Acta Cryst. (1994). D**50**, 900–909

Error Detection in Crystallographic Models

BY MIKE CARSON, TIMOTHY W. BUCKNER, ZI YANG AND STHANAM V. L. NARAYANA

Center for Macromolecular Crystallography, University of Alabama at Birmingham, Birmingham, AL 35294, USA

AND CHARLES E. BUGG

BioCryst Pharmaceuticals, Inc., 2190 Parkway Lake Drive, Birmingham, AL 35244, USA

(*Received* 13 *October* 1993; *accepted* 5 *May* 1994)

Abstract

A variety of criteria were tested for identifying errors in protein crystal coordinates. Statistical analysis was based on comparisons of a highly refined crystal structure and several preliminary models derived from molecular replacement. A protocol employing temperature factors, real-space fit residuals, geometric strains, dihedral angles and shifts from the previous refinement cycle is developed. These results are generally applicable to the detection of errors in partially refined protein crystal structures.

1. Introduction

Brändén & Jones (1990) note that crystallographic analysis is both objective and subjective. Modern crystallographic software such as *X-PLOR* (Brünger, 1992*a*) and *O* (Jones, Zou, Cowan & Kjeldgaard, 1991) contain assorted tools for more objective analysis. Guss, Bartunik & Freeman (1992) have studied the effects of the refinement strategy on the accuracy and precision of coordinates. We investigate various methods to assess the validity of a protein crystal structure. The metrics are B factor, real-space fit (omit map and $2F_o - F_c$), sliding R-window, R free, geometric strain energy, φ/ψ 'energy', ω-angle, χ-angle 'energy', r.m.s. shift in refinement, deviation from a database, exposed surface and three-dimensional folding profile.

Our test case is the structure of the serine protease factor D determined by MIR methods. The crystal structure was refined to 2.0 Å to an R factor of 0.188 (Narayana *et al.*, 1994). It is compared to models built by homology using methods similar to Greer's (1990), which 'solved' the structure by molecular-replacement methods and refined with *X-PLOR* (with no human intervention) to reasonable R factors and geometries. In order to better understand the errors that might arise from the use of homology models, we investigated the differences between the structure of factor D and the original and refined homology models. Details of the model building, refinements and differences are presented in the accompanying paper (Carson, Bugg, DeLucas & Narayana, 1994).

The accuracy of models constructed from homology is of fundamental concern when the models cannot be confirmed by experimental methods. An experienced modeler can generate a structure that would be deemed correct on geometric grounds. Programs such as *PROCHECK* (Morris, Macarthur, Hutchinson & Thornton, 1992) or *GEOM* (Cohen, 1993) which assess all geometric features cannot adequately establish the validity of a model. Empirical diffraction data is also required.

There is also concern about the general validity of refined crystal structures obtained through molecular-replacement techniques employing homology models. There is always the question of model bias when no empirically determined phases are available. A procedure to reliably identify suspect regions of the structure is needed.

We take the refined crystal structure of factor D to be correct, and use the deviations of the models as error functions. Correlations of various criteria against the errors are evaluated. Statistical analysis suggests a linear model of five variables: temperature factor, real-space fit, geometric strain, dihedral-angle value and shift from the previous refinement cycle. A protocol to identify model errors based on these crystallographic, energetic and geometric grounds is presented. Grossly incorrect residues are identified with approximately 90% accuracy.

2. Methods

2.1. *Models*

Homology modeling created a full-atom model of human blood complement factor D based on serine protease structures deposited in the Brookhaven Protein Data Bank (PDB) (Bernstein *et al.*, 1977). A manual model built using techniques based on a presentation by Greer (1988) gave a preliminary solution of the crystal structure *via* molecular replacement. The crystallographic refinement of factor D to 2.0 Å based on MIR data is described by Narayana *et al.* (1994). Model phases were used only to locate heavy-atom sites. This refined crystal structure is of high quality; the coordinates are thus assumed to be correct and provide the

basis for the comparisons that follow. Factor D is a single polypeptide chain of 228 residues which crystallizes in the triclinic space group P_1 with two independent molecules per asymmetric unit. The final coordinates of the two independent factor D monomers are referred to as 'FDA' and 'FDB', for the *A* and *B* chains.

An automated model was created by the *Protein Design* module of *Quanta* (Polygen Molecular Simulations, Incorporated, 200 Fifth Avenue, Waltham, MA 02254) for comparison. The initial homology models are referred to as 'FDM' for the manually (M) constructed and 'FDQ' for the *Quanta*-generated (Q) coordinates. Two copies (*A*, *B*) of these models were placed in the unit cell and subjected to simulated-annealing (SA) refinement (Brünger, Kuriyan & Karplus, 1987) with *X-PLOR* (X), followed by individual atomic temperature factor refinement against the 2.0 Å data set of 23 249 reflections. This produced the models FDMAX, FDMBX, FDQAX and FDQBX for the two independent monomers refined from each starting homology model. The original molecular-replacement model was subjected to SA refinement as new native data sets became available. The monomers in this model, now subjected to five iterations (I), are referred to as FDIAX and FDIBX.

The modeling and refinement procedures that created the six model coordinate sets above are fully documented in the accompanying paper (Carson *et al.*, 1994). Each partially refined model set (FDMX, FDQX, FDIX) has roughly the same deviation between sets as deviation from the crystal structure. The average root-mean-square (r.m.s.) deviations of these sets from the crystal structure is 1.2 Å for main-chain and 2.8 Å for side-chain atoms. However, these deviations are far from uniformly distributed. About 3/4 of the main chains and 1/2 of the side chains are considered to be correct within experimental error. About 1/7 of the main chain has deviations greater than 1.0 Å (see Tables 5 and 6 of the accompanying paper). Most of these gross errors are localized in key substrate-binding loops.

