Acta Cryst. (1998). D**54**, 391–399

Determination of the Relative Precision of Atoms in a Macromolecular Structure

GENFA ZHOU,† JUNFENG WANG†, ERIC BLANC AND MICHAEL S. CHAPMAN*

Department of Chemistry and Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306–3015, USA. E-mail: chapman@sb.fsu.edu

(*Received 12 March 1997; accepted 27 August 1997*)

Abstract

Several real-space indices and temperature factors are compared with respect to their correlation with atomic positional error and their ability to indicate atoms and residues with the worst of subtle errors. The best index, r^{ED}, is a correlation coefficient between model and map electron densities, similar to one proposed earlier, but incorporating two improvements. Firstly, resolution is accounted for explicitly by calculating the model electron density by Fourier transformation of resolution-truncated scattering factors. Secondly, the deviation between model and map electron densities is assigned to neighboring atoms according to their contribution to the electron density of each grid point. With maps of various qualities, r^{ED} is the single index with best correlation to atomic error with grouped or individual atoms, and it is the most reliable indicator of poor residues. With poorer omit maps, imprecision of individual atoms is best diagnosed by a combination of low r^{ED} or high B factor. With the improved methods, 60–70% of the least precise atoms can detected in a fully refined structure. Similarly, 40–80% of the least precise atoms of an unrefined model can be detected by comparison with an isomorphous replacement map. This is useful in assessing and improving the quality of a model, but not sufficient to confidently validate all atoms of a structure at sub-atomic resolution.

1. Introduction

The difficulty of estimating standard errors for individual macromolecular atomic parameters (Sheldrick, 1996; Ten Eyck, 1996), means that the most commonly used indicators are global. That is, they offer an assessment of the average error of all coordinates (Brünger, 1997; Luzzati, 1952). These assessments are based on the consistency of atomic parameters with the experimental structure amplitudes, ($|F|$). They are, therefore, fundamentally limited by a Fourier relationship, that makes each $|F|$ dependent in part on all atoms, making atomic errors interdependent (Ten Eyck, 1996). Jones (Brändén & Jones, 1990) circumvented this difficulty by comparing the model with the local values of the electron density, suggesting the use of either a 'real-space' (or 'residue') R factor, $R^{residue}$. A linear correlation coefficient, $r^{residue}$, is currently preferred, because, in principle, it is not dependent on scale constants (Read, 1986). Thus, particularly strong or weak density does not, in itself, lead to a poor index value.

An inherent limitation of real-space indices is the dependence of the experimental electron-density map on the phases. Use of (inaccurate) multiple isomorphous replacement (MIR) phases would be likely to underestimate the precision of a refined structure, but the use of phases calculated from the model carries the danger of underestimating the error with the bias likely to be of map to model, even with '$2F_o - F_c$' syntheses (Bhat & Cohen, 1984; Luzzati, 1953). It is an overstatement to claim (Kleywegt & Jones, 1996) that use of a simulated-annealing omit map (Hodel *et al.*, 1992) constitutes a real-space equivalent of Brünger's free R factor (Brünger, 1992) (because $|F|$'s are used for both refinement and map calculation), but Kleywegt & Jones (1996) correctly assert that such a map, with little model bias, is a good choice. These measures of real-space fit have been used successfully to pin-point regions of the model where the sequence is out of register (Jones *et al.*, 1991).

Previously reported real-space indices were not claimed to be sensitive to subtle errors. When several models of *P*2 myelin protein were built, it was only models with root-mean-square (r.m.s.) deviations from the refined structure (R = 15.7%, 2.7 Å) exceeding 2 Å that had elevated average residue R factors (Jones *et al.*, 1991, Table 4). Smaller errors are usually detected indirectly, through deviation from ideal stereochemistry. Crystallographic refinement of poor model introduces distortions to the stereochemistry that can be recognized (Laskowski *et al.*, 1993) as the model moves to fit the experimental data. In a comprehensive analysis of many possible local indices, Carson *et al.* (1994) found that the real-space R factors and B factors were the single most reliable indicators, but that combinations of five parameters were more effective. Again, the objective was to detect relatively gross discrepancies between test structures with a root-mean-square difference of 2.8 Å after refinement at 2 Å resolution.

† The first two authors contributed equally to this work.

The work described here evaluates various (real-space) indicators as direct detectors of model errors that are more subtle than the mis-aligned sequences studied by Jones *et al.* (1991) and about $\frac{1}{4}$ of the magnitude studied by Carson *et al.* (1994). Although conceptually similar to the indicators tested earlier, for the indices tested here, electron density is calculated from the model in a different way. Jones *et al.* (1991), like Diamond (1971), used spherical Gaussian functions to approximate the electron density surrounding each atom directly in real space. The resolution limit is accounted for by searching for empirical smearing constants that give the best correlation (Jones & Kjeldgaard, 1996). Here, by contrast, the electron density is calculated from the Fourier transform of the atomic scattering factor (Chapman, 1995) which can be explicitly truncated at the appropriate resolutions. The practical difference at medium resolution is that grid points farther away from an atom can be used, because the tails of electron-density functions of neighboring atoms are calculated more precisely.

2. Methods

The following real-space indicators were compared.

(i) The residue R factor R^{residue}, calculated with spherical Gaussians according to Jones *et al.* (1991).

(ii) The residue correlation coefficient r^{residue}, also calculated according to Jones *et al.* (1991).

