Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Errors and reproducibility in electron-density map interpretation

Sherry L. Mowbray,[a]* Charlotte Helgstrand,[b] Jill A. Sigrell,[b] Alexander D. Cameron[b] and T. Alwyn Jones[b]

[a]Department of Molecular Biology, Swedish University of Agricultural Sciences, Biomedical Centre, Box 590, S-751 24 Uppsala, Sweden, and [b]Department of Molecular Biology, Uppsala University, Biomedical Centre, Box 590, S-751 24 Uppsala, Sweden

Correspondence e-mail: mowbray@xray.bmc.uu.se

Received 30 October 1998
Accepted 12 April 1999

Three investigators, with varying levels of experience, independently built and refined the structure of *Escherichia coli* ribokinase at 2.6 Å resolution. At the end of the refinement/rebuilding processes the models had essentially converged, although each had its own particular pattern of remaining errors. The subsequent refinement of the same structure at 1.8 Å resolution allowed an overall quality check of each of the lower resolution models, and an analysis of which graphics-based tools were generally most efficient in locating these errors. Criteria which are useful in the application of Ramachandran, main-chain and side-chain database and real-space fit analyses are presented.

1. Introduction

Like fine furniture, protein structures used to be lovingly built and patiently polished by master craftsman before meeting the eyes of the world. By contrast, many new crystallographic structures are being built and refined by inexperienced people with varying degrees of expert guidance. The repetitive but demanding nature of refinement, the pressure to obtain the final product and the use of often sub-optimal crystallographic data can combine to result in errors by any investigator. The types of errors which occur vary in frequency and severity (Brändén & Jones, 1990; Kleywegt & Jones, 1997*a*). Fortunately, it seems to be rare for completely wrong models to be published and deposited in the Protein Data Bank (Bernstein *et al.*, 1977). Less dramatic errors, such as the sequence being out of register with the electron density in a portion of the structure, frequently occur during the early stages of model building, but are usually corrected during the refinement process. It is expected that few models with such errors are found in the current release of the PDB, although stereochemical and hydrogen-bonding considerations suggest that small changes might be appropriate in many deposited structures (Hooft *et al.*, 1996).

As when the large-scale production of furniture replaced the individual construction of each piece, it has been necessary to analyse and reproduce the key elements of the process of protein model-building, as well as to establish quality control. No-one wishes to discover (or much worse, have someone else discover) errors in their structures after deposition, nor do they wish that biologists be misled in the conclusions drawn from them. The development of methods which prevent and/or detect errors, both during refinement and in the final product, has thus become an important area of research. One fundamental lesson gleaned from the study of high-resolution structures is that proteins closely follow the stereochemical

rules previously established for small molecules. The use of main-chain and side-chain databases represents one attempt at forcing the application of these rules during protein structural refinement in general (Jones *et al.*, 1991; Zou & Mowbray, 1994). The benefits to be gained are largest during low-resolution refinement, where the data are not adequate to fully resolve side-chain and main-chain conformations. The application of such tools in model refinement to some extent invalidates their use as a criterion for the final structure quality, but the gains in model accuracy outweigh the loss in their value as an indicator. It has also become clear that there is a real need for simple methods which can be applied simultaneously with the rebuilding process, as well as for statistics supporting the utility of these methods in practice.

In the present paper, we describe a study in which three investigators with varying levels of experience each produced a refined structure of the same protein at medium resolution. Each person worked with the same 3.0 Å MIR map and 2.6 Å reflection data, but solved, built and refined the structure independently. We demonstrate that the refinements converged to give similar structures, all of which were very similar to the final structure refined to 1.8 Å resolution. The various models were then used to assess a number of methods for evaluating model quality. As an analysis of the results suggested which tools that are easy to use at the graphics terminal during rebuilding sessions are most likely to be applied, we had a particular interest in determining the most efficient tools for use in *O* (Kleywegt & Jones, 1997*a*).

2. Methodology

The real-space correlation coefficient (RSCC), peptide orientation analysis (pepflip) score and rotamer side-chain fit (RSC) for each residue were obtained in *O* (Jones *et al.*, 1991; Jones & Kjeldgaard, 1997*a*). In calculations of the RSCC, c is a constant chosen to remove systematic differences between F_o and F_c, and A_o is a 'zero-temperature radius' (Deisenhofer & Steigemann, 1975). Values of $c = 0.82$ and $A_o = 0.90$ were used for the 2.6 Å maps, and $c = 1.04$ and $A_o = 0.90$ were used for the MIR and 1.84 Å maps; temperature factors were set to 20.0 Å^2 for all atoms. The RSC calculations made use of a newer rotamer library (Kleywegt & Jones, 1998). The latest versions of *O* include on-the-fly graphing and Ramachandran display options.