There are some significant conformational differences between the independent monomers in the crystal structure. About half of the gross errors in the models occur

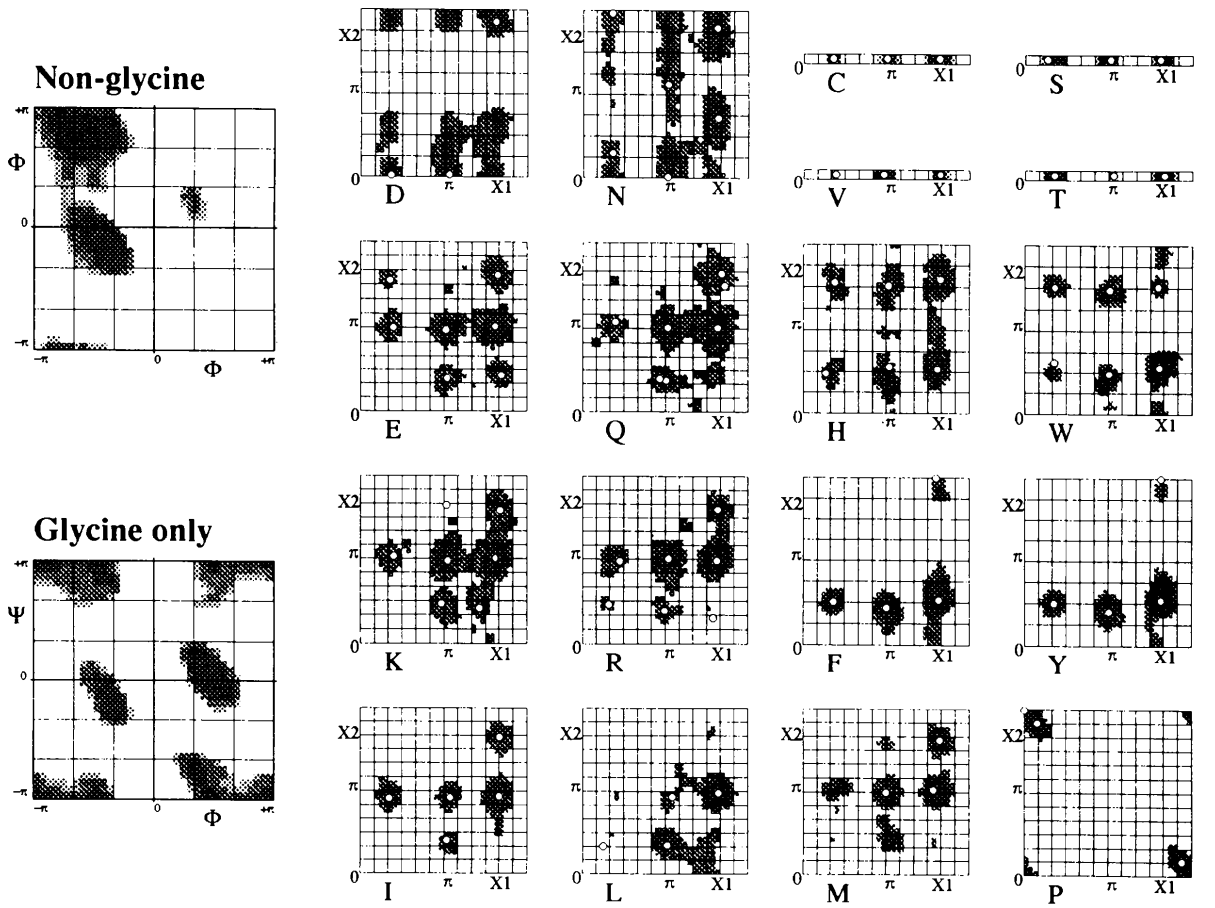

Fig. 1. Observed dihedral preferences. The frequency of occurrence in the PDB in grids of 10° increment has been tabulated, and the grids have been smoothed with a Gaussian filter. The shading levels (light to dark) represent probabilities greater than 0.5, 2.0 and 5.0 standard deviation units above the mean. Data for all main-chain dihedrals (excluding glycine) and glycine only are shown, as well as side-chain dihedrals for each amino acid with at least one dihedral. The rotamer library of Ponder & Richards is shown on the side-chain dihedral plots as open circles.

in these regions. We adopted the criterion that a model agrees if the r.m.s. deviation limit is within twice the error suggested by the Luzzati plots (see accompanying paper). We take the portions where the superimposed FDA and FDB models agree within this limit to represent a restricted set of data with highest confidence. We also use the entire crystal structure for error analysis. Analysis of the crystal structure with the methodology developed here reveals that the FDB monomer fits the experimental data better, in particular the residue range 41–48. Comparisons with only the FDB subunit provide another test set of data.

2.2. *Error evaluation criteria*

A variety of criteria of structural correctness suggested by other authors were applied to the models above. The large number of models allows for some statistical analysis.

2.2.1. *Conventional R factors and Luzzati plots.* The standard crystallographic R factor ($\Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$) is a widely used indication of structure quality. Luzzati (1952) proposed a method for estimating coordinate errors by analysis of plots of R factor as a function of resolution, assuming all errors are explained by the error in coordinate positions. Both methods give a single overall value for the structure.

It has been noted that R factors alone are a rather poor indicator of structure quality (Brändén & Jones, 1990). It is generally conceded that Luzzati plots underestimate the coordinate error; alternative methods have been proposed (Read, 1986).

2.2.2. *Real-space R factor per residue.* The real-space fit residual (Jones *et al.*, 1991) measures agreement between an electron-density map and the atomic model. For each residue, all or a subset of the atoms are converted into a 'map' by treating each atom as a Gaussian density function with a given temperature factor. The model map is compared to the observed map over the grid points around the atoms. This R value ranges from 0.0 (good) to 1.0 and is generally about 0.25 for correct structures, but may exceed 0.50 for flexible side chains extending into the solvent. Incorrectly folded structures are identified by runs of residues with high R factors.

The real-space fit is usually evaluated with maps having Fourier coefficients of $2F_o - F_c$ and phases calculated from the final model. This evaluation criterion is referred to as rsr. This R factor may also be calculated from *OMITMAPS* (Bhat & Cohen, 1984) in an attempt to reduce the model bias. This evaluation criterion is referred to as orsr. These rsr methods were calculated with the program *O* (Jones *et al.*, 1991).

A complementary approach is to calculate the R factor with a sliding 'window' of omitted residues (Rao, 1991). Each of five successive residues are omitted from the model for structure-factor calculations. If a given region has significant errors, the R factor will improve with the window omitted.

An *X-PLOR* script was written to execute this procedure. An overall fast Fourier transform (FFT) scale factor and R factor are determined for use in all subsequent calculations. The R factor is then calculated for each omitted window of residues. The difference from the overall R factor is taken and normalized by the mass of the atoms in the window. This index is then negated, so that poor regions have high values. This evaluation criterion is referred to as rwin.

The windowing method can also be applied to a test set of reflections, which have been omitted from the SA refinement using the free-R value protocol recently advocated by Brünger (1992*b*). This evaluation criterion is referred to as rfree.

