free R value than NCS constraints, then use constraints. If NCS restraints are to be employed, then use the experimental map to design a suitable NCS-restraint scheme (Kleywegt, 1999). Avoid the temptation to model alternative conformations in low-resolution maps or to place putative solvent molecules in every local maximum of the ($F_o - F_c$, α_c) difference map. In other words, be conservative and remember that the maxim 'where freedom is given, liberties are taken' is highly applicable to refinement programs (Hendrickson & Konnert, 1980; Kleywegt & Jones, 1995*b*).

(5) Adopt methodological advances as soon as they become available. Several innovations have only been slowly accepted by the mainstream (*e.g.* the use of databases in building and rebuilding, the use of the free R value, the use of electron-density averaging in molecular-replacement cases, bulk-solvent modelling). The most prominent recent development is the use of likelihood-based refinement programs (Bricogne & Irwin, 1996; Pannu & Read, 1996; Murshudov *et al.*, 1997; Adams *et al.*, 1997; Pannu *et al.*, 1998). These programs produce better models and maps and considerably reduce over-fitting (as assessed by the difference between the free and conventional R values).

(6) Most importantly, the crystallographer should be hyper-critical towards the fruits of his or her own labour. Every intermediate model is a hypothesis to be shot down (Jones & Kjeldgaard, 1994). The crystallographer should be more critical than the supervisor, the supervisor more critical than the referee and the referee more critical than the casual reader. It goes without saying that the reader, casual or not, should have access to model coordinates, experimental data and electron-density maps!

21.1.6. Final model

Once the refinement is finished [*i.e.* once the ($F_o - F_c$, α_c) difference map is featureless (Cruickshank, 1950) and parameter shifts in further refinement cycles are negligibly small], three tasks remain: validation of the final model, description and analysis of the structure, and deposition of the model coordinates and the crystallographic data with the Protein Data Bank (Bernstein *et al.*, 1977).

Until a few years ago, validation of the final model typically entailed calculating the conventional R value, r.m.s. deviations from ideal values of bond lengths and angles, average temperature factors, and a Luzzati-type estimate of coordinate error. Kleywegt & Jones (1995*b*) showed that these statistics are not necessarily even remotely related to the actual quality of a model. Based on these criteria, a backwards-traced protein model was of higher apparent quality than a carefully refined correct model. After this, the realisation sunk in that the best validation criteria are those that assess aspects of the model that are 'orthogonal' to the information used during model refinement and rebuilding. For instance, the main-chain φ and ψ torsion angles are usually not restrained during refinement; this makes the Ramachandran plot such a powerful validation tool (Kleywegt & Jones, 1996*b*, 1998). Other examples of useful independent tests include the profile method of Eisenberg and co-workers (Lüthy *et al.*, 1992), the directional atomic contact analysis method of Vriend & Sander (1993) and the threading-potential method of Sippl (1993).

In general, all quality checks provide necessary, but in themselves insufficient, indications as to whether or not a model is essentially correct. A truly good model should make sense with respect to what is currently known about physics, chemistry, crystallography, protein structures, statistics and (last, but not least) biology and biochemistry (Kleywegt & Jones, 1995*a*). A good model will typically score well on most if not all validation criteria, whereas a poor one will score poorly on many criteria. The same is true at the level of residues: a poor or erroneous region in a model will be characterized by violations of many residue-level quality criteria (Kleywegt & Jones, 1997).

21.1.7. A compendium of quality criteria

In this section, some of the quality and validation criteria that have been used by macromolecular crystallographers are summarized (for more detailed information, the reader is referred to the primary literature). When judging how useful or powerful these criteria are in a certain case, one should keep in mind that any criterion that has been used explicitly or implicitly during model refinement (*e.g.* geometric restraints) or rebuilding (*e.g.* rotamer libraries) does *not* provide a truly independent check on the quality of the model.

Many, but not all, of the criteria discussed below pertain specifically to protein models. Comparatively little work has been performed on the validation of nucleic acid models, although there are indications that there is a need for such procedures (Schultze & Feigon, 1997). The situation would appear to be even worse for hetero-entities (*e.g.* ligands, inhibitors, cofactors, covalent attachments, saccharides, metals, ions; van Aalten *et al.*, 1996; Kleywegt & Jones, 1998).

21.1.7.1. *Data quality*

Although many quality and validation criteria have been developed for assessing coordinate sets of protein models, comparatively few criteria are available for assessing the quality of the crystallographic data.

