Main (1998), and secondly, information about effects such as pseudosymmetry could be exploited.

Another argument for the use of intensities rather than amplitudes is relevant to least squares, where the derivative for amplitude-based refinement with respect to F_{calc} is singular when F_{calc} is equal to zero (Schwarzenbach *et al.*, 1995). This is not the case for intensity-based least squares. In applying maximum likelihood, this problem does not arise (Pannu & Read, 1996; Murshudov *et al.*, 1997).

Finally, while there may be some advantages in refining against intensities, Fourier syntheses always require structure-factor amplitudes.

18.4.4.2. Restraints on coordinates and ADPs

For a good small-molecule crystal the experimental X-ray data extend to ~0.8 Å spacing and the structure can be refined against the X-ray data alone. The resulting accuracy of the atomic coordinates will generally be better than 0.01 Å. However, even for small-molecule structures, disordered regions require the imposition of stereochemical restraints (or constraints) if the chemical integrity is to be preserved and the ADPs are to be realistic.

The typical situation for protein crystals is quite different, with atomic resolution being the exception rather than the rule. Thus, for proteins the geometry of the atomic model needs to be restrained, both in terms of geometry and ADPs. The geometric target values have been established from a set of amino-acid and small-peptide structures (see Chapter 18.3 by Engh & Huber), for which the bond-length r.m.s.d. is about 0.02 Å. In the present context, we restrict the discussion to bond lengths, but this is representative of the other restraints. Clearly, the relative contribution of the X-ray data and the restraints on the final parameters varies as a function of resolution. The restraints dominate at low resolution, while by the time 0.8 Å spacing is achieved, the restraints will be essentially irrelevant for well ordered regions. The imposition of bond-length restraints with target deviations of ~0.02 Å means that the distribution of bond lengths in the final model will be of the same order, independent of the resolution of the X-ray data. However, this must not be taken to imply that the accuracy of the atomic parameters is invariant with the overall resolution and, more importantly, the atomic displacement. To be explicit, the accuracy of the atomic positions decreases (1) as the resolution becomes worse (*i.e.*, the number of X-ray observations decreases) and (2) as the ADPs become larger. While this should be obvious to the practicing crystallographer, it may not be so apparent to the less expert user of the PDB (Wlodawer *et al.*, 2008).

Jaskolski *et al.* (2007) analysed ten structures from the PDB refined at ultra-high (better than 0.8 Å) resolution to investigate appropriate geometrical restraints. They confirmed the general correctness of values in the Engh & Huber dictionary and showed that the mean observed deviations in these ten structures from the target bond lengths were indeed roughly 0.02 Å. They therefore postulated that 0.02 Å was an appropriate value to use in applying stereochemical restraints to protein structures in general. There has been some dispute about this value (Tickle, 2007) but we believe it to be appropriate. A more detailed analysis of this issue has been performed, suggesting that target values differ depending on the structural context (Karplus *et al.*, 2008); this may lead to some fine adjustments in the target dictionary, as predicted earlier (EU 3-D Validation Network, 1998).

In analogy to the geometrical restraints based on the Engh & Huber dictionary, anisotropic ADP restraints were established in the *SHELX* program suite (Sheldrick, 2008). For example, they prevent atoms from becoming unrealistically anisotropic and restrain the shapes of the ellipsoids of bonded atoms to be not too dissimilar. Riding hydrogen atoms are assigned constrained isotropic ADPs based on those of the parent atoms.

A more theoretical justification for use of restraints is that refinement can be considered as Bayesian estimation. From this point of view, all available and usable prior knowledge should be exploited, as it should not harm the parameter estimation during refinement. Bayesian estimation shows asymptotic behaviour (Box & Tiao, 1973), *i.e.*, when the number of observations becomes large, the experimental data override the prior knowledge. In this sense, the purpose of the experiment is to enhance our knowledge about the molecule, and the procedure should be cumulative, *i.e.*, the result of the old experiment should serve as prior knowledge for the design and treatment of new experiments (Box & Tiao, 1973; Stuart *et al.*, 1999; O'Hagan, 1994). However, there are problems in using restraints. For example, the probability distribution reflecting the degree of belief in the restraints is not good enough. Use of a Gaussian approximation to distributions of distances, angles and other geometric properties has not been justified. Firstly, the distribution of geometric parameters depends strongly on ADPs. Secondly, different geometric parameters are correlated. Thirdly, many geometric parameters (*e.g.* bond angles, torsion angles) are dependent on the conformation, configuration and environment of the molecule in question (Karplus *et al.*, 2008; Gelbin *et al.*, 1996). This problem should be the subject of further investigation.

In summary, the atomic resolution structures to date confirm that a mean deviation of bond lengths from target values of 0.02 Å (and comparable values for other restraint types) is appropriate, but may be subject to minor adjustments. These levels of restraint should be applied at all resolutions: the stereochemistry should be neither over- nor under-restrained.

18.4.4.3. Partial occupancy

It may be necessary to refine one additional parameter, the occupancy factor of an atomic site, for structures possessing regions that are spatially or temporally disordered, with some atoms lying in more than one discrete site. The sum of the occupancies for alternative individual sites of a protein atom must be 1.0.

For macromolecules, the occupancy factor is important in several situations, including the following:

(1) when a protein or ligand atom is present in all molecules in the lattice, but can lie in more than one position due to alternative conformations;
(2) for the solvent region, where there may be overlapping and mutually exclusive solvent networks;
(3) when ligand-binding sites are only partially occupied due to weak binding constants, and the structures represent a mixture of native enzyme with associated solvent and the complex structure;
(4) when there is a mixture of protein residues in the crystal, due to inhomogeneity of the sample arising from polymorphism, a mixture of mutant and wild-type protein, or other causes.

Unfortunately, the occupancy parameter is highly correlated with the ADP, and it is difficult to model these two parameters at resolutions less than atomic. Even at atomic resolution, it can