Uppsala Software Factory programs (http://alpha2.bmc.uu.se/~gerard/usf/) were used for much of the analysis. *MOLEMAN*2 (Kleywegt, 1995) was applied to obtain Ramachandran plots and to calculate per-residue temperature factors. *LSQMAN* (Kleywegt, 1996; Kleywegt & Jones, 1997*b*) was used to calculate overall r.m.s. differences for C$^\alpha$ or side-chain atoms, as well as to fix the nomenclature for Glu, Asp, Tyr and Phe residues, to ensure that the correct atoms were compared. *LSQMAN* was also utilized to calculate differences in the φ and ψ angles and the circular variance of the χ_1 and χ_2 values of pairs of structures, where the circular variance of two observations of the dihedral angle χ is defined as

$$1.0 - 0.5\{[\textstyle\sum \cos(\Delta\chi)]^2 + [\sum \sin(\Delta\chi)]^2\}^{1/2}.$$

The circular variance lies between 0 and 1, with lower values indicating greater similarity between the observations of the angle. Calculation of per-residue values for side-chain r.m.s. differences (including the C$^\beta$ atom) between different models combined use of *LSQMAN* and the program *RMSPDB* (G. Kleywegt, unpublished work). All figures were generated with *CA-Cricket Graph* III (Computer Associates International, Inc.); the same program was used for calculation of most residue-based statistics.

3. Building and refinement of the models

The subject of the present study was the *Escherichia coli* ribokinase, a 309-residue enzyme which phosphorylates ribose at O5 in the presence of ATP and Mg^{2+}. Ribokinase was chosen because it represented an MIR solution of a protein from a previously uncharacterized structural family. The native data were collected from a crystal containing ribose and nucleotide and were 97% complete to 2.6 Å resolution, with an $R_{\rm merge}$ of 6.9% and an average $I/\sigma(I)$ of 20.3 (Sigrell *et al.*, 1997). The MIR map was calculated at 3.0 Å resolution using two derivatives, a selenomethionine-substituted protein and a mercury-acetate-soaked native. The MIR phases were improved using solvent flattening and histogram matching (Sigrell *et al.*, 1998); the map calculated based on the resulting phases is referred to here as the 'original' MIR map for convenience.

Initial models were built independently from this map by three different researchers. One (*A*) had solved and refined MIR structures previously. The other two (*B* and *C*) were novices with a basic knowledge of protein structures and of the tools which could be used to build and analyse them obtained through local courses and suggested reading. Attempts were made to minimize interaction during the whole process, but the total absence of any type of communication was not an attainable goal. The initial models (designated as model 0 in each case) were built from skeletons with the *Proleg* libraries of main-chain and side-chain conformations found in *O* (Jones & Thirup, 1986; Jones & Kjeldgaard, 1997*b*). *A* and *B* built simultaneously without comparing the results; after the first building cycle, it was confirmed only that the same connectivity had been obtained before proceeding. Builder *C* built at a later date, but knew only that the map could be interpreted, not what the structure looked like.

Subsequently, each investigator was given the same reflection data set to 2.6 Å resolution and proceeded with refinement in *X-PLOR* (Brünger, 1992). Each refinement consisted of a similar Powell minimization, slowcool, Powell minimization and group temperature-factor refinement (the latter being usual in our laboratories at this resolution). All used the same randomly chosen subset of reflections (10%) in calculating $R_{\rm free}$. Each investigator used a starting temperature of 2000 K for the slowcool and chose W_A based on an optimization of $R_{\rm free}$. Each carried out three macro-cycles including inspection and rebuilding at the graphics terminal, followed by

Table 1
Analysis of various 2.6 Å models and comparisons to the final high-resolution structure.