21.1.7.1.1. *Merging R values*

Possibly the most common mistake in papers describing protein crystal structures is an incorrectly quoted formula for the merging R value (calculated during data reduction),

$$R_{\rm merge} = \sum_h \sum_i |I_{h,i} - \langle I_h \rangle| / \sum_h \sum_i I_{h,i},$$

where the outer sum (h) is over the unique reflections (in most implementations, only those reflections that have been measured more than once are included in the summations) and the inner sum (i) is over the set of independent observations of each unique reflection (Drenth, 1994). This statistic is supposed to reflect the spread of multiple observations of the intensity of the unique reflections (where the multiple observations may derive from symmetry-related reflections, different images or different crystals). Unfortunately, $R_{\rm merge}$ is a very poor statistic, since its value increases with increasing redundancy (Weiss & Hilgenfeld, 1997; Diederichs & Karplus, 1997), even though the signal-to-noise ratio of the average intensities will be higher as more observations are included (in theory, an N-fold increase of the number of independent observations should improve the signal-to-noise ratio by a factor of $N^{1/2}$). At high redundancy, the value of $R_{\rm merge}$ is directly related to the average signal-to-noise ratio (Weiss & Hilgenfeld, 1997): $R_{\rm merge} \simeq 0.8/\langle I/\sigma(I) \rangle$.

Diederichs & Karplus (1997) have suggested a number of alternative measures that lack most of the drawbacks of $R_{\rm merge}$. Their statistic $R_{\rm meas}$ is similar to $R_{\rm merge}$, but includes a correction for redundancy (m),

$$R_{\rm meas} = \sum_h [m/(m-1)]^{1/2} \sum_i |I_{h,i} - \langle I_h \rangle| / \sum_h \sum_i I_{h,i}.$$

Another statistic, the pooled coefficient of variation (PCV), is defined as

$$\mathrm{PCV} = \sum_h \{[1/(m-1)] \sum_i (I_{h,i} - \langle I_h \rangle)^2\}^{1/2} / \sum_h \langle I_h \rangle.$$

Since PCV = $1/\langle I/\sigma(I)\rangle$, this quantity also provides an indication as to whether the standard deviations $\sigma(I)$ have been estimated appropriately. Finally, the statistic $R_{\rm mrgd\text{-}F}$, used for assessing the quality of the reduced data, enables a direct comparison of this merging R value with the refinement residuals R and $R_{\rm free}$.

Ideally, merging statistics should be quoted for all resolution shells (which should not be too broad), as well as for the entire data set. However, as a minimum, the values for the two extreme (low- and high-resolution) shells and for the entire data set should be reported.

21.1.7.1.2. *Completeness*

Data completeness can be assessed by calculating what fraction of the unique reflections within a range of Bragg spacings that could in theory be observed has actually been measured. Ideally, completeness should be quoted for all resolution shells (which should not be too broad), as well as for the entire data set. However, as a minimum, the values for the two extreme (low- and high-resolution) shells and for the entire data set should be reported.

21.1.7.1.3. *Redundancy*

Redundancy is defined as the number of independent observations (after merging of partial reflections) per unique reflection in the final merged and symmetry-reduced data set. Ideally, average redundancy should be quoted for all resolution shells (which should not be too broad), as well as for the entire data set. However, as a minimum, the values for the two extreme (low- and high-resolution) shells and for the entire data set should be reported.

21.1.7.1.4. *Signal strength*

The average strength or significance of the observed intensities can be expressed in different ways. Values that are often quoted include the percentage of reflections for which $I/\sigma(I)$ exceeds a certain value (usually 3.0) and the average value of $I/\sigma(I)$. Ideally, these numbers should be quoted for all resolution shells (which should not be too broad), as well as for the entire data set. However, as a minimum, the values for the two extreme (low- and high-resolution) shells and for the entire data set should be reported.

21.1.7.1.5. *Resolution*

The nominal resolution limits of a data set are chosen by the crystallographer, usually at the data-processing stage, and ought to reflect the range of Bragg spacings for which useful intensity data have been collected. Unfortunately, owing to the subjective nature of this process, resolution limits cannot be compared meaningfully between data sets processed by different crystallographers. Careful crystallographers will take factors such as shell completeness, redundancy and $\langle I/\sigma(I)\rangle$ into account, whereas others may simply look up the minimum and maximum Bragg spacing of all observed reflections. Bart Hazes (personal communication) has suggested defining the effective resolution of a data set as that resolution at which the number of observed reflections would constitute a 100% complete data set. Alternatively, Vaguine *et al.* (1999) define the effective (or optical) resolution as the expected minimum distance between two resolved peaks in the electron-density map and calculate this quantity as $2\Delta_P/2^{1/2}$, where Δ_P is the width of the origin Patterson peak. One day, hopefully, the term 'resolution' will be replaced by an estimate of the information content of data sets. Randy Read (personal communication) has carried out preliminary work along these lines.

