To assess the 'geometric strain' in a model on a per-residue basis, the refinement program *X-PLOR* (Brünger, 1992*b*) can produce geometric pseudo-energy plots. In such a plot, the ratio of $E_{\rm geom}(i)$/r.m.s.$(E_{\rm geom})$ is calculated as a function of the residue number i. The pseudo-energy term $E_{\rm geom}$ consists of the sums of the geometric and stereochemical pseudo-energy terms of the force field ($E_{\rm geom} = E_{\rm bonds} + E_{\rm angles} + E_{\rm dihedrals} + E_{\rm impropers}$), involving only the atoms of each residue.

It has been observed that the more high-resolution protein structures become available, the more 'well behaved' proteins turn out to be, *i.e.* the distributions of conformational torsion angles and torsion-angle combinations become even tighter than observed previously and the numerical averages tend to shift somewhat (Ponder & Richards, 1987; Kleywegt & Jones, 1998; EU 3-D Validation Network, 1998; MacArthur & Thornton, 1999; Walther & Cohen, 1999).

21.1.7.2.3. *C^α-only models*

Validation of C^α-only models may be necessary if such a model is retrieved from the PDB to be used in molecular replacement or homology modelling exercises; however, not many validation tools can handle such models (Kleywegt, 1997). The C^α backbone can be characterized by C^α—C^α distances (∼2.9 Å for a *cis*-peptide and ∼3.8 Å for a *trans*-peptide), C^α—C^α—C^α pseudo-angles and C^α—C^α—C^α—C^α pseudo-torsion angles (Kleywegt, 1997). The pseudo-angles and torsion angles turn out to assume certain preferred value combinations (Oldfield & Hubbard, 1994), much like the backbone φ and ψ torsions, and this can be employed for the validation of C^α-only models (Kleywegt, 1997). In addition to these straightforward methods, the mean-field approach of Sippl (1993) is also applicable to C^α-only models.

21.1.7.2.4. *Contacts and environments*

Hydrophobic, electrostatic and hydrogen-bonding interactions are the main stabilizing forces of protein structure. This leads to packing arrangements where hydrophobic residues tend to interact with each other, where charged residues tend to be involved in salt links and where hydrophilic residues prefer to interact with each other or to point out into the bulk solvent. Serious model errors will often lead to violations of such simple rules of thumb and introduce non-physical interactions (*e.g.* a charged arginine residue located inside a hydrophobic pocket; Kleywegt *et al.*, 1996) that serve as good indicators of model errors. Directional atomic contact analysis (Vriend & Sander, 1993) is a method in which these empirical notions have been formalized through database analysis. For every group of atoms in a protein, it yields a score which in essence expresses how 'comfortable' that group is in its environment in the model under scrutiny (compared with the expectations derived from the database). If a region in a model (or the entire model) has consistently low scores, this is a very strong indication of model errors. The *ERRAT* program is based on the same principle, but it is less specific in that it assesses only six types of non-bonded interactions (CC, CN, CO, NN, NO and OO; Colovos & Yeates, 1993).

Hydrogen-bonding analysis can often be used to determine the correct orientation of asparagine, glutamine and histidine residues (McDonald & Thornton, 1995). Similarly, an investigation of unsatisfied hydrogen-bonding potential can be used for validation purposes (Hooft *et al.*, 1996*b*), as can calculation of hydrogen-bonding energies (Morris *et al.*, 1992; Laskowski, MacArthur *et al.*, 1993).

Finally, a model should not contain unusually short non-bonded contacts. Although most refinement programs will restrain atoms from approaching one another too closely, if any serious violations remain they are worth investigating, since they may signal an underlying problem (*e.g.* erroneous omission of a disulfide restraint or incorrect side-chain assignment).

21.1.7.2.5. *Noncrystallographic symmetry*

Molecules that are related by noncrystallographic symmetry exist in very similar, but not identical, physical environments. This implies that their structures are expected to be quite similar, although different relative domain orientations and local variations may occur (*e.g.* owing to different crystal-packing interactions; Kleywegt, 1996). Many criteria have been developed to quantify the differences between (NCS) related models. Some, such as the r.m.s. distance (*e.g.* on all atoms, backbone atoms or C^α atoms) are based on distances between equivalent atoms, measured after a (to some extent arbitrary; Kleywegt, 1996) structural superpositioning operation has been performed. Others are based on a comparison of torsion angles, be it of main-chain φ, ψ angles [*e.g.* $\Delta\varphi$, $\Delta\psi$ plot (Korn & Rose, 1994); multiple-model Ramachandran plot (Kleywegt, 1996); $\sigma(\varphi)$, $\sigma(\psi)$ plot (Kleywegt, 1996); circular variance (Allen & Johnson, 1991) plots of φ and ψ (G. J. Kleywegt, unpublished results); Euclidian φ, ψ distances (Carson *et al.*, 1994) or pseudo-energy values (Carson *et al.*, 1994)] or side-chain χ_1, χ_2 angles [*e.g.* multiple-model χ_1, χ_2 plot (Kleywegt, 1996); $\sigma(\chi_1)$, $\sigma(\chi_2)$ plots (Kleywegt, 1996); circular variance (Allen & Johnson, 1991) plots of χ_1 and χ_2 (G. J. Kleywegt, unpublished results); Euclidian χ_1, χ_2 distances (Carson *et al.*, 1994) or pseudo-energy values (Carson *et al.*, 1994)]. Still other methods are based on analysing differences in contact-surface areas (Abagyan & Totrov, 1997), temperature factors (Kleywegt, 1996) or the geometry of the C^α backbone alone (Flocco & Mowbray, 1995; Kleywegt, 1996). Many of these methods can also be used to compare the structures of related molecules in different crystals or crystal forms (*e.g.* complexes, mutants).