2.2.3. *Motion and temperature factor per residue.* Experience has shown that correctly positioned model coordinates will not shift significantly, even upon heating to 4000 K, with the *X-PLOR* SA protocol. This suggests that coordinate shifts during a refinement cycle may be an appropriate error function. An *X-PLOR* script is used to compute r.m.s. differences in atomic coordinates (Å) between two models on a per-residue basis, taking into account the symmetry of Asp, Glu, Tyr and Phe residues. This is a useful tool for describing the differences between two structures. This evaluation criterion is referred to as rms.

It is general knowledge that higher atomic temperature factors (B factors) correlate with exposed surface area, and that high B factors may indicate error, as well as their intended purpose to model thermal disorder. An *X-PLOR* script is used to compute r.m.s. values of the B factors (Å^2) on a per-residue basis. This evaluation criterion is referred to as bf. Another script determines the accessible surface area (Å^2) (Lee & Richards, 1971). This evaluation criterion is referred to as surf.

2.2.4. *Preferred conformations.* A model can be evaluated by examining how well it conforms to expected torsion-angle values. The freely rotatable dihedral angles of the protein backbone (φ/ψ) which define the secondary structure are known to have favored values. This is explained on energetic grounds from the study of model peptides (Ramachandran, Ramakrishnan & Sasisekharan, 1963), and leads to the popular representation commonly known as the Ramachandran plot.

The freely rotatable dihedral angles of protein side chains (χ's) are also expected to have favored values on energetic grounds (*e.g.* staggered *versus* eclipsed conformations). An investigation of highly refined structures (with no dihedral constraints applied) from the PDB has established a 'rotamer library' of side-chain conformations in proteins (Ponder & Richards, 1987).

We used the rotamer library to select the 'best' structures from the PDB for a subsequent analysis. A structure was selected if 80% of the side chains with

dihedrals matched one of the preferred rotamers. Of the slightly more than 1000 independent protein chains in the PDB at the time, 382 were selected. Using this set of protein structures, φ/ψ and $\chi 1/\chi 2$ values were mapped to 36 × 36 element (10° interval) grids creating distributions for each amino-acid type. These grids give empirical Ramachandran-like plots presented in Fig. 1.

These distributions are used to assign a single value for main-chain or side-chain dihedrals. The grids are converted to standard deviation units. The value for an observed dihedral is determined from bilinear interpolation of the grid. To plot these single dihedral values, we adopt the following convention: if the grid value is 0.0 (no observations), it is set to −1.0; if the grid value is greater than 1.0, it is clamped to 1.0. Finally, the dihedral value is set to 1.0 minus the grid value. This dihedral value then has a range from 0.0 (very good) to 2.0. This gives a mock potential-energy function with steep walls for the disallowed regions. For non-glycine main-chain dihedrals, lookup is based on the combined table. For glycine, the best value of the combined table and the glycine table is chosen. For side-chain dihedrals, there is a table for each amino acid. This evaluation criterion is referred to as dihe.

Jones & Thirup (1986) have shown that proteins may be built from fragments of unrelated proteins, based on initial C_α positions. A preliminary presentation of the program *O* (Jones, Bergdoll & Kjeldgaard, 1989) outlined several methods for locating potential errors in a model. These ideas were implemented into *Atom*, our local version of *FRODO* (Jones, 1978). Each run of five consecutive C_α are fitted against a library of 62 highly refined protein structures. The top ten matches, based on r.m.s. deviations of C_α positions, are saved. The r.m.s. of the ten deviations is computed. A low value is generally obtained for common secondary-structure types. A high value indicates an unusual conformation or possible error in the structure. In a similar fashion, the position of the carbonyl O atom at the middle of the five-residue run is compared with the corresponding O atom in the ten best fits. A deviation of over 2.0 Å suggests that this peptide plane is incorrectly modeled and the carbonyl should be flipped. The evaluation criterion based on the structural database is referred to as db.

2.2.5. *Potential energy and correctness of folding.* In a classic study (Novotny, Bruccoleri & Karplus, 1984), proteins were deliberately misfolded and energy minimized. One could not distinguish the correct structure from the incorrect structure from the potential-energy functions alone. Thus the commonly used potential energy functions would not be a sufficient test of correctness.

Eisenberg and coworkers (Bowie, Lüthy & Eisenberg, 1991; Lüthy *et al.*, 1992) recently published a method to identify misfolded structures based on '3D-profiles.' These are computed from observed statistics involving secondary structure and solvent-accessible surface of the side chains, obtained from analysis of structures in the PDB. A value is computed for each residue, then a smoothing window of 21 residues applied. This score averages about 0.4 for correctly folded structures, but will fall below zero in incorrect structures. This evaluation criterion is referred to as 3D.

In *PROLSQ* (Hendrickson, 1980) and *X-PLOR*, the minimization function seeks to balance the contributions from the structure factors and the ideal geometry (*PROLSQ*) or potential energy (*X-PLOR*). The average deviation in bonds or angles for the entire structure is often given as an indicator of quality. Brünger has proposed that the geometric strain energy in a residue correlates with the error in the structure. This evaluation criterion is referred to as geom.

2.3. *Error models*

We take the refined crystal structure to be correct. Any difference between a model structure and the crystal coordinates is regarded as error. These errors are tabulated on a per-residue basis, using all atoms, main chain only, or side chain only. The residues are partitioned into two sets: the entire molecule and a restricted set of highest confidence (see *Models*). Two measures of the error function are explored, denoted as rms and del-dihe.

The r.m.s. difference in coordinates between the crystal structure and a given model is determined as previously described. The difference between two structures may also be expressed in terms of dihedral angles. The differences in φ/ψ or $\chi 1/\chi 2$ pairs are computed as euclidian distances in radians. For example, the main-chain del-dihe is $[(\Delta\varphi)^2 + (\Delta\psi)^2]^{1/2}$.

The relationships between the error functions and the error evaluation criteria are monitored with the standard correlation coefficient. For the data sets a and b,

$$\text{correl}(a,b) = \langle ab - \langle a\rangle\langle b\rangle\rangle / (\langle a_2 - \langle a\rangle^2\rangle\langle b^2 - \langle b\rangle^2\rangle)^{1/2}.$$

A general linear model was constructed to assess the independence of the various criteria. The error model has the form,

$$\text{Error} = a_0 + a_1\text{Criterion}_1 + a_2\text{Criterion}_2 + \ldots + a_n\text{Criterion}_n.$$

The a_i are the scalar coefficients.