Model	Number of C^{α} atoms (% of structure omitted)	R.m.s. bond, angle deviations (Å, °)	R.m.s.d. C^{α}/ side-chain atoms to final (Å)	Number (%) C^{α} atoms near final position†	Mean pepflip (% in model >2.0 Å)	Mean RSC score (% in model >1.0 Å)	$\langle\lvert\Delta\varphi\rvert\rangle$, $\langle\lvert\Delta\psi\rvert\rangle$ from final structure	Torsion-angle variance to final χ_1, χ_2	Mean RSCC to $2F_o - F_c$ map (% with poor fit)‡	Mean RSCC to MIR map
*A*0	307 (0.6)	—	0.81/1.87	172 (56)	0.92 (11)	0.37 (7.8)	25.0, 24.0	—	0.60 (18)	0.60
*A*1	307 (0.6)	0.006, 1.41	0.53/1.68	281 (92)	0.74 (7.2)	0.45 (14)	14.3, 13.1	—	0.80 (6.5)	0.69
*A*2	307 (0.6)	0.006, 1.30	0.39/1.55	292 (95)	0.71 (5.2)	0.41 (11)	11.5, 10.5	—	0.81 (4.9)	0.69
*A*3	307 (0.6)	0.006, 1.31	0.34/1.40	300 (98)	0.70 (4.6)	0.40 (11)	9.8, 9.1	0.06, 0.13	0.81 (5.5)	0.70
*B*0	304 (1.6)	—	1.29/2.50	152 (50)	1.22 (16)	0.64 (29)	39.5, 37.8	—	0.62 (38)	0.62
*B*1	286 (7.4)	0.011, 2.12	0.95/2.14	244 (80)	0.93 (11)	0.52 (18)	20.1, 19.9	—	0.80 (8.0)	0.69
*B*2	306 (1.0)	0.010, 2.01	0.90/2.11	270 (88)	0.85 (8.8)	0.55 (18)	17.0, 17.0	—	0.81 (9.8)	0.68
*B*3	301 (2.6)	0.011, 1.63	0.30/1.43	298 (97)	0.68 (4.3)	0.50 (15)	9.6, 8.9	0.08, 0.17	0.81 (4.7)	0.70
*C*0	294 (4.9)	—	1.02/2.17	176 (58)	1.16 (16)	0.43 (10)	34.4, 35.2	—	0.66 (26)	0.66
*C*1	285 (7.8)	0.008, 1.57	0.85/1.96	251 (82)	0.85 (7.7)	0.44 (13)	15.6, 16.2	—	0.80 (11)	0.71
*C*2	297 (3.9)	0.007, 1.35	0.32/1.53	287 (94)	0.69 (4.4)	0.44 (14)	9.8, 8.3	—	0.84 (3.3)	0.72
*C*3	299 (3.2)	0.019, 1.98	0.25/1.12	294 (96)	0.70 (5.0)	0.48 (12)	8.6, 7.6	0.05, 0.15	0.85 (1.7)	0.72
Final, 1.84 Å	306 (1.0)	0.011, 1.03	—	—	0.66 (4.2)	0.39 (9)	—		0.93 (0.7)	0.72

† Calculated using *LSQMAN*, using a 1.8 Å cutoff for initial structures, 0.8 Å for other structures. ‡ Using a cutoff of 0.5 for the MIR map, 0.7 for the 2.6 Å $2F_o - F_c$ maps and 0.8 for the final 1.84 Å $2F_o - F_c$ map.

crystallographic refinement (group temperature factors were reset to 20 $Å^2$ at the beginning of each cycle); the models, as output by *X-PLOR*, are designated further with the descriptors 1–3. In *A*'s models, residue 91 was correctly built as a lysine, while in those of *B* and *C* this residue was built as an arginine owing to a book-keeping error. None of the models had water or ligands at the completion of the third refinement cycle, although appropriate electron density was present in the active site and elsewhere.

The primary criterion used in rebuilding was the visual assessment of fit to the electron density of $2F_o - F_c$ maps. Builder *A* habitually checked hydrogen bonding and *B* and *C* also did so, but less consistently. As a result of a casual conversation with *B*, builder *C* decided to omit regions which were not clear in the electron density, while builders *A* and *B* had generally included best guesses even for poorer regions. Builder *A* did not use the programs available in Uppsala for finding structural problems (Kleywegt & Jones, 1996*a*, 1997*a*), with the explicit intention of testing the relative benefits of the tools *versus* experience. Builders *B* and *C* knew about these tools and were told to use them, but were not forced to do so. It became clear at the end of the refinements that neither had actually applied the tools. This was an important lesson and spurred us to think about what sorts of tools *would* get used. Some of these ideas have been incorporated into the most recent version of *O*.