21.1.7.2.6. *Solvent molecules*

Solvent molecules provide an excellent means of 'absorbing' problems in both the experimental data and the atomic model. Neither their position nor their temperature factor are usually restrained (other than by the data and restraints that prevent close contacts) and sometimes even their occupancy is refined. At a resolution of ∼2 Å, crystallographers tend to model roughly one water molecule for every amino-acid residue and at 1.0 Å resolution this number increases to ∼1.6 (Carugo & Bordo, 1999). When waters are placed, it should be ascertained that they can actually form hydrogen bonds, be it to protein atoms or to other water molecules. Considering that several ions that are isoelectronic with water (Na^+, NH_4^+) are often used in crystallization solutions, one should keep in mind the possibility that some entities that have been modelled as water molecules could be something else (Kleywegt & Jones, 1997). A method to check if water molecules could actually be sodium ions, based on the surrounding atoms, has been published (Nayal & Di Cera, 1996).

21.1.7.2.7. *Miscellaneous*

Many other coordinate-based methods for assessing the validity or correctness of protein models have been developed. These include the profile method of Eisenberg and co-workers (Bowie *et al.*, 1991; Lüthy *et al.*, 1992), the inspection of atomic volumes (Pontius *et al.*, 1996), and the use of threading and other potentials (Sippl, 1993; Melo & Feytmans, 1998; Maiorov & Abagyan, 1998). Some of these methods are described in more detail elsewhere in this volume. The program *WHAT IF* (Vriend, 1990) contains a large array of quality checks, many of which are not available in other programs, that span the spectrum from administrative checks to global quality indicators (Hooft *et al.*, 1996). During the refinement

process, coordinate shifts can be used as a rough indication of 'quality' or, rather, convergence (Carson *et al.*, 1994; Kleywegt & Jones, 1996*a*). Crude models tend to undergo much larger changes during refinement than models that are essentially correct and complete. Also at the residue level, large coordinate shifts indicate residues that are worth a closer look.

Laskowski *et al.* (1994) have formulated single-number geometrical quality criteria, which they dubbed '*G* factors' in analogy to crystallographic *R* values. These *G* factors combine the results of a number of quality checks (covalent geometry, main-chain and side-chain torsion angles *etc.*) in a single number.

21.1.7.3. *Model quality, temperature factors*

In crystallographic refinement, atomic displacement parameters (ADPs; often referred to as temperature factors or *B* factors) model the effects of static and dynamic disorder. Except at high resolution (typically better than 1.5 Å), where there are sufficient observations to warrant refinement of anisotropic temperature factors, ADPs are usually constrained to be isotropic. The isotropic temperature factor *B* of an atom is related to the atom's mean-square displacement $\langle \Delta r^2 \rangle$ according to $B = 8\pi^2 \langle \Delta r^2 \rangle / 3$. Compared with the atomic coordinates, there are usually comparatively few restraints on temperature factors during refinement. Therefore, particularly at low resolution, temperature factors often function as 'error sinks' (Read, 1990). They absorb not only the effects of static and dynamic disorder, but also of various kinds of model errors.

Compared with the wealth of statistics that can be used to check and validate coordinates, there are relatively few methods available to assess how reasonable a model's temperature factors are. One obvious check is to see how well the average temperature factor of the model matches the value calculated from the data, using either a Wilson plot (Wilson, 1949) or the Patterson origin peak (Vaguine *et al.*, 1999). Since the average temperature factor of a model is usually not restrained, this is a useful check that has been used on several occasions to justify high average *B* factors. One should keep in mind that a low average *B* factor, *per se*, is not necessarily an indication of high model quality. For instance, a backwards-traced protein model can have a considerably lower average *B* factor than a correct model at a similar resolution (Kleywegt & Jones, 1995*b*). Average (and minimum and maximum) temperature-factor values can also be listed separately for various groups of atoms (*e.g.* individual protein or nucleic acid molecules, ligands, solvent molecules). A simple plot of residue-averaged temperature factors as a function of residue number may reveal regions of the molecule that have consistently high *B* factors, which may be a consequence of problems in the model (Kleywegt *et al.*, 1996).