3. Results

3.1. *R factors and Luzzati plots*

The final R factors and estimated coordinate errors for the models are given in Table 4 of the accompanying paper. The experimental crystal structure clearly provides the best model, but the differences between it and the homology models are not dramatic. All estimated errors are less than 0.32 Å.

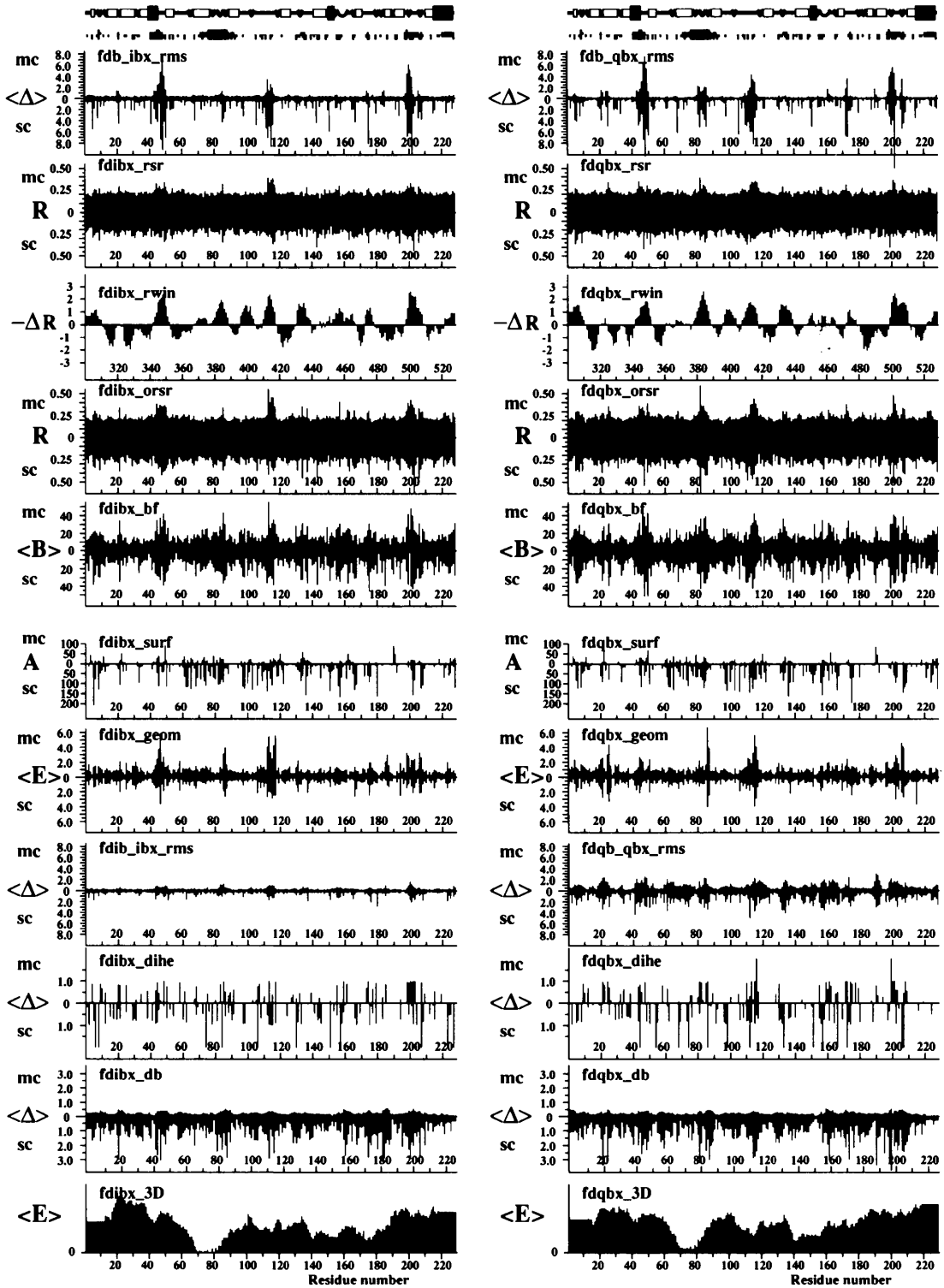

Fig. 2. Model error *versus* criteria. The data for FDIBX (left) and FDQBX (right) are shown. The top of each figure has the same two schematic insets. The first shows the secondary structure observed for factor D with helices as thick filled rectangles, sheets as open rectangles, turns as 'u' curves, and coils as lines. The second shows the r.m.s. deviations between superimposed FDA and FDB as three levels with a zero value defining the restricted set of data < 0.46 Å, the small value being < 1.5 Å, and the large value representing deviations > 1.5 Å. The first per-residue plot is the deviation of the model with the crystal structure (taken as the error measurement). The remaining plots are each criterion as discussed in the text. All except 3D have separate data for main-chain (mc) and side-chain (sc) atoms. For the db criterion, the side chain refers to the carbonyl.

3.2. *Correlation of errors with various criteria*

Fig. 2 presents plots of the criteria on a per-residue basis for the FDIBX and FDQBX models. The top graph of each figure is the observed r.m.s. difference between the crystal structure and the model. This is taken as the measure of error for each residue. Inset are diagrams of the secondary structure and of the regions where FDA and FDB differ.

Table 1 lists the correlation coefficient between each criterion and the r.m.s. coordinate error per residue for FDIBX and FDQBX. Fig. 3 presents the mean values of the correlation coefficients and their standard deviations calculated with the six models. Both Table 1 and Fig. 3 present the values for the data set with all residues and for the restricted data set (only those residues that are the same in both FDA and FDB).

The temperature factor generally has the greatest correlation with model error. This is true for both test sets of data. The correlation with surface area is higher using all the data, as many of the differences between the data sets involve surface residues. The correlation between bf and surf for the side chains of the various data sets averaged about 0.5. The bf criterion is chosen as a test for error.

Comparing the measures of real-space fit, those using *OMITMAP* have a slightly, but not significantly, higher value than those based on $2F_o - F_c$ maps. (The *OMITMAP* procedure produced an average phase-angle change of only about 12° for the models.) The *R*-window

Table 1. *Correlations of deviations to criteria*

The correlation coefficients ($\times 10^3$) are given for the criteria discussed in *Methods*. The results are for the data plotted in Fig. 2. For geom, rsr's, bf, surf and rms, data were compared against the respective data for all atoms, main-chain (mc) atoms and side-chain (sc) atoms. The remaining criteria are not so divided. FDIBX (I) and FDQBX (Q) data for all 228 residues are given in the upper portion of the table. The lower portion of the table is for the 124 residues (all), 175 residues (mc), and 120 residues (sc) where the FDB subunit superimposes within 0.46 Å upon FDA.