21.1.7.1.6. *Unit-cell parameters*

The accuracy of unit-cell parameters has been shown to be grossly overestimated for small-molecule crystal structures (Taylor & Kennard, 1986). Not intimidated by this observation, some macromolecular crystallographers routinely quote unit-cell axes of 100–200 Å with a precision of 0.01 Å. An analysis of several high-resolution protein crystal structures has revealed that surprisingly large errors in the unit-cell parameters appear to be quite common (at least if synchrotron sources are used for data collection; EU 3-D Validation Network, 1998). Such errors can be detected *a posteriori* by checking if the bond lengths in a model show any systematic, perhaps direction-dependent, variations from their target values.

21.1.7.1.7. *Symmetry*

From the symmetry of the diffraction pattern, the point-group symmetry of the crystal lattice can usually be derived. It is important to merge the data in the point group with the highest possible symmetry (usually assessed using merging statistics) in order to minimize the chance of making an incorrect space-group assignment (Marsh, 1995, 1997; Kleywegt *et al.*, 1996). Once the first data set has been processed, it is always useful to compute a self-rotation function. A non-origin peak of comparable strength to the origin peak will indicate that the true space group has higher symmetry. [Similarly, a self-Patterson function can be calculated at this stage to detect any purely translational NCS (Kleywegt & Read, 1997).] Once the final model is available, a search for possibly missed higher symmetry can be carried out, *e.g.* using the method developed by Hooft *et al.* (1994).

Sometimes crystallographic symmetry breaks down (pseudo-symmetry): an apparent higher symmetry at low resolution does not hold at higher resolution. In some cases, this is a consequence of the chemistry of the system studied (*e.g.* an asymmetric ligand bound by a symmetric protein dimer). In other cases, it may go undetected and complicate space-group determination and solution and refinement of the structure.

When it comes to space-group determination, many of the lessons learned by small-molecule crystallographers also apply to macromolecular crystallography (Marsh, 1995; Watkin, 1996).

21.1.7.2. *Model quality, coordinates*

Many criteria (and computer programs) are available to check for structural outliers based only on analysis of Cartesian coordinate sets.

21.1.7.2.1. *Geometry and stereochemistry*

The covalent geometry of a model can be assessed by comparing bond lengths and angles to a library of 'ideal' values. In the past, every refinement and modelling program had its own set of 'ideal' values. This even made it possible to detect (with 95% accuracy) with which program a model had been refined, simply by inspecting its covalent geometry (Laskowski, Moss & Thornton, 1993). Nowadays, standard sets of ideal bond lengths and bond angles derived from an analysis of small-molecule crystal structures from the CSD (Allen *et al.*, 1979) are available for proteins (Engh & Huber, 1991; Priestle, 1994) and nucleic acids (Parkinson *et al.*, 1996). For other entities, typical bond lengths and bond angles can be taken from tables of standard values (Allen *et al.*, 1987) or derived by other means (Kleywegt & Jones, 1998; Greaves *et al.*, 1999).

For bond lengths, the r.m.s. deviation from ideal values is invariably quoted. Deviations from ideality of bond angles can be expressed directly as an angular r.m.s. deviation or in terms of angle distances (*i.e.* the angle $\angle ABC$ is measured by the 1–3 distance $|AC|$; note that this distance is also implicitly dependent on the bond

lengths $|AB|$ and $|BC|$). There are some indications that protein geometry cannot always be captured by assuming unimodal distributions (*i.e.* geometric features with only a single 'ideal' value). For example, Karplus (1996) found that the main-chain bond angle τ_3 ($\angle$N—C^α—C) varies as a function of the main-chain torsion angles φ and ψ.