(iii) A weighted real-space R factor R^{ED}, using the resolution-dependent electron-density function of Chapman (1995) in contrast to the spherical Gaussian functions used by Jones *et al.* (1991),

$$R_j^{\text{ED}} = \frac{\sum_{x \text{ near } j} w_{x,j} |S\rho_{\text{map}}(\mathbf{x}) + k - \rho_{\text{model}}(P, \mathbf{x})|}{\frac{1}{2} \sum_{x \text{ near } j} w_{x,j} |S\rho_{\text{map}}(\mathbf{x}) + k + \rho_{\text{model}}(P, \mathbf{x})|}, \quad (1)$$

where $\mathbf{x}$ is the set of grid points near the j subset of all J atoms (such as those of a single amino acid). S and k are optimized scale constants, and P is the set of atomic parameters (x, y, z, B...). The factor of $\frac{1}{2}$ in the denominator that makes R^{ED} analogous to reciprocal-space R factors, but doubles the expected numerical value relative to that of Jones *et al.* (1991). Unlike the prior work, grid points are weighted,

$$w_{x,j} = \frac{|\rho_{j,\text{model}}(\mathbf{x})|}{\sum_J |\rho_{J,\text{model}}(\mathbf{x})|} \quad (2a)$$

or

$$w_{x,j} \simeq \frac{|\rho_{j,\text{model}}(\mathbf{x})|}{\sum_J \rho_{J,\text{model}}(\mathbf{x})}. \quad (2b)$$

Without weighting, grid points within the cutoff radii of several atoms would contribute more to the statistics. Our weighting proportionately distributes residual error to all neighboring atoms such that the total contribution to all atoms of each grid point is the same [$\sum_J w_{x,j} = 1$, from (2*a*)]. The (default) approximation in (2*b*) does not account for the small negative 'ripple' from distant atoms (because of resolution truncation), but the approximation affects the R factor little and improves computational efficiency. R factors for individual residues were calculated using scaling constants S and k, determined only once from the entire protein.

Table 1. *Determination of optimal parameters for index calculation for MBP*

R^{ED}, r^{ED} and Γ depend on two cutoff radii: the contribution to ρ_{model} is ignored for atoms $> r_{\text{calc}}^{\max}$ from any grid point, while indices are calculated using only grid points within $r_{\text{ref}}^{\max}$ of indexed atoms, j. Their choice balances statistical reliability, noise limitation and computational expediency. Details are given in Chapman (1995), but briefly, optimal values of $r_{\text{ref}}^{\max}$ are expected to be between typical van der Waals radii and the high-resolution limit, while $r_{\text{calc}}^{\max}$ should be larger than $r_{\text{ref}}^{\max}$ by at least a van der Waal's radius. In Table 2 for MBP, $r_{\text{ref}}^{\max}$ = 1.5 to 1.8 Å and $r_{\text{calc}}^{\max}$ = 3.5 Å are shown to be optimal for the calculation of r^{ED}, and for calculation of the R^{ED} with greatest sensitivity to coordinate change (column 3). Values of 1.5 and 3.5 Å were used. Constants A_0 and C, used to determine the radii of Gaussian electron-density functions in R^{residue} and r^{residue}, were similarly optimized for the maps at different resolutions, but were close to the values expected (Jones & Kjeldgaard, 1996).

$r_{\text{ref}}^{\max}$ (Å)	$r_{\text{calc}}^{\max}$ (Å)	$(R_{0.3}^{\text{ED}} - R_0^{\text{ED}})/\sigma(R_0^{\text{ED}})$	r_0^{ED}
1.5	2.0	1.717	0.876
1.2	2.5	1.625	0.859
1.5	3.5	1.765	0.881
1.8	3.5	1.672	0.888
2.1	3.5	1.651	0.882
2.4	3.5	1.378	0.872
1.5	5.0	1.712	0.885

(iv) A weighted linear correlation coefficient, r^{ED}, calculated using *RSRef*,

$$r_j^{\text{ED}} = \frac{\langle \rho_{\text{map}} \rho_{\text{model}} \rangle - \langle \rho_{\text{map}} \rangle \langle \rho_{\text{model}} \rangle}{[(\langle \rho_{\text{map}}^2 \rangle - \langle \rho_{\text{map}} \rangle^2)(\langle \rho_{\text{model}}^2 \rangle \langle \rho_{\text{model}} \rangle^2)]^{1/2}}, \quad (3)$$

from Spiegel (1975), equation 8.32. In contrast to the familiar unweighted form used for r^{residue}, means are weighted in a conventional manner,

$$\langle \rho_{j,\text{map}}^m \rho_{j,\text{model}}^n \rangle = \frac{\sum_{\mathbf{x} \text{ near } j} w_{x,j} \rho_{j,\text{map}}^m(\mathbf{x}) \rho_{j,\text{model}}^n(P, \mathbf{x})}{\sum_{\mathbf{x} \text{ near } j} w_{k_j}} \quad (4)$$

$$\text{for } (m, n) \subset (0, 1, 2).$$

With weighting there is no need to subtract the contribution of neighboring atoms, a process that reintroduces dependence on scale factors into r^{residue}, and unreasonable degrees of freedom by allowing scaling that need not be consistent with the linear regression implicit in the correlation coefficient of indexed atoms. r^{ED} also differs in the incorporation of resolution in the calculation of electron density.

(v) The magnitude of the derivative of the real-space residual, calculated for all atoms i in set j,

$$\Gamma_j = \tfrac{1}{j}\sum_{i\subset j}\left[\left(\frac{\partial\Omega_i^p}{\partial x}\right)^2+\left(\frac{\partial\Omega_i^p}{\partial y}\right)^2+\left(\frac{\partial\Omega_i^p}{\partial z}\right)^2\right]^{1/2} \quad (5)$$

where

$$\Omega_i^p(P) = \sum_{\mathbf{x}\ \text{near}\ i}[S\rho_{\text{map}}(\mathbf{x}) + k - \rho_{\text{model}}(P,\mathbf{x})]^2. \quad (6)$$

Like an index suggested by Wodak *et al.* (1996), Γ_j is based on difference electron density, and like the stereochemical screening of Tronrud & Ten Eyck (1992) it uses first derivatives.