Criteria	I.all	Q.all	I.mc	Q.mc	I.sc	Q.sc
geom	605	570	467	416	582	546
rsr	667	570	645	615	544	530
orsr	669	600	638	592	573	559
rwin	446	464	536	539	455	459
bf	692	692	596	673	616	665
surf	431	437	181	298	411	405
rms	580	516	649	473	504	392
ω	−65	−3	−81	−92	−73	−9
φ/ψ	368	384	432	458	351	371
$\chi 1,2$	366	442	29	136	369	459
db-Cα	376	336	410	420	363	311
db-O	275	339	382	487	244	342
3D	−19	55	−109	−23	−36	55
geom	559	579	498	610	529	441
rsr	646	583	683	698	518	553
orsr	645	687	671	709	556	602
rwin	279	295	413	417	273	280
bf	673	686	565	681	562	635
surf	227	338	289	412	236	266
rms	644	446	682	356	612	251
ω	−43	−13	−131	−86	−80	−22
φ/ψ	256	276	294	382	189	225
$\chi 1,2$	525	513	111	224	543	586
db-Cα	183	271	398	372	139	184
db-O	120	265	388	532	80	241
3D	79	120	48	74	81	169

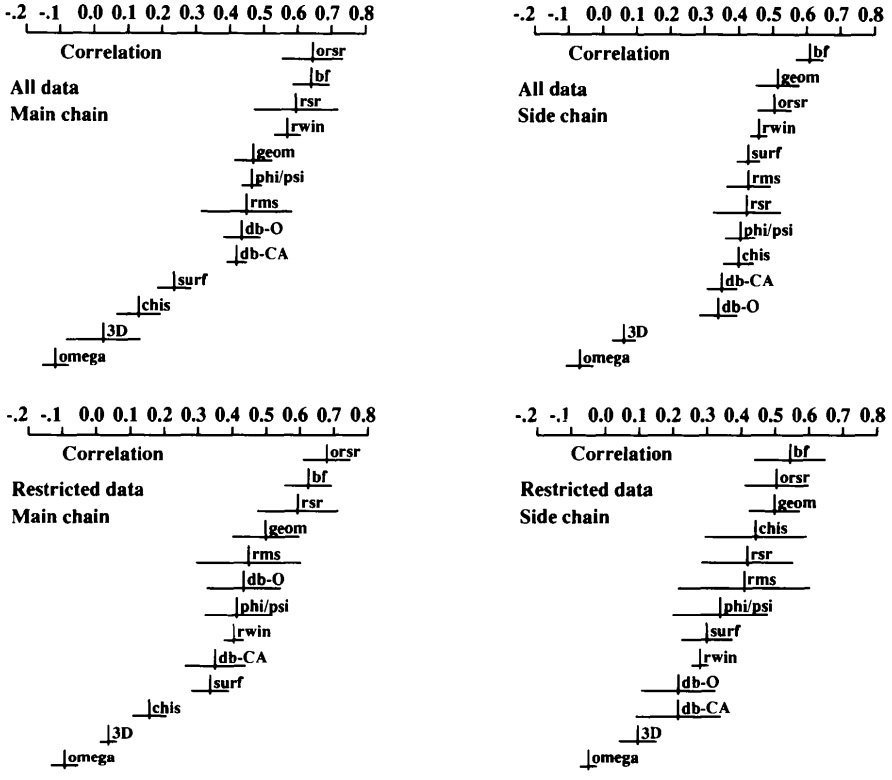

Fig. 3. Correlation of errors with criteria. The correlation coefficient of the deviation error with each criterion is shown. The vertical bar gives the average over each of the six models, and the horizontal bar plots $\pm 1.0\sigma$. All data involve correlations over all 228 residues. Restricted data involve correlations over only the 175 main-chain and 120 side-chain residues of the FDB subunit that superimpose within 0.46 Å upon FDA.

Table 2. *Average correlation between criteria*

The averages over the six models of the correlation coefficients ($\times 10^3$) of each final criterion with one another are shown. The upper half of the table is for all 228 residues. The lower half is for the restricted data, as in Table 1. The upper-right triangle gives main-chain–main-chain values and the lower-left triangle gives side-chain–side-chain values.

MC\SC	bf	rsr	geom	rms	dihe
bf	—	329	361	383	280
rsr	563	—	261	151	139
geom	412	432	—	329	426
rms	421	344	339	—	297
dihe	407	380	461	297	—

MC\SC	bf	rsr	geom	rms	dihe
bf	—	196	203	317	167
rsr	367	—	169	147	55
geom	334	335	—	351	425
rms	357	224	312	—	278
dihe	290	269	361	223	—

method gives similar correlations considering all the data, but drops significantly when evaluated against the restricted data. The free-R value evaluated in the same fashion (data not shown) gave a slightly worse correlation than the R-window method. (A random 10% of the reflections was used as the free-R test set.) These sliding R-factor methods are computationally expensive. The real-space fit residual based on the $2F_o - F_c$ maps, rsr, is adopted as the best criterion in this group, due to the computational simplicity of the process.

The geometric strain energy is the only other criterion with a correlation coefficient greater that 0.5. The r.m.s. shift shows the highest correlation with the iteratively refined FDI model. This would be the more common situation. The FDM and FDQ models had only been subjected to one cycle of SA refinement, so they have not had a chance to 'settle in'. (The r.m.s. differences based on a subsequent SA run on each gave significantly higher correlations.) Both these criteria, geom and rms, have been adopted for use.

Considering main-chain atoms, the φ/ψ dihedral values, C_α fit to database and C=O fit to database show correlations of about 0.5 with all the data. These values all drop (the first two significantly) when compared to the restricted data. This implies Ramachandran plot agreement alone may not be a sufficient criterion. Comparison of the carbonyl direction to the database appears to be a slightly, but not significantly, better measure of these three. However, this measure may give false negatives when glycines or prolines are involved. There is essentially no correlation with the side-chain dihedrals.

Considering side-chain atoms, there is a fair correlation (about 0.35) with the main-chain values cited above considering all the data. However, these correlations fall dramatically when using only the restricted set. The correlation with the $\chi 1/\chi 2$ dihedral values increases to 0.45 with the restricted data. As the dihedral criterion, dihe, works well for both the main chain and side chain and involves only a table lookup, it is adopted for subsequent analysis.