Chirality is another important criterion in the case of biomacromolecules: most amino-acid residues will have the L configuration for their C^α atom. Also, the C^β atoms of threonine and isoleucine residues are chiral centres (IUPAC–IUB Commission on Biochemical Nomenclature, 1970; Morris *et al.*, 1992). Chirality can be assessed in terms of improper torsion angles or chiral volumes. For example, to check if the C^α atom of any residue other than glycine has the L configuration, the improper (or virtual) torsion angle C^α—N—C—C^β should have a value of about +34° (a value near −34° would indicate a D-amino acid). The torsion angle is called improper or virtual because it measures a torsion around something other than a covalent bond, in this case the N—C 'virtual bond'. The chiral volume is defined as the triple scalar product of the vectors from a central atom to three attached atoms (Hendrickson, 1985). For instance, the chiral volume of a C^α atom is defined as

$$V_{C^\alpha} = (\mathbf{r}_N - \mathbf{r}_{C^\alpha}) \cdot [(\mathbf{r}_C - \mathbf{r}_{C^\alpha}) \times (\mathbf{r}_{C^\beta} - \mathbf{r}_{C^\alpha})],$$

where $\mathbf{r}_X$ is the position vector of atom X. It should be noted that the chiral volume also implicitly depends on the bond lengths and angles involving the four atoms.

Another issue to consider is that of moieties that are necessarily planar (*e.g.* carboxylate groups, phenyl rings; Hooft *et al.*, 1996*a*). Again, planarity can be assessed in two different ways: by inspecting a set of (possibly improper) torsion angles and calculating their r.m.s. deviation from ideal values (*e.g.* all ring torsions in a perfectly flat phenyl ring should be 0°) or by fitting a least-squares plane through each set of atoms and calculating the r.m.s. distance of the atoms to that plane. Note that for double bonds, *cis* and *trans* configurations cannot be distinguished by deviations from a least-squares plane, but they can be distinguished by an appropriately defined torsion angle.

21.1.7.2.2. *Torsion angles (dihedrals)*

The conformation of the backbone of every non-terminal amino-acid residue is determined by three torsion angles, traditionally called φ (C_{i-1}—N_i—C_i^α—C_i), ψ(N_i—C_i^α—C_i—N_{i+1}) and ω (C_i^α—C_i—N_{i+1}—C_{i+1}^α). Owing to the peptide bond's partial double-bond character, the ω angle is restrained to values near 0° (*cis*-peptide) and 180° (*trans*-peptide). *Cis*-peptides are relatively rare and usually (but not always) occur if the next residue is a proline (Ramachandran & Mitra, 1976; Stewart *et al.*, 1990). The average ω-value for *trans*-peptides is slightly less than 180° (MacArthur & Thornton, 1996), but surprisingly large deviations have been observed in atomic resolution structures (Sevcik *et al.*, 1996; Merritt *et al.*, 1998). The ω angle therefore offers little in the way of validation checks, although values in the range of ±20 to ±160° should be treated with caution in anything but very high resolution models. The φ and ψ torsion angles, on the other hand, are much less restricted, but it has been known for a long time that owing to steric hindrance there are several clearly preferred combinations of φ, ψ values (Ramakrishnan & Ramachandran, 1965). This is true even for proline and glycine residues, although their distributions are atypical (Morris *et al.*, 1992). Also, an overwhelming majority of residues that are not in regular secondary-structure elements are found to have favourable φ, ψ torsion-angle combinations (Swindells *et al.*, 1995). For these reasons, the Ramachandran plot (essentially a φ, ψ scatter plot) is an extremely useful indicator of model quality (Weaver *et al.*, 1990; Laskowski, MacArthur *et al.*, 1993; MacArthur & Thornton, 1996; Kleywegt & Jones, 1996*b*; Kleywegt, 1996; Hooft *et al.*, 1997). Residues that have unusual φ, ψ torsion-angle combinations should be scrutinized by the crystallographer. If they have convincing electron density, there is probably a good structural or functional reason for the protein to tolerate the energetic strain that is associated with the unusual conformation (Herzberg & Moult, 1991). As a rule, the residue types that are most often found as outliers are serine, threonine, asparagine, aspartic acid and histidine (Gunasekaran *et al.*, 1996; Karplus, 1996). The quality of a model's Ramachandran plot is most convincingly illustrated by a figure. Alternatively, the fraction of residues in certain predefined areas of the plot (*e.g.* core regions) can be quoted, but in that case it is important to indicate which definition of such areas was used. Sometimes, one may also encounter a Balasubramanian plot, which is a linear φ, ψ plot as a function of the residue number (Balasubramanian, 1977).