The real-space indicators were also compared with the refined B factors. The simple mean B factor among atoms in a residue was used, because it was better correlated to positional error than the root-mean-square B factor.

The indices depend on constants such as cutoff radii for the calculation of atomic electron density. These constants can be optimized (Table 1).

2.1. *Test systems*

(*a*) Most tests were performed with the ytterbium complex of mannose binding protein (MBP) A, for which a high-resolution (1.8 Å) structure and map were available (Burling *et al.*, 1996). An accurate map had been calculated with model-independent multi-wavelength anomalous dispersion (MAD) phases (Hendrickson, 1991), but the structure had been refined against only the structure amplitudes (Burling *et al.*, 1996). The availability of coordinates from a prior 2.3 Å refinement of an isomorphous holmium complex (Weis *et al.*, 1991) allowed examination of the correlation of quality indices with the discrepancy between the refinements at different resolutions (see below).

(*b*) Low-resolution tests involved canine parvovirus (CPV) (Tsao *et al.*, 1991; Xie & Chapman, 1996). The structure had been refined in real space (Chapman & Rossmann, 1996) at 2.9 Å. It was compared with a 3.25 Å model-independent map that was of high quality (a correlation coefficient of 0.91 was similar to MBP's). This had resulted from application of 60-fold non-crystallographic symmetry (Kim *et al.*, 1989; Tsao *et al.*, 1992).

(*c*) α-Amylase inhibitor Hoe-467 A (Pflugrath *et al.*, 1989) was used for tests against a (poorer) MIR map. Phases to 2.5 Å had been determined from four derivatives with a mean figure of merit of 0.63. The structure had been refined in reciprocal space to an R factor of 20% to 2 Å resolution.

2.2. *Three methods of model error simulation*

2.2.1. *Rigid translations.* For all test systems, the model was displaced from the refined optimum with rigid-body translations. Although often inappropriate for reciprocal-space tests, rigid-body translations are more stringent test in real space than random displacements because good stereochemistry is maintained. Thus, overlapping electron density of neighboring atoms does not become an unrealistically easy way to diagnose error. Furthermore, the even distribution of error throughout the structure resembles that of a real structure determination. The rigid translations were repeated for different randomly chosen directions and different size displacements. Interpretation was non-trivial. Because of the varying strength of electron density, there is inherent variation in real-space quality indices. What index values suggest inaccurate model? More importantly, how do different indices compare in revealing error? Let ζ be one of the quality indices. Its sensitivity, s, is defined here as the displacement of coordinates required to change ζ by one standard deviation, σ_0, such that,

$$\zeta(P + s.\mathbf{r}) - \zeta(P) = \sigma_0[\zeta(P)] \quad (7)$$

where $\mathbf{r}$ is a unit vector of random direction.

2.2.2. *Homologous lower resolution structures.* More realistic tests took advantage of a second MBP structure deposited in the PDB. The pair of structures have different lanthanide substituents: ytterbium for the 1.8 Å structure (Burling *et al.*, 1996), and holmium for the 2.3 Å structure (Weis *et al.*, 1991), but are otherwise isomorphous. Determined at 1.8 Å resolution, with more recent anomalous dispersion phasing methods, it is likely that the Yb structure is more precise. However, the Ho structure was also exhaustively refined (R = 18%, 2.3 Å) and is therefore representative of the errors that might remain in a typical structure refined at medium to high resolution. It was assumed that the difference between the two coordinate sets would be proportionate to the error of the Ho structure. Although the Yb structure was determined with high precision using an accurate experimental electron-density map, it is unlikely to be error-free, and thus the assumption is a first-order approximation only. Cryo-data collection had been used for the Yb structure (only), so the Ho coordinates were first overlaid on the Yb coordinates by least-squares rigid-body supposition using all atoms except a few residues near the termini of each chain. The r.m.s. displacement upon supposition was 0.66 Å.

Quality indices for the superposed Ho structure were calculated with respect to the three maps. The first two were calculated with the high-quality MAD-phases map of Burling *et al.* (1996) at 1.8 and 2.3 Å resolution, respectively. The third was a 2.3 Å σ_A-weighted (Read, 1986) simulated-annealing omit map (Hodel *et al.*, 1992) calculated from the Yb-MBP structure amplitudes and phases from the superposed Ho-MBP structure refined at 2.3 Å (Weis *et al.*, 1991). Similar maps could be

calculated at the end of the most structure determinations. 46 omit maps were calculated, each omitting a window of five amino acids plus immediate neighbors. Indices were calculated using the appropriate omit map for each amino acid.

2.2.3. *Models degraded by molecular dynamics.* Additional tests for the α-amylase inhibitor used derivatives of the refined model (Pflugrath *et al.*, 1989) that were degraded by torsion angle molecular dynamics (Rice & Brünger, 1994) in the absence of X-ray restraints. Similarly degraded models have been used to test refinement protocols (Adams *et al.*, 1997). To represent situations typical near the start and end of the model building into MIR maps, two models were used with r.m.s. backbone errors of 1.4 and 0.6 Å, respectively.

2.3. *Evaluation of indices as diagnostic indicators of error*

Indices were evaluated graphically, through correlation to the atomic error, and for their ability to detect the residues (or atoms) with highest error. Residues/atoms with quality indices in the worst tenth percentile were compared to residues/atoms with coordinate errors in the worst tenth percentile. Those belonging to both groups were designated true positives. True negatives belonged to neither group. False negatives were those with coordinate error in the worst tenth percentile, but quality index in the top 90%. False positives were among the best 70% regarding coordinate error, but had indices among the 10% worst. For models degraded by molecular dynamics, the worst tenth percentile had very large coordinate errors (up to 3.2 Å), so the tests were modified to test the effectiveness in diagnosing coordinate errors greater than 1 or 1.5 Å.