A comparison of the tables reveals, somewhat surprisingly, that the peptide plane ω angle is anti-correlated with the error. This may be a consequence of our using special '*PROLSQ*'-like potentials for the peptide planes. This results in substantially more planar peptides than the default *X-PLOR* parameters. The three-dimensional folding test is not correlated, but this method is primarily for the identification of grossly misfit models. These criteria will not be considered further.

While the del-dihe method is a good way to monitor differences between structures (see accompanying paper), the r.m.s. measure is generally better as an error function. The dihe and db criteria correlations with del-dihe were all higher than their correlations with r.m.s. The del-dihe error function is significantly less correlated with most of the evaluation criteria. This was especially true for the side-chain correlations. The r.m.s. deviation is thus used as the error measure for the remainder of the paper.

3.3. *Correlation of the various criteria to each other*

The following five criteria previously described were adopted on the basis of their correlation to error in the model: bf, rsr, geom, rms and dihe. These may all be applied separately to the main-chain and the side-chain atoms of each residue. These criteria were correlated with one another (Table 2) to determine their independence.

Table 2 reveals positive correlation between all the criteria, but these correlations are much less than the correlations of the criteria with the error. The only value greater than 0.5 was for bf to rsr, computed over all data. The average criteria–criteria correlation is 0.28 for the restricted data. The average deviation–criteria correlation is 0.49 for the same data.

3.4. *Combination of criteria*

To further test the independence of these five criteria, linear models were constructed and subjected to singular value decomposition tests. The routine 'svdfit' of Press, Teukolsky, Vetterling & Flannery (1992) was employed for multiple-regression analysis. The best linear models were obtained using the log of the r.m.s. deviation as the error function. Individual t tests suggest a high probability that each individual criterion is required for modeling the error (the largest value was $P = 0.0017$ for side-chain r.m.s.). This was true with the analysis based on all six models, on only the three B-chain models, and on the restricted set of the six models.

The individual coefficients vary considerably, though they are all on the same order of magnitude. For example, the coefficient for rsr is about threefold that for r.m.s. with the restricted side-chain data, while they are about equal for the restricted main-chain data. We do not

wish to suggest that the coefficients obtained from this particular study should be used – only that these criteria are independent predictors of error.

We will attempt to combine them to give an overall score. Each correlation and model analysis thus far has been against the raw data. In order to put the disparate data (such as B factor and R factor) on the same scale, they are converted into standard deviation units relative to the mean.

3.5. *Correlation of combined criteria*

The r.m.s. difference between FDB and FDIBX (again, assumed to be the true error) is presented in Fig. 4. The sum of the five criteria and each individual criterion, each in standard deviation units, are also shown. Data are presented separately for main-chain and side-chain atoms. The sum of the five independent criteria in standard deviation units was chosen as our final error criterion due to its consistency, as explained below.

The desired error-detection method must identify 'incorrect' residues and give no false negatives. An arbitrary testing cutoff must be selected for this criterion. Another arbitrary cutoff is used to select residues in error based on the r.m.s. deviation of the model from the crystal structure. Table 3 gives results for main-chain and side-chain atoms with a variety of error (Å) and criterion (σ) cutoffs. The values are the averages over all 228 residues of the three models corresponding to FDB. The number of incorrect residues are noted, and the percent correctly identified by application of the criterion cutoff.

Table 3. *Identification of gross errors*

The Å error gives the r.m.s. deviation cutoff that defines a gross error. The results are averages of all residues in the three models corresponding to the FDB crystal structure. The σ-cut gives the sum-of-criteria cutoff used to flag errors. The # bad counts residues deviating by more than the given error. The % hits are the bad residues correctly flagged. The # false are correct residues (within twice the Luzzati error) flagged as being in error.

Atoms	Å error	σ-cut	# bad	% hits	# false
mc	1.0	1.00	27	89	3
mc	1.0	0.67	27	92	7
mc	1.5	1.00	21	97	3
sc	1.0	1.00	89	42	0
sc	1.0	0.67	89	53	1
sc	1.5	1.00	74	48	0
sc	1.5	0.67	74	60	1
sc	2.0	0.67	53	69	1
sc	2.5	0.67	39	82	1
sc	3.0	0.67	32	88	1