4. The three series of structures converged to each other and to the final model

The progress of R and $R_{\rm free}$ during the three refinements is shown in Fig. 1. In each case, $R_{\rm free}$ after the third refinement cycle was 33–34% and the conventional R was approximately 28%. The largest drop in $R_{\rm free}$ was during refinement cycle 1 in all cases. The r.m.s. bond and angle deviations for each model are reported in Table 1. The final 2.6 Å models were not considered to be exhaustively refined, but rather to have converged, based on their R factors. The r.m.s. difference between their C^{α} positions was 0.25–0.35 Å, as low as or lower than our expected level of coordinate error.

At this point, the various 2.6 Å resolution models were compared to a later model of the same protein refined at 1.84 Å resolution, which included ribose, nucleotide, two phosphates and 181 water molecules (Sigrell *et al.*, 1998). The corresponding X-ray data were 95.5% complete, with an $R_{\rm merge}$ of 5.5% and an average $I/\sigma(I)$ of 30.2. This high-resolution model (entry code 1rkd in the Protein Data Bank), which was thoroughly checked for hydrogen bonding and other problems, is referred to as the 'final' one in the following

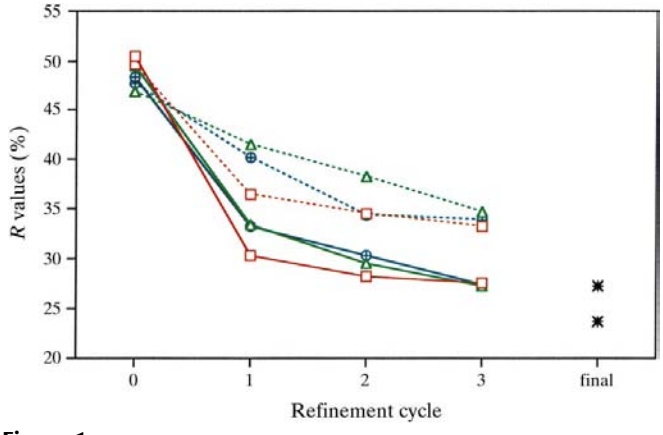

Figure 1
Progress of R (solid lines) and $R_{\rm free}$ (broken lines) for the three 2.6 Å resolution refinement series. Values for the models of builders *A* (red squares), *B* (green triangles) and *C* (blue circles) are shown at the different stages of refinement, along with those for the final 1.84 Å model (black stars).

discussion; the corresponding R and R_{free} were 22.0 and 25.8%, respectively. Fig. 2 shows that all of the final 2.6 Å models had an r.m.s. difference on C^{α} position of 0.25–0.35 Å from the high-resolution structure (see Table 1 for earlier models), while their r.m.s. difference from their respective starting C^{α} positions ranged from 0.8 to 1.3 Å. Thus, not only were the final 2.6 Å models very similar to each other, they were all very similar to the high-resolution structure.

Based on numerous distributions of differences in C^{α} position, we have used a cutoff of 0.8 Å to define portions of each 2.6 Å structure which are significantly different from the final model. By this criterion, roughly half of the residues were correctly placed in the first models to be built, while nearly all were correct in the third models (Table 1). The largest changes from the original structures occurred during refinement cycle 1 in each case. In terms of approach to the final structures, the greatest progress was found during refinement 1 for *A*, after refinement 3 for *B* and after refinement 2 for *C*. These particular trends reflect the correction of some parts of models *B* and *C* at these time-points. The effect ofmaking these changes could also be seen in corresponding drops in R_{free} (Fig. 1).

More residues were correctly placed in all *A* models than in the equivalent *B* or *C* models. However, builder *A* included all residues for which plausible density was present, including some portions which did not become well behaved until the ligands were added (*i.e.* after the 2.6 Å refinements described here). As a result, more atoms were included in *A* models, but not all were equally dependable; a few isolated C^{α} positions were misplaced by as much as 2 Å. Builder *B* had similar difficulty. The problematic residues were omitted by *C* (see Table 1); model *C*3 thus contained fewer residues, but all were close to the correct positions.

The temperature factors resulting from group temperature-factor refinement of the lower resolution structures were highly correlated with the average atomic temperature factors of the final high-resolution structure. After accounting for the fact that the overall temperature factor of the high-resolution data set was approximately 6 $Å^2$ higher than that of the lower resolution set, the per-residue temperature factors of the final structure had a correlation coefficient of 0.65–0.75 with those of models 1 and of about 0.80 for all models 3. The practice of resetting group temperature factors to 20 $Å^2$ at the beginning of each cycle of temperature-factor refinement resulted in absolute values which were very similar to those of the high-resolution structure (otherwise residues with higher temperature factors tended to receive ever-higher temperature factors during refinement, despite very similar atomic positions). The temperature factors of the 2.6 Å models were also highly correlated with each other (correlation coefficients 0.86–0.90 for the models 3).