Other statistics pertain to the r.m.s. differences in *B* factors between atoms that are somehow related, for example through a chemical bond (r.m.s. $\Delta B_{\rm bonded}$), through a 1–3 interaction or through noncrystallographic symmetry (possibly after correcting for any differences between the average *B* factors of the NCS-related molecules). Sometimes these statistics are calculated separately for main-chain and side-chain atoms. If the *B* factors of such related atoms have been restrained to be similar during refinement, these checks do not provide a convincing indication of the quality of the model. On the other hand, the *B* factors of atoms that have non-bonded interactions are usually not restrained to be similar, which renders the r.m.s. *B*-factor difference between such atoms (r.m.s. $\Delta B_{\rm non\text{-}bonded}$) slightly more informative.

Since proteins tend to consist of a tightly packed core with more flexible regions at the surface, a radial *B*-factor plot (*i.e.* a plot of the average *B* factor of all atoms in a certain distance range from the centre of the molecule as a function of the distance) is expected to be shaped roughly like a half-parabola. Kuriyan & Weis (1991) used a ten-parameter isotropic rigid-molecule model of the mean-square atomic displacement (Schomaker & Trueblood, 1968). After obtaining values for the ten parameters (either by refinement against the structure-factor data or by fitting to the refined *B* factors of the model), the *B* factor of any atom can be calculated and depends only on its coordinates. They found that regions with large discrepancies between the refined and fitted *B* factors tend to be associated with errors or problems in a model.

Validation of anisotropic ADPs (Merritt, 1999), non-unit occupancies and H atoms, all of which are usually associated with high-resolution data, is still in its infancy. The validity of modelling anisotropic ADPs can be assessed by comparing the reduction of the conventional and free *R* values. If occupancies are used for multiple conformations of, for example, a side chain, the sum of the occupancies should be unity.

21.1.7.4. *Model versus experimental data*

21.1.7.4.1. *R values*

The traditional statistic used to assess how well a model fits the experimental data is the crystallographic *R* value,

$$R = \sum w\big||F_o| - k|F_c|\big| \Big/ \sum |F_o|.$$

This statistic is closely related to the standard least-squares crystallographic residual $\sum w(|F_o| - k|F_c|)^2$ and its value can be reduced essentially arbitrarily by increasing the number of parameters used to describe the model (*e.g.* by refining anisotropic ADPs and occupancies for all atoms) or, conversely, by reducing the number of experimental observations (*e.g.* through resolution and σ cutoffs) or the number of restraints imposed on the model. Therefore, the conventional *R* value is only meaningful if the number of experimental observations and restraints greatly exceeds the number of model parameters. In 1992, Brünger introduced the free *R* value ($R_{\rm free}$; Brünger, 1992*a*, 1993, 1997; Kleywegt & Brünger, 1996), whose definition is identical to that of the conventional *R* value, except that the free *R* value is calculated for a small subset of reflections that are not used in the refinement of the model. The free *R* value, therefore, measures how well the model predicts experimental observations that are not used to fit the model (cross-validation). Until a few years ago, a conventional *R* value below 0.25 was generally considered to be a sign that a model was essentially correct (Brändén & Jones, 1990). While this is probably true at high resolution, it was subsequently shown for several intentionally mistraced models that these can be refined to deceptively low conventional *R* values (Jones *et al.*, 1991; Kleywegt & Jones, 1995*b*; Kleywegt & Brünger, 1996). Brünger suggests a threshold value of 0.40 for the free *R* value, *i.e.* models with free *R* values greater than 0.40 should be treated with caution (Brünger, 1997). Tickle and coworkers have developed methods to estimate the expected value of $R_{\rm free}$ in least-squares refinement (Tickle *et al.*, 1998). Since the difference between the conventional and free *R* value is partly a measure of the extent to which the model over-fits the data (*i.e.* some aspects of the model improve the conventional but not the free *R* value and are therefore likely to fit noise rather than signal in the data), this difference $R_{\rm free} - R$ should be small (Kleywegt & Jones, 1995*a*; Kleywegt & Brünger, 1996). Alternatively, the $R_{\rm free}$ ratio (defined as $R_{\rm free}/R$; Tickle *et al.*, 1998) should be close to unity. Various practical aspects of the use of the free *R* value have been discussed by Kleywegt & Brünger (1996) and by Brünger (1997).

Self-validation is an alternative to cross-validation and in the case of crystallographic refinement, the Hamilton test (Hamilton, 1965) is a prime example of this. This method enables one to assess whether a reduction in the *R* value is statistically significant given

references