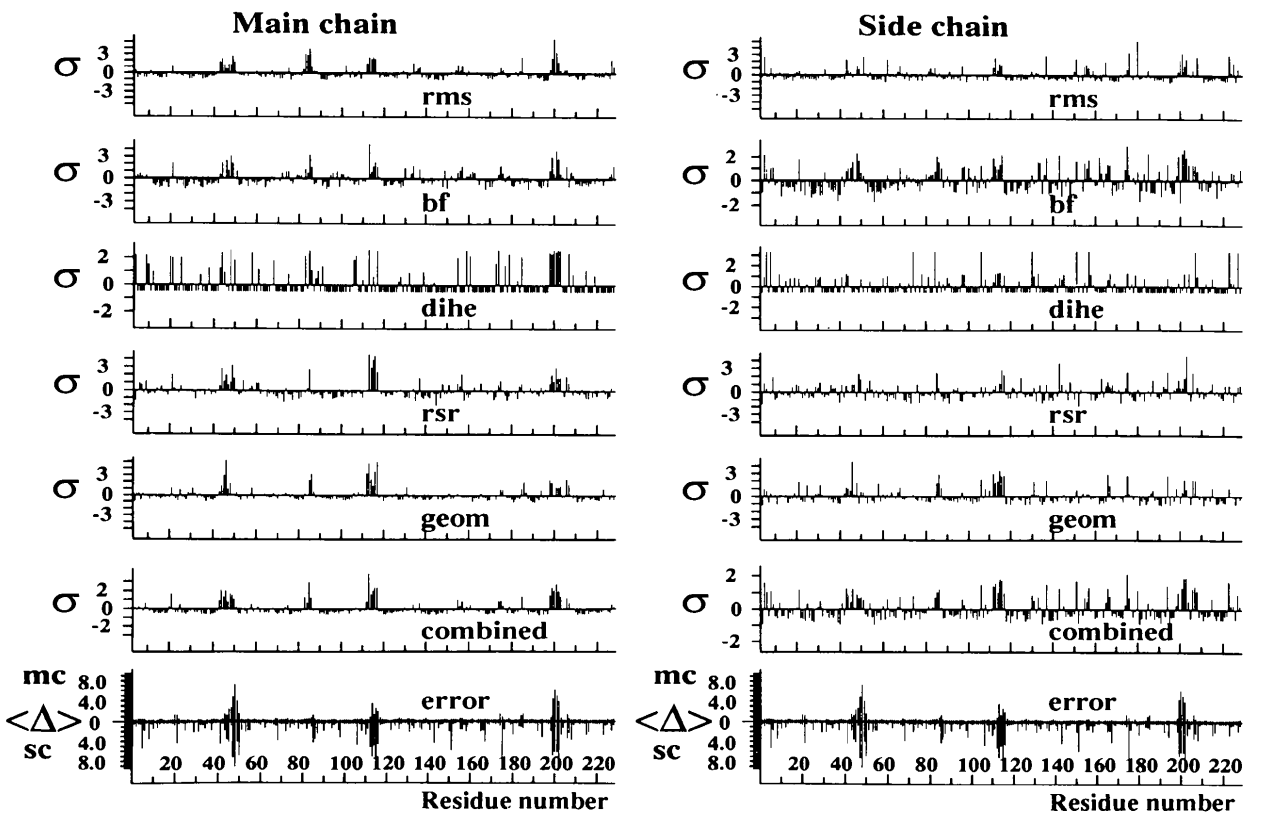

Fig. 4. Model error *versus* final combined criteria. Data for FDIBX are given. The presentation is similar to that of Fig. 2. Here only the best criteria in standard deviation units are given, as well as the sum of these criteria. The sum, cast in standard deviation units, is chosen as the ultimate error-detection function.

The number of false positives count residues flagged as being in error, yet agreeing to the crystal structure within twice the Luzzati error.

These analyses were carried out against each individual criterion and various combinations. The results using the five combined criteria consistently located the greatest percentage of residues in error, and was even more impressive in producing the smallest number of false positives. The method works very well for identifying problems in the main chain. Nearly 10% of the residues in the molecular-replacement models are grossly in error. Using 1.0 Å as the cutoff for errors and 1.0σ for the cutoff for the criterion test, the method identifies nearly 90% of the problem residues. Only about 1% of the protein give a true false negative. An additional 2% of residues having errors greater than twice the Luzzati limit, but less than the cutoff, were also flagged. These are taken as true, but less severe, errors.

These same cutoffs applied to side-chain residues give a less impressive result. Nearly 40% of the residues are in error by the 1.0 Å distance cutoff, but only 42% are identified as such. However, there were no true false negatives reported. Using 2.0 Å as the error cutoff, nearly a quarter of the side chains are grossly in error. With a 0.67σ criteria cutoff, the method identifies nearly 70% of the problem residues and only one residue as a false positive. An additional 7% of residues with less severe errors are also flagged.

4. Discussion

4.1. *Confidence in models*

The molecular-replacement models had reasonable *R* factors and geometries. However, about 10% of the main chain was grossly in error and almost all of these errors were located in the active-site and substrate-binding regions of factor D. These are precisely the residues that must be known accurately to understand the structure/function relationship in this enzyme. A tool to assess the quality of each residue is needed.

A protocol is presented that locates almost all of these grossly misfit residues, while falsely identifying a minimal number of well fitted residues for our test case. Variants of this protocol have been successfully used in this laboratory for numerous crystal refinements. Here we attempt to formalize the protocol, allowing the crystallographer to assess the confidence in an atomic model on a per-residue basis with a single error criterion.

4.2. *Previous applications*

Our previous work on identifying and illustrating features and potential problems in protein structure (Carson & Bugg, 1988) used essentially the same types of criteria as described here. Ribbon drawings were color-coded by residue *B* factors, degree of fit to density, or a mapping to the Ramachandran plot. Side chains not matching the rotamer library and color coding of the atomic structure by potential energy were shown. These points are discussed further in the description of the graphics program *Ribbons* 2.0 (Carson, 1992). A new feature is the assignment of a single pseudo-energy value to represent the φ/ψ or the $\chi1/\chi2$ dihedral-angle pair, based on the observed distribution in the PDB.

These methods were recently used in the refinement of the crystal structure of peroxynitrite-modified superoxide dismutase (SOD) by Smith *et al.* (1992). Here the PDB structure '2SOD' (Tainer, Getzoff, Beem, Richardson & Richardson, 1982) was used for a trivial molecular-replacement solution. The standard SA protocol refined this model to an *R* factor of 0.216 to 2.5 Å. Our criteria identified 43 potential problems in the 151 residues; 34 residues were manually adjusted to the maps. The subsequent SA step reduced the *R* factor to 0.190 and led to improvement in the criteria. The r.m.s. atomic shift from the original PDB structure was 1.22 Å for all atoms and 0.73 Å for C_α atoms.

4.3. *Overview of the protocol*

The protocol detects errors in partially refined crystallographic structures. A variety of standard tools and methods developed largely by others in the field are employed. These tools, *i.e. X-PLOR* and *FRODO*, are accessible to the majority of macromolecular crystallographers. Assorted utilities of the *Ribbons* package are also required (freely available through ftp).

The protocol requires coordinates of the latest refined structural model (with individual *B* factors) in PDB format, as well as the same structure from the previous round of model building or refinement. Suitable electron-density maps in the *FRODO* format must be present (these may be made with *X-PLOR*). *X-PLOR* scripts are used to determine average *B* factors, average shifts between the two models, and average geometric strain energy for the main-chain and side-chain atoms on a per-residue basis. *Ribbons* utilities compute the real-space fit residual and dihedral angle probabilities for each residue. The five criteria are then converted to standard deviations and averaged, giving a single goodness-of-fit value for the main-chain and side-chain atoms in each residue.

The recent implementation of the real-space fit as part of the *Ribbons* package has produced some interesting results. Here, the user is expected to produce both a $2F_o - F_c$ and an F_c map in units of e Å^{-3}. The summation is then performed over all grid points within 2.2 Å of any atom of interest. The implementation in *O* inputs only the first map, and calculates the model map on the fly by adding atomic Gaussian densities (with a uniform temperature factor) to the grid.

The correlations of rsr with r.m.s. error now approach 0.7 for both main chain and side chain with the new implementation, significantly higher than before. How-

ever, the correlation of rsr to B factor has also increased significantly, as might be expected given the inclusion of individual B factors in the map calculations. This may influence the independence of these two variables in the final error function, but these two are the most correlated with error. The new rsr implementation in effect gives B factor a higher weight. One could always calculate the F_c map with an overall temperature factor to avoid this bias.