In protein structures, the plane of the peptide bond can have two different orientations (approximately related by a 180° rotation around the virtual C^α—C^α bond) that are both compatible with a *trans* configuration of the peptide (Jones *et al.*, 1991). The correct orientation can usually be deduced from the density of the carbonyl O atom or from the geometric requirements of regular secondary-structure elements (in α-helices, all carbonyl O atoms point towards the C-terminus of the helix; in β-strands, carbonyl O atoms usually alternate their direction). In other cases, *e.g.* in loops with poor density, the correct orientation may be more difficult to determine and errors are easily made. By comparing the local C^α conformation to a database of well refined high-resolution structures, unusual peptide orientations can be identified and, if required, corrected (through a 'peptide flip'; Jones *et al.*, 1991; Kleywegt & Jones, 1997, 1998). Since flipping the peptide plane between residues i and $i+1$ changes the ψ angle of residue i and the φ angle of residue $i+1$ by ∼180°, erroneous peptide orientations may also lead to outliers in the Ramachandran plot (Kleywegt, 1996; Kleywegt & Jones, 1998).

All amino-acid residues whose side chain extends beyond the C^β atom contain one or more conformational side-chain torsion angles, termed χ_1 (N—C^α—C^β—X^γ, where X may be carbon, sulfur or oxygen, depending on the residue type; if there are two γ atoms, the χ_1 torsion is calculated with reference to the atom with the lowest numerical identifier, *e.g.* $O^{\gamma 1}$ for threonine residues), χ_2 (C^α—C^β—X^γ—X^δ) *etc.* Early on, it was found that the values that these torsion angles assume in proteins are similar to those expected on the basis of simple energy calculations and that in addition certain combinations of χ_1, χ_2 values are clearly preferred (so-called rotamer conformations; Janin *et al.*, 1978; James & Sielecki, 1983; Ponder & Richards, 1987). Analogous to Ramachandran plots, χ_1, χ_2 scatter plots can be produced that show how well a protein's side-chain conformations conform to known preferences (Laskowski, MacArthur *et al.*, 1993; Carson *et al.*, 1994). Alternatively, a score can be computed for each residue that shows how similar its side-chain conformation is to that of the most similar rotamer for that residue type. This score can be calculated as an r.m.s. distance between corresponding side-chain atoms (Jones *et al.*, 1991; Zou & Mowbray, 1994; Kleywegt & Jones, 1998) or it can be expressed as an r.m.s. deviation of side-chain torsion-angle values from those of the most similar rotamer (Noble *et al.*, 1993).

Other torsion angles that have been used for validation purposes include the proline φ torsion (restricted to values near −65° owing to the geometry of the pyrrolidine ring; Morris *et al.*, 1992) and the χ_3 torsion in disulfide bridges (defined by the atoms C^β—S—S′—$C^{\beta\prime}$ and restricted to values near +95 and −85°; Morris *et al.*, 1992). In addition to the torsion-angle values of individual residues, pooled standard deviations of χ_1 and/or χ_2 torsions have been used for validation purposes (Morris *et al.*, 1992; Laskowski, MacArthur *et al.*, 1993).

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

the increase in the number of degrees of freedom. Application of this test in the case of macromolecules is compounded by the difficulty of estimating the effect of the combined set of restraints on the (effective) number of degrees of freedom, but some information can nevertheless be gained from such an analysis (Bacchi *et al.*, 1996).

21.1.7.4.2. *Real-space fits*

The fit of a model to the data can also be assessed in real space, which has the advantage that it can be performed for arbitrary sets of atoms (*e.g.* for every residue separately). Jones *et al.* (1991) introduced the real-space *R* value, which measures the similarity of a map calculated directly from the model (ρ_c) and one which incorporates experimental data (ρ_o) as

$$R = \sum |\rho_o - \rho_c| / \sum |\rho_o + \rho_c|,$$

where the sums extend over all grid points in the map that surround the selected set of atoms. The real-space fit can also be expressed as a correlation coefficient (Jones & Kjeldgaard, 1997), which has the advantage that no scaling of the two densities is necessary. Chapman (1995) described a modification in which the density calculated from the model is derived by Fourier transformation of resolution-truncated atomic scattering factors.

The program *SFCHECK* (Vaguine *et al.*, 1999) implements several variations on the real-space fit. The normalized average displacement measures the tendency of groups of atoms to move away from their current position. The density correlation is a modification of the real-space correlation coefficient. The residue-density index is calculated as the geometric mean of the density values of a set of atoms, divided by the average density of all atoms in the model. It therefore measures how high the electron-density level is for the set of atoms considered (*e.g.* all side-chain atoms of a residue). The connectivity index is identical to the residue-density index, but is calculated only for the N, C^{α} and C atoms. It thus provides an indication of the continuity of the main-chain electron density.