3. Results

3.1. *Validation of the electron-density calculation*

Correlation coefficients of r = 0.92 and 0.89 for protomers A and B had previously been reported (Burling *et al.*, 1996). Burling *et al.* calculated these statistics by comparing experimental and model electron-density values for all pixels within the van der Waals volume of the protein, and using model electron density calculated by Fourier transformation. Using the same MAD-phased map, and similar cutoffs ($r_{\rm ref}^{\rm max}$ = 1.8, $r_{\rm calc}^{\rm max}$ = 5.0 Å), but using the electron-density calculation method of Chapman (1995) gives an overall unweighted r = 0.912. Thus, for near-complete data sets, our method of electron-density calculation does not limit the precision of comparison, even for maps of the highest quality.

The high correlation coefficients were obtained despite two known discrepancies between calculated and experimental electron densities. Firstly, neither the Burling *et al.* (1996) method, nor that used here, accounts in the calculated density for the blurring of electron-density maps due to phase error and/or figure of merit (FOM) weighting. Secondly, unlike Burling's method, our method ignores the effect on the map of missing reflections. *RSRef* was modified to account for the spherically averaged effects of FOM-weighting and/or missing data by attenuation of the form factors used for electron-density calculation. Our crude correction had slightly deleterious effects, and was not used further. (Attenuation by the fraction of data observed at each resolution likely overestimates the effect of unobserved reflections, because they are likely to be systematically weak.)

3.2. *Comparison of the current indices with those calculated using spherical Gaussian functions*

Fig. 1 shows that the electron-density functions used for $R^{\rm ED}$ and $r^{\rm ED}$ are more consistent with the electron-

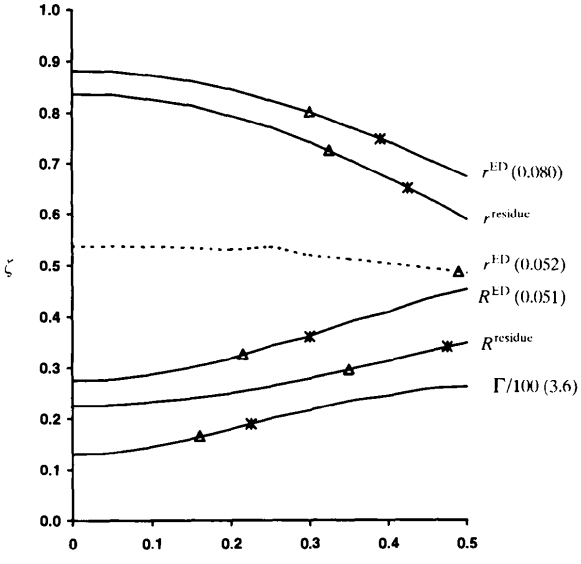

Fig. 1. Comparison of indicators (ζ) using the 1.8 Å MAD-phased map of MBP (solid lines, Burling *et al.*, 1996) and the 2.5 Å MIR map of the α-amylase inhibitor (dashed line, Pflugrath *et al.*, 1989). Error was introduced into the respective refined structures by rigid translations of all atoms in randomly chosen directions. Quality indices were then calculated for each amino acid (all atoms) and averaged. Δs mark s values (the solutions to equation 7), the points at which the change in quality index equals the standard deviation (σ) of the index of the unmoved structure (given in parentheses, above). * mark 1.65σ, corresponding to the 95% limit of a normal probability distribution. Lower error values for Δ and * indicate greater sensitivity, so this figure confirms that indices calculated with resolution-dependent electron-density functions are superior to those calculated with spherical Gaussians. Furthermore, although $R^{\rm ED}$ is expected to be twice $R^{\rm residue}$ (due to the 1/2 in the denominator of equation 3), it is less than that, indicating better agreement between experimental and calculated electron density, as also seen with the correlation coefficients.

density map than the spherical Gaussian functions used in R^{residue} and r^{residue}. r^{ED} is clearly higher than r^{residue} (*i.e.* in better agreement) and it falls sharply with coordinate error. The sensitivities marked in Fig. 1 are useful for comparing between different R factors, or between different correlation coefficients (r), but less useful in comparing between Γ, R and r. The standard deviation of the correlation coefficients $\sigma(r)$ is nearly independent of the coordinate error, whereas $\sigma(\Gamma)$ is strongly dependent, making r^{ED} the most sensitive indicator with more realistic non-uniform distributions of coordinate error (Table 2). Parallel calculations with CPV at 3.25 Å showed that the indices are only modestly dependent on resolution with r^{ED}, R^{ED} and Γ sensitivities of 0.39, 0.29 and 0.29 Å, respectively. Phase quality is more important. The change in r^{ED} with coordinate error is much less marked for the 2.5 Å MIR α-amylase inhibitor map (Fig. 1), but a sensitivity of 0.49 Å indicates that r^{ED} will still be useful when building a crude model, when MIR maps are typically used.

