINE*. The use of fast Fourier algorithms gives these programs some speed advantage over *SHELXL*, but this is now of less importance given the power of contemporary computers. The key advantage of *SHELXL* is the ability to perform full- (or at least block-) matrix inversion, and hence extract error estimates for each individual parameter.

It is not straightforward to provide hard and fast rules for the optimum strategy. The power of refinement at atomic resolution lies in the sheer number of X-ray observations which drive the minimization to a global minimum. The dangers of bias in the final model and of false minima can largely be ignored. As a guideline the following steps would seem to be a common-sense approach when data to 1.2 Å or better are available.

(1) Refine an initial isotropic model at modest resolution with one of the conventional programs to ensure the correct global minimum has been reached. This step was probably more appropriate in the days when computing resources were more limiting, and could probably be omitted today.
(2) Introduce an anisotropic model with riding hydrogen atoms using one of the maximum-likelihood programs such as *REFMAC*, *PHENIX.REFINE* or *SHELXL* (with the conjugate-gradient option), and refine against the experimental data, omitting around 1000–2000 reflections for validation by $R_{\rm free}$. It is *not* necessary to omit 5% of the data, particularly at this resolution.
(3) Complete the water model using one of the autobuild options within or associated with (*e.g. ARP/wARP* or *COOT*) the refinement program. This is an area amenable to considerable automation in the future to allow partially occupied sites and overlapping networks to be defined without user intervention.
(4) Introduce alternative conformations for the protein. Much of this is also amenable to automation, especially for the side chains, where programs such as *COOT* can already suggest likely alternate rotamers.
(5) Inspect appropriately weighted electron-density maps for features to be added, deleted or moved. This includes ligands, metal atoms, ions, cryprotectants and other additives. Current developments in programs such as *COOT* are increasingly automating this step. Software libraries that provide accurate descriptions of the ligand chemistry and geometry are essential, and are the subject of intense development. It is vital that structures are deposited in the PDB with the correct geometrical restraints. To some extent, steps (3)–(5) go in parallel, and indeed their order is not vital.
(6) Consider the *SHELXL* full-matrix option for the final refinement cycles. This will provide estimates of the accuracy of the individual parameters, not available from other programs.
(7) As a final step, run some cycles with all reflections including those previously used for $R_{\rm free}$, but without changing the model parameters. At all stages, it will remain necessary to impose restraints to avoid poor behaviour of any flexible regions. As the resolution goes beyond 1 Å, the effect of such restraints on good parts of the structure becomes minimal.

In summary, at atomic resolution models require special attention to detail, reflecting the wealth of information with regard to features such as alternative conformations of both side and main chains, and extensive ordered water structure often with partial occupancy and overlapping networks. With the present software, this can be a lengthy operation, and indeed this can persuade some scientists to work at lower, less informative, resolution to speed up and simplify the analysis. Many of the presently tedious manual tasks in terms of model construction are, however, amenable to extensive automation. As the number of atomic resolution structures increases, we can expect the refinement and modelling process to be greatly simplified with new software algorithms.

References

Afonine, P. V., Grosse-Kunstleve, R. W., Adams, P. D., Lunin, V. Y. & Urzhumtsev, A. (2007). *On macromolecular refinement at subatomic resolution with interatomic scatterers*. *Acta Cryst.* D**63**, 1194–1197.
Agarwal, R. C. (1978). *A new least-squares refinement technique based on the fast Fourier transform algorithm*. *Acta Cryst.* A**34**, 791–809.
Allen, F. H., Bellard, S., Brice, M. D., Cartwright, B. A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B. G., Kennard, O.,