21.1.7.4.3. *Coordinate error estimates*

Since a measurement without an error estimate is not a measurement, crystallographers are keen to assess the estimated errors in the atomic coordinates and, by extension, in the atomic positions, bond lengths *etc.* In principle, upon convergence of a least-squares refinement, the variances and covariances of the model parameters (coordinates, ADPs and occupancies) may be obtained through inversion of the least-squares full matrix (Sheldrick, 1996; Ten Eyck, 1996; Cruickshank, 1999). In practice, however, this is seldom performed as the matrix inversion requires enormous computational resources. Therefore, one of a battery of (sometimes quasi-empirical) approximations is usually employed.

For a long time, the elegant method of Luzzati (1952) has been used for a different purpose (namely, to estimate average coordinate errors of macromolecular models) than that for which it was developed (namely, to estimate the positional changes required to reach a zero *R* value, using several assumptions that are not valid for macromolecules; Cruickshank, 1999). A Luzzati plot is a plot of *R* value *versus* $2 \sin\theta/\lambda$, and a comparison with theoretical curves is used to estimate the average positional error. Considering the problems with conventional *R* values (discussed in Section 21.1.7.4.1), Kleywegt *et al.* (1994) instead plotted free *R* values to obtain a cross-validated error estimate. This intuitive modification turned out to yield fairly reasonable values in practice (Kleywegt & Brünger, 1996; Brünger, 1997). Read (1986, 1990) estimated coordinate error from σ_A plots; the cross-validated modification of this method also yields reasonable error estimates (Brünger, 1997).

Cruickshank, almost 50 years after his work on the precision of small-molecule crystal structures (Cruickshank, 1949), introduced the diffraction-component precision index (DPI; Dodson *et al.*, 1996; Cruickshank, 1999) to estimate the coordinate or positional error of an atom with a *B* factor equal to the average *B* factor of the whole structure. In several cases for which full-matrix error estimates are available, the DPI gives quantitatively similar results. *SFCHECK* (Vaguine *et al.*, 1999) calculates both the DPI and Cruickshank's 1949 statistic (now termed the 'expected maximal error') based on the slope and the curvature of the electron-density map.

21.1.7.4.4. *Noncrystallographic symmetry*

Despite the multitude of criteria for assessing conformational differences between related molecules, there was until recently no objective way to assess whether such differences were a true reflection of the experimental data or a manifestation of refinement artifacts (Kleywegt & Jones, 1995*b*; Kleywegt, 1996). However, it has been found that electron-density maps calculated with experimental phases (or, at least, phases that are biased as little as possible by the model) and amplitudes can be used to correlate expected similarities (based on the data) with observed ones (manifest in the final refined models; Kleywegt, 1999). This method uses a local density-correlation map, as introduced by Read (Vellieux *et al.*, 1995), to measure the local similarity of the density of two or more models on a per-atom or per-residue basis. By comparing these values to the observed structural differences in the final models, it is relatively easy to check if the latter differences are warranted by the information contained in the experimental data (Kleywegt, 1999).

21.1.7.4.5. *Difference density quality*

van den Akker & Hol (1999) described a method (called *DDQ*, standing for difference density quality) to assess the local and global quality of a model based on analysis of an ($F_o - F_c$, α_c) map calculated after omission of all water molecules. In this method, the map and model are used to calculate several scores. One score assesses the presence or absence of favourably positioned water peaks near polar and apolar atoms. Other scores provide a measure for the presence or absence of positive and negative shift peaks that may indicate incorrect coordinates, temperature factors or occupancies. The scores can be averaged per residue or for an entire model and can be used to detect problems in models. The method appears to be applicable to $\sim$3 Å resolution.

21.1.7.5. *Accountancy*

The opinions as to what constitutes an error in a model vary somewhat in the community [compare Hooft *et al.* (1996) and Jones *et al.* (1996), for instance], but most people would agree that a crystallographic error is one that requires access to the experimental data for its verification, and whose correction alters the calculated structure factors (*e.g.* position, *B* factor, occupancy or scattering factors of one or more atoms). In addition to this, there are nomenclature rules and conventions to which a model that is made publicly available should adhere. Separate from this is the issue of the (more clerical) validation of public database entries ('PDB files'; Hooft *et al.*, 1994, 1996) which, while important to maintain the integrity of these databases, ultimately ought to be the responsibility of the database curators (Jones *et al.*, 1996; Keller *et al.*, 1998; Abola *et al.*, 2000).

references