3.3. *Correlation of indices with coordinate error*

Table 2 shows that r^{ED} is the index best correlated with coordinate error and that it is a modest improvement upon r^{residue} (Jones *et al.*, 1991). The correlation between r^{ED} and model error is illustrated graphically in Fig. 2. Indices are improved by resetting all B factors to a constant (Table 2). The correlation is best when side chains are excluded, and when the effects of intrinsic variation in electron-density quality of different side-chain types do not matter. Perhaps surprisingly, reasonable rank correlation is obtained with r^{ED} for individual atoms, even with sub-atomic resolution maps. With the 2.3 Å resolution omit map, r^{ED} is a slightly better predictor of error than individual B factors, but when atoms are grouped, or with better

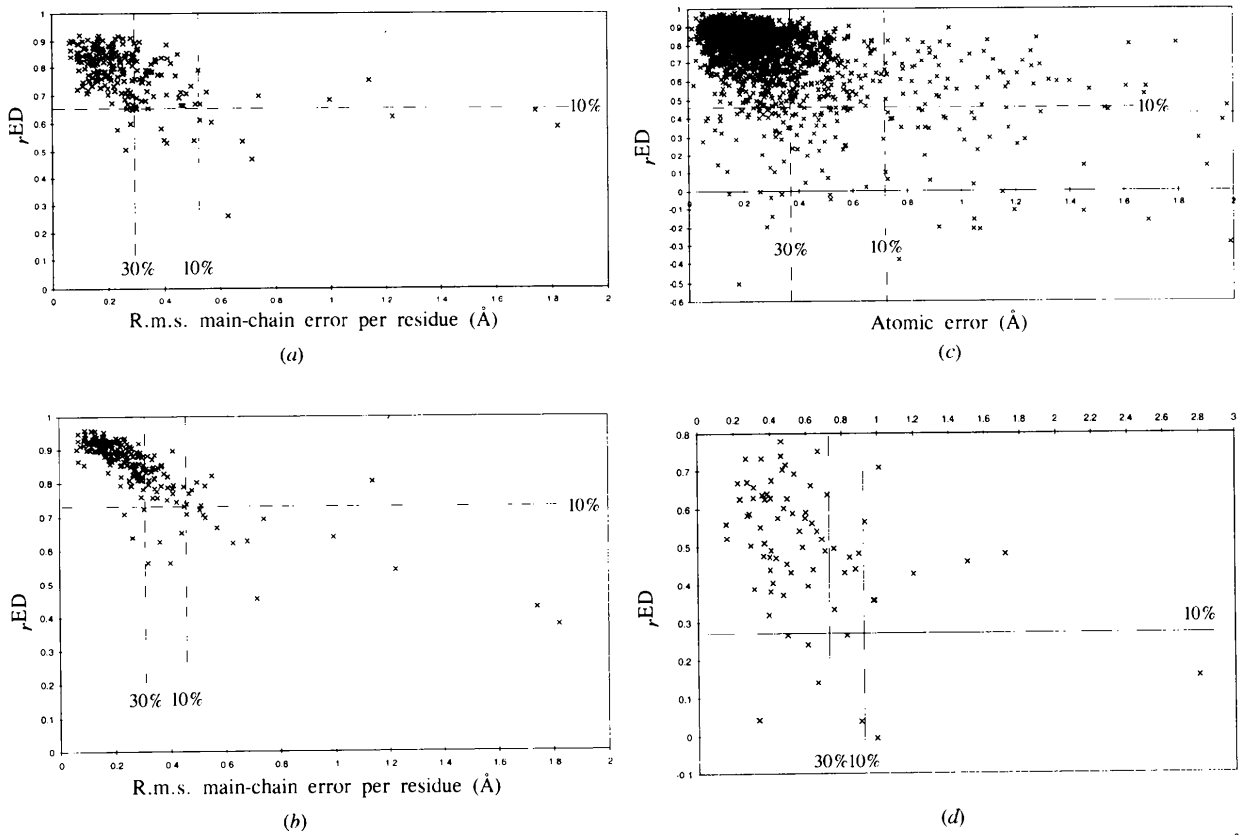

Fig. 2. Correlation of r^{ED} with coordinate error. (*a*) The Ho-MBP structure is compared with a simulated-annealing omit map at 2.3 Å resolution. There is a clear correlation between r^{ED} and coordinate error, but there is considerable scatter. Dotted lines indicate the percentile limits for r^{ED} and coordinate discrepancy (between Ho-MBP and Yb-MBP) used to test r^{ED} as a diagnostic indicator of error. Most, but not all, of the least accurate residues are correctly identified as those with lowest r^{ED}. A 2 Å cutoff has been used as in (*b*) and (*c*) excluding three residues with errors between 2.5 and 6 Å. (*b*) The Ho-MBP structure is now compared with the 1.8 Å Yb-MBP MAD map. The scatter is much reduced, implying that the limitation in (*a*) is map quality. Remaining scatter can be attributed to the index and to the unknown residual error in the 1.8 Å structure that has been assumed to be zero in using coordinate discrepancy as a measure of error. The correlation is good enough that errors of a few tenths of an Å can be reliably detected. (*c*) and (*d*) show more challenging cases with r^{ED} calculated for: (*c*) individual atoms against a 2.3 Å omit-map and (*d*) for residues against a 2.5 Å MIR map. There is considerable scatter, but the trend is visible that the atoms with worst index have the worst error.

Table 2. *Correlation of quality indices with coordinate error for MBP*

Spearman (Spiegel, 1975) rank correlation coefficients are shown parenthetically next to the more familiar linear correlation coefficients. Rank correlation measures how similarly the coordinates are ordered in terms of index and error. Rank coefficients are more appropriate, because they make no assumption about a linear relationship between error and index. They test whether the index orders atoms/residues in order of their error. Improvements seen after setting B factors to a constant likely have several causes: (1) Refinement tends to increase the B factors of poorly modeled regions, artificially lowering the discrepancy between map and model density. (2) Disordered regions have less precisely refined atomic positions leading to a larger estimate of error when calculated from the difference between two structures. Lowering the temperature factors of these regions amplifies the discrepancies between map and model electron density. (3) Overfit B factors disrupt the electron density scaling of the whole molecule, affecting R^{ED} but not $r^{I:D}$. Note that Jones *et al.* (1996) achieve an effect, similar to making B factors constant, through their transformation of B factors into atomic radii. This implicitly adds 11 Å^2 to all B factors and reduces variation between them.