Sevcik, Hill & Dauter (1993) have proposed the 'discriminator' for each atom as its temperature factor divided by its electron density in the final $2F_o - F_c$ map, or B (Å^2)/e (Å^{-3}). They monitor this function on a per-residue basis to assess quality. Their method is consistent with the results presented here. A referee offered the gut opinion that it all comes down to B factors and difference maps. This is basically true. The rsr described above is the best single criterion, but it is not as powerful in identifying errors as is the full protocol.

4.4. *Future work*

We plan to accumulate partially refined models from a variety of crystal structures, analyze them by the methods presented, and monitor the deviations from the final result. This will provide the data for a neural network program to 'learn' to recognize errors in macromolecular crystal structure.

We have shown here that the incorrect residues can largely be identified. Fig. 9 of the accompanying paper shows the FDIBX molecular-replacement model with the FDB crystal structure and the computed OMITMAPS based only on the model and the native diffraction data. It shows weak disconnected density for the current model, and a parallel stretch of similar density several ångströms away where the atoms should be. Knowing that the model is grossly in error at that point would provide the impetus to make major changes in these residues. The unanswered question is whether the correct structure could have been attained without resorting to MIR methods and many rounds of manual refitting on graphics. The *Ribbons++* program under development will seek to provide answers.

4.5. *Program availability*

The *Ribbons* utilities are freely available *via* anonymous ftp to xtal.cmc.uab.edu. The code is written in the C language and produces ASCII and PostScript output. UNIX versions for Silicon Graphics and Evans & Sutherland workstations as well as a VAX/VMS version are available. A UNIX version of Bhat's *OMITMAP* program is also available (see Carson, 1991). The graphics display program *Ribbons* 2.5 is available at nominal charge to academics (contact carson@luna.cmc.uab.edu).

We gratefully acknowledge NASA grant NAGW-813 and Public Health Sevice grant AI32949 for support.

References

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F. JR, BRICE, M. D., ROGERS, J. R., KENNARD, O., SHIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.

BHAT, T. N. & COHEN, G. H. (1984). *J. Appl. Cryst.* **17**, 244–248.

BOWIE, J. U., LÜTHY, R. & EISENBERG, D. (1992). *Science*, **253**, 164–170.

BRÄNDÉN, C. & JONES, A. T. (1990). *Nature (London)*, **343**, 687–689.

BRÜNGER, A. T (1992*a*). *X-PLOR*. Version 3.1. *A System for X-ray Crystallography and NMR*. Yale Univ., New Haven, CT, USA.

BRÜNGER, A. T (1992*b*). *Nature (London)*, **355**, 472–474.

BRÜNGER, A. T., KURIYAN, J. & KARPLUS, M. (1987). *Science*, **235**, 458–460.

CARSON, M. (1991). *J. Appl. Cryst.* **24**, 958–961.

CARSON, M. (1992). *J. Appl. Cryst.* **25**, 327–328.

CARSON, M. & BUGG, C. E. (1988). Am. Crystallogr. Assoc. Meet., Abstract G2.

CARSON, M, BUGG, C. E., DELUCAS, L. J. & NARAYANA, S. V. L. (1994). *Acta Cryst.* D**50**, 889-898.

COHEN, G. H. (1993). *J. Appl. Cryst.* **26**, 495.

GREER, J. (1988). Am. Crystallogr. Assoc. Meet., Abstract I3.

GREER, J. (1990). *Proteins*, **7**, 317–334.

GUSS, J. M., BARTUNIK, H. D. & FREEMAN, H. C. (1992). *Acta Cryst.* B**48**, 790–811.

HENDRICKSON, W. A. (1980). *Proceedings of the Daresbury Study Weekend*, November 15–16, Daresbury, England, edited by P. A. MACHIN, J. W. CAMPBELL & M. ELDER, pp. 1–8.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

JONES, T. A., BERGDOLL, M. & KJELDGAARD, M. (1989). *O. A Macromolecule Modeling Environment*. In *Crystallographic and Modeling Methods in Molecular Design*, edited by C. E. BUGG & S. E. EALICK, pp. 189–199. New York: Springer-Verlag.

JONES, T. A. & THIRUP, S. (1986). *EMBO J.* **5**, 819–822.

JONES, T. A., ZOU, J.- Y., COWAN, S. W. & KJELDGAARD, M. (1991). *Acta Cryst.* A**47**, 110–119.

LEE, B. & RICHARDS, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.

LÜTHY, R., BOWIE, J. U. & EISENBERG, D. (1992). *Nature (London)*, **356**, 83–85.

LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.

MORRIS, A. L., MACARTHUR, M. W., HUTCHINSON, E. G. & THORNTON, J. (1992). *Proteins*, **12**, 345–364.

NARAYANA, S. V. L., CARSON, M., EL-KABBANI, O., KILPATRICK, J. M., MOORE, D., CHEN, X., BUGG, C. E., VOLANAKIS, J. E. & DELUCAS, L. J. (1994). *J. Mol. Biol.* **235**, 695–708.

NOVOTNY, J., BRUCCOLERI, R. & KARPLUS, M. (1984). *J. Mol. Biol.* **177**, 787–818.

PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T. & FLANNERY, B. P. (1992). *Numerical Recipes in C. The Art of Scientific Computing*. New York: Cambridge Univ. Press.

PONDER, J. W. & RICHARDS, F. M. (1987). *J. Mol. Biol.* **193**, 775–791.

RAMACHANDRAN, G. N., RAMAKRISHNAN, C. & SASISEKHARAN, V. (1963). *J. Mol. Biol.* **7**, 95–99.

RAO, J. K. M. (1991). Am. Crystallogr. Assoc. Meet., p.35, Abstract G07.

READ, R. (1986). *Acta Cryst.* A**42**, 140–149.

SEVCIK, J., HILL, C. P., DAUTER, Z. & WILSON, K. S. (1993). *Acta Cryst.* D**49**, 257–271.

SMITH, C., CARSON, M., VAN DER WOERD, M., CHEN, J., ISCHIROPOLOUS, H., & BECKMAN, J. (1992). *Arch. Biochem. Biophys.* **299**, 350–355.

TAINER, J. A., GETZOFF, E. D., BEEM, K. M., RICHARDSON, J. S. & RICHARDSON, D. C. (1982). *J. Mol. Biol.* **160**, 181–217.