5. Fit to the electron density

Visual fit to the electron density of $2F_o - F_c$ maps was the major criterion used by all builders. The agreement between the coordinates and the corresponding electron-density maps can be described quantitatively for each residue (Jones *et al.*, 1991). As implemented in *O*, this function can take the form of an *R* factor or a real-space correlation coefficient (RSCC). The advantages of the latter have been discussed previously (Zou & Mowbray, 1994) and this form was used in the present analysis. In agreement with the R_{free} results, the average RSCC calculated using the appropriate $2F_o - F_c$ map improved during all refinements (Table 1). The distribution of RSCC values also became sharper and peaked at a higher value as the refinements proceeded (Fig. 3). This is an effect of both increased resolution and improved fit to the electron-density maps. It was also noted that maps calculated using data in the 2.6–7.5 Å resolution range had a similar average RSCC, but a lower median value than the maps shown, which were calculated using data in the 2.6–10 Å resolution range.

Low RSCC values were correlated with errors in C^{α} position. About half of the residues with RSCCs less than 0.7 in the early models had differences in C^{α} position from the final

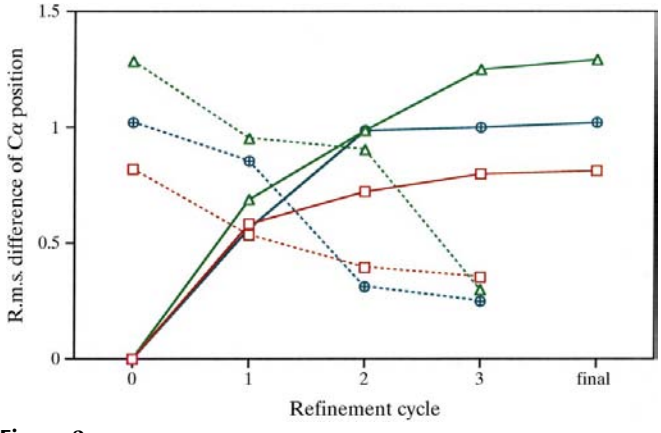

Figure 2
R.m.s. difference of the C^{α} positions for each model to those of the initial one of the series (solid lines) and to the final 1.84 Å structure (broken lines). Values for the models of builders *A* (red squares), *B* (green triangles) and *C* (blue circles) are shown at the different refinement cycles. Values are also given for the final 1.84 Å model.

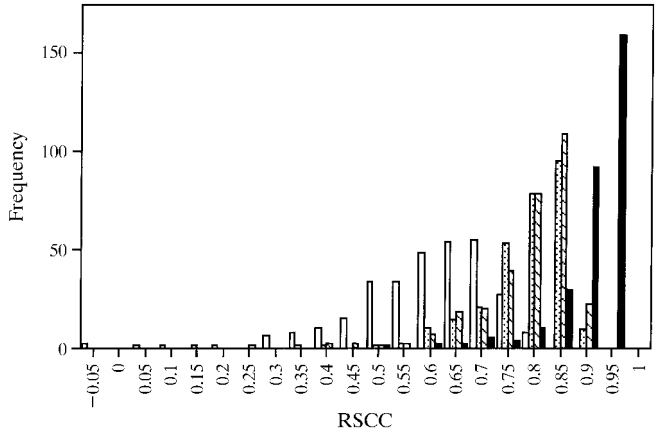

Figure 3
Histogram showing the frequency of different ranges of RSCC values for models *B*0 (open bars), *B*1 (lightly shaded bars), *B*3 (striped bars) and the final model (solid bars), each calculated using the corresponding $2F_o - F_c$ map.

1.8 Å structure which were greater than 0.8 Å (illustrated for model *A*1 in Fig. 4). For example, the residues of the ill-defined loops mentioned above have lower RSCCs than the average in the structure. A few residues have poor RSCC values but are close to their correct locations. In these cases, the density is often broken in the 2.6 Å map.

Temperature factors have been used by many investigators as an indicator of fit to the electron density. It has been reported previously that RSCC values are in better agreement with a visual assessment of electron density than temperature factors (Zou & Mowbray, 1994). The two quantities are inversely correlated; correlation coefficients vary in the range −0.7 to −0.85, with the value decreasing during refinement (Fig. 5).

Like low RSCC values, high temperature factors