		1.8 Å MAD map			2.3 Å σ_A-weighted SA-omit map		
		Grouping atoms in residue		Individual	Grouping atoms in residue		Individual
Method		Main chain	All atoms	atoms	Main chain	All atoms	atoms
R^{ED}		0.52 (0.67)	0.35 (0.59)	0.31 (0.55)	0.11 (0.29)	0.03 (0.24)	0.13 (0.31)
R^{ED}	B's set to 20 Å^2	0.70 (0.78)	0.60 (0.75)	0.48 (0.62)	0.42 (0.49)	0.37 (0.53)	0.28 (0.42)
$R^{residue}$	(O)	0.79 (0.49)	0.68 (0.66)	N/A	0.39 (0.43)	0.43 (0.56)	N/A
r^{ED}		−0.86 (0.83)	−0.67 (0.81)	−0.54 (0.70)	−0.56 (0.53)	−0.43 (0.57)	−0.35 (0.48)
r^{ED}	B's set to 20 Å^2	−0.85 (0.82)	−0.69 (0.81)	−0.56 (0.67)	−0.51 (0.54)	−0.45 (0.61)	−0.36 (0.48)
$r^{residue}$	(O)	−0.84 (0.74)	−0.68 (0.74)	N/A	−0.37 (0.46)	−0.39 (0.58)	N/A
Γ		−0.17 (−0.06)	−0.32 (0.26)	−0.15 (0.03)	−0.24 (0.11)	−0.37 (0.36)	−0.17 (0.10)
B factor (Å^2)					0.57 (0.40)	0.51 (0.39)	0.55 (0.45)
		2.3 Å MAD map					
$r^{I:D}$	B's set to 20 Å^2	−0.84 (0.75)	−0.66 (0.74)	−0.50 (0.60)			

Table 3. *Correlation of quality indices with coordinate error for α-amylase inhibitor*

α-Amylase inhibitor: 2.5 Å MIR map, B's set to 20 Å^2.

	R.m.s. backbone error = 0.6 Å	Grouping atoms in residue		Individual
Method	R.m.s. backbone error (Å)	Main chain	All atoms	atoms
R^{ED}	0.6	0.36 (0.33)	0.27 (0.37)	0.49 (0.44)
r^{ED}	0.6	−0.38 (−0.36)	−0.36 (−0.43)	−0.62 (−0.45)
R^{ED}	1.4	0.58 (0.50)	0.52 (0.47)	0.36 (0.46)
r^{ED}	1.4	−0.70 (−0.64)	−0.64 (−0.57)	−0.48 (−0.54)

maps, r^{ED} is substantially better. Table 2 and Fig. 2 show, not surprisingly, that the correlation is best for high-quality maps. Comparisons of r^{ED} calculated with MAD and omit maps at various resolutions in Table 2 and Fig. 2 show, not surprisingly, that the correlation is best for high-quality maps. Comparisons of r^{ED} calculated with MAD and omit maps at various resolutions in Table 2 with an MIR map in Table 3 show that the quality of phases (MAD > omit > MIR) is more important than map resolution.

3.4. *Quality indices as diagnostic indicators of error*

In judging the indices as a diagnostic tool (Table 4), the most important statistics are the proportion of true positives, which ideally should be 10%, and also the proportion of false negatives (ideally 0%) that are the residues/atoms with high error that were undetected by the quality index. Table 4 shows that r^{ED} is the most reliable indicator of the worst errors, although the improvement over other indices is modest. In most cases, indices are more reliable with constant B factors. Comparison of the test with MAD and omit maps in Table 4 and against an MIR map in Table 5 show once again that phase quality is more important than resolution. With MIR maps, tests are not sufficiently reliable to diagnose small errors in a refined model. However, they are effective in earlier stages of structure determination when MIR maps are usually used, and the unrefined models typically contain large errors (Table 5).

For individual atoms, Table 4 shows that temperature factors can be a reasonable diagnostic indicator, but r^{ED} is better with the high-quality 1.8 Å MAD map. B factors and $r^{I:D}$ are complementary indicators in several ways. $r^{I:D}$ is sensitive to map/phase quality that is not relevant to B factors. Table 4 shows that the sensitivity of r^{ED} to map quality is overcome by grouping atoms, averaging out the effects of phase error. B factors embody both disorder and model quality, whereas real-space indicators are more specific for quality, because in disordered regions, both calculated and observed density is weak and in agreement. B factors and r^{ED} are (only) partially correlated ($r = -0.53$ for individual atoms *versus* omit map). So, by combining the information, is a more reliable diagnostic achieved? Table 4 shows the results of selecting atoms with either the highest B factors or lowest r^{ED}. To select a total 10% of that atoms, for the omit map, the worst 6.5% for both B factor and $r^{I:D}$, were selected, but for the MAD maps, 7.2% was the cutoff. For the high-

Table 4. *Indices as diagnostic indicators of error as tested with MBP*

Residues or atoms with the worst 10% indices and worst 10% discrepancies between the two MBP stuctures were identified. True positives were those belonging to both groups, false negatives were those with discrepancies in the worst 10% that were not identified by the index, and false positives were those incorrectly flagged by the index whose discrepancy was actually within the best 70%. For the combined r^{ED}/B-factor tests, the 10% limit was changed to 6.5% (omit-map) of 7.2% (MAD maps).

			1.8 Å MAD map			2.3 Å σ_A-weighted SA-omit map		
			Grouping atoms in each residue (%)		Individual	Grouping atoms in each residue (%)		Individual
Method	*B* factors		Main chain	All atoms	atoms (%)	Main chain	All atoms	atoms (%)
R^{ED}	Set to 20 Å^2	true +	6.6	5.3	6.5	5.3	4.0	4.1
		true −	86.3	85.0	86.5	85.0	83.7	84.1
		false +	0.4	1.8	1.4	2.6	2.2	4.2
		false −	3.5	4.8	3.5	4.8	6.2	5.9
$R^{residue}$		true +	6.2	6.2	N/A	4.8	4.0	N/A
(*O*)		true −	85.9	85.9	N/A	84.6	83.7	N/A
		false +	1.3	0.4	N/A	2.6	1.8	N/A
		false −	4.0	4.0	N/A	5.3	6.2	N/A
r^{ED}		true +	7.0	5.7	6.3	5.3	2.6	4.5
		true −	86.8	85.5	86.3	85.0	82.4	84.5
		false +	0.9	1.8	1.4	2.6	3.1	2.9
		false −	3.1	4.4	3.7	4.8	7.5	5.5
r^{ED}	Set to 20 Å^2	true +	6.6	6.6	6.6	5.7	4.4	4.7
		true −	86.3	86.3	86.6	85.5	84.1	84.7
		false +	0.9	0.9	1.2	2.6	1.8	2.8
		false −	3.5	3.5	3.4	4.4	5.7	5.3
$r^{residue}$		true +	6.6	5.7	N/A	4.4	2.6	N/A
(*O*)		true −	86.3	85.5	N/A	84.1	82.4	N/A
		false +	1.3	1.3	N/A	2.6	2.6	N/A
		false −	3.5	4.4	N/A	5.7	7.5	N/A
B factor		true +				4.4	4.4	6.1
		true −				84.1	84.1	86.1
		false +				2.6	3.5	2.2
		false −				5.7	5.7	3.9
			MAD map					
			Individual atoms (%) 2.3 Å	Individual atoms (%) 1.8 Å				
B factor or r^{ED}	Set to 20 Å^2	true+	6.8	6.9				6.1
		true −	86.6	86.9				86.1
		false +	1.6	1.5				2.0
		false −	3.2	3.1				3.9

quality MAD maps (at either 2.3 or 1.8 Å resolution), the combined diagnostic is an improvement over either alone (Table 4). With the poorer omit map, the combined diagnostic appears equivalent to the use of *B* factors alone. These statistics might underestimate the contribution of r^{ED} to determining model quality, because our estimate of error, the difference between homologous structures, does not distinguish between experimental error and imprecision resulting from disorder. Relative to *B* factors alone, the combined index will better highlight atoms with greatest experimental error, to which r^{ED} is most sensitive.

4. Conclusions

New methods have been derived for the calculation of real-space correlation coefficients that are improved indicators of model quality. As an indicator of the relative precision of all residues/atoms (accurately and inaccurately positioned), r^{ED} is consistently the best parameter with all types of map tested, although the improvement over previous indicators (Jones & Kjeldgaard, 1996; Jones *et al.*, 1991) is sometimes modest. r^{ED} is a useful indicator of precision not only in entire residues, but in individual atoms.

The correlation coefficients obtained here (−0.56 to −0.86 depending on map) are commensurate with the 0.64 obtained by Carson *et al.* (1994) using 2 Å maps. As the r.m.s. difference between the models tested here (0.66 Å) is $< \frac{1}{4}$ that tested in the earlier work (2.8 Å), these statistics indicate that the new indices are more sensitive to subtle errors.

The detection of the largest errors in highly refined structures at sub-atomic resolution still remains challenging. In the detection of residues with worst main-chain error, r^{ED} is superior to prior real-space indices or *B* factors. Detection of residues with worst overall error is possible with maps of exceptional quality, but although r^{ED} is the single best indicator, they are not sufficiently reliable with omit maps. It is more reliable to highlight individual atoms with the combination of lowest r^{ED} or highest *B* factor. Using these indicators,

Table 5. *Indices as diagnostic indicators of error as tested with α-amylase inhibitor*

The tests are similar to those in Table 4, except that less stringent criteria for erroneous atoms were used as indicated in the table, because of the large coodinate errors of the models. This table shows that MIR maps arenot of sufficient quality to detect subtle errors in the relatively good 0.6 Å model, but that over $\frac{3}{4}$ of the worst errors can be detected in crude (unrefined) models typically built into an MIR map and represented by the 1.4 Å model. α-Amylase inhibitor: 2.5 Å MIR map, *B*'s set to 20 Å^2

Method	R.m.s. backbone error (Å)		Grouping atoms in each residue (%) Error criteria	Main chain	All atoms	Individual atoms (%) Error criteria	
r^{ED}	0.6	true +	Worst 10%;	2.7	2.7	Worse than 1.0 Å	7.9
	0.6	true −	errors: 0.9 Å (main chain),	81.1	81.1	(worst 20%)	71.6
	0.6	false +	1.3 Å (all atoms)	4.1	5.4		7.5
	0.6	false −		8.1	8.1		11.1
r^{ED}	1.43	true +	Worse than 1.5 Å	28.4	37.8	Worse than 1.5 Å	31.8
	1.43	true −	worst 37% (main chain),	52.7	37.8	(worst 58%)	36.3
	1.43	false +	50% (all atoms)	9.5	12.2		22.6
	1.43	false −		9.5	12.2		9.3

60–70% of the worst residues/atoms can be detected (depending on map quality). Success rates of 40–90% were reported by Carson *et al.* (1994), but in easier tests to diagnose gross errors (>1 Å) in 2 Å refinements. The 60–70% detection is an improvement on the 40% of residues detected by prior correlation indices against an omit map, but the new indices still leave 30–40% escaping detection. Other stereochemical indicators (Jones *et al.*, 1991; Laskowski *et al.*, 1993; Tronrud & Ten Eyck, 1992) should also be used, but it is likely that no combination of these can be guaranteed to highlight all errors at sub-atomic resolution.

Relative to prior treatments, the indices are diagnosing small errors. The tenth percentile used in the MBP tests corresponded to 0.46, 0.87, and 0.72 Å for main-chain, all-residue and individual atom errors, respectively. These are more subtle errors than the gross errors (such as sequence misalignment) for which real-space *R* factors were originally advocated (Jones *et al.*, 1991). Smaller errors can also be detected, especially with high-quality maps as can be seen by the correlation of r^{ED} and error, even for errors of a few tenths of an Å (Fig. 2).

What types of errors lead to the worst r^{ED}? For the 12 MBP residues successfully flagged for overall error, five were from the termini known to be disordered (Burling *et al.*, 1996) and seven had side chains that were built in conformations that were different (and presumably incorrect) in the 2.3 Å structure as compared to the 1.8 Å structure. The *B* factors were not consistently the very highest, but were well above average. Interestingly, it was presumptive model errors that were detected rather than sites with reported multiple side-chain conformers (Burling *et al.*, 1996).

Finally, the improved correlation of the Ho-MBP r^{ED} for the MAD map compared to the simulated-annealing omit map (Table 2, Fig. 2) is an additional testament to the value of high-quality experimental phases. Even with a high-quality phasing model (R = 18%, 2.3 Å) (Weis *et al.*, 1991), and the use of state-of-the-art omit-map methodology (Bhat & Cohen, 1984; Hodel *et al.*, 1992) the map calculated with model phases is a considerably less reliable indicator of where model improvements are required than is the map calculated with anomalous dispersion phases (Burling *et al.*, 1996; Hendrickson, 1991).

We are grateful to Temple Burling and Axel Brünger for offering use of the Yb-MBP coordinates and diffraction data prior to release by the Protein Data Bank. This work was generously supported by the National Science Foundation (MSC; BIR9418741). The new indices have been programmed into *RSRef*, distributed from http://www.sb.fsu.edu/~rsref.

References

Adams, P. D., Pannu, N. S., Read, R. J. & Brünger, A. T. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 5018–5023.
Bhat, T. N. & Cohen, G. H. (1984). *J. Appl. Cryst.* **17**, 244–248.
Brändén, C.-I. & Jones, T. A. (1990). *Nature (London)*, **343**, 687–689.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Brünger, A. T. (1997). *Methods Enzymol.* **277**, 366–396.
Burling, F. T., Weis, W. I., Flaherty, K. M. & Brünger, A. (1996). *Science*, **271**, 72–77.
Carson, M., Buckner, T. W., Yang, Z., Narayana, S. V. L. & Bugg, C. E. (1994). *Acta Cryst.* D**50**, 900–909.
Chapman, M. S. (1995). *Acta Cryst.* A**51**, 69–80.
Chapman, M. S. & Rossmann, M. G. (1996). *Acta Cryst.* D**52**, 129–142.
Diamond, R. (1971). *Acta Cryst.* A**27**, 436–452.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Hodel A., Kim, S.-H. & Brünger, A. T. (1992). *Acta Cryst.* A**48**, 851–858.
Jones, A. T. & Kjeldgaard, M. (1996). *O, The Manual* 5.10 edition, Uppsala University, Uppsala, Sweden.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kim, S., Smith, T. J., Chapman, M. S., Rossman, M. G., Pevear, D. C., Dutko, F. J., Felock, P. J., Diana, G. D. & McKinlay, M. A. (1989). *J. Mol. Biol.* **210**, 91–111.
Kleywegt, G. J. & Jones, A. T. (1996). *Acta Cryst.* D**52**, 829–832.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Luzzati, V. (1952). *Acta Cryst.* **5**, 802.
Luzzati, V. (1953). *Acta Cryst.* **6**, 142–152.
Pflugrath, J. W., Wiegand, G., Huber, R. & Vertesey, L. (1989). *J. Mol. Biol.* **189**, 383–386.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Rice, L. M. & Brünger, A. T. (1994). *Proteins Struct. Funct. Genet.* **19**, 277–290.
Sheldrick, G. M. (1996). Macromolecular Crystallography Computing School, Western Washington University, USA.
Spiegel, M. R. (1975). *Probability and Statistics*. New York: McGraw-Hill.
Ten Eyck, L. F. (1996). Macromolecular Crystallography Computing School, Western Washington University, USA.
Tronrud, D. E. & Ten Eyck, L. F. (1992). *TNT Refinement Package*, Release 5-A, University of Oregon, Eugene, Oregon, USA.
Tsao, J., Chapman, M. S., Agbandje, M., Keller, W., Smith, K., Wu, H., Luo, M., Smith, T. J., Rossmann, M. G., Compans, R. W. & Parrish, C. (1991). *Science*, **251**, 1456–1464.
Tsao, J., Chapman, M. S., Wu, H., Agbandje, M., Keller, W. & Rossman, M. G. (1992). *Acta Cryst.* B**48**, 75–88.
Weis, W. I., Kahn, R., Fourme, R., Drickamer, K. & Hendrickson, W. A. (1991). *Science*, **254**, 1608–1615.
Wodak, S., Pontius, J., Vaguine, A. & Richelle, J. (1996). Macromolecular Crystallography Computing School, University of Western Washington, USA.
Xie, Q. & Chapman, M. S. (1996). *J. Mol. Biol.* **264**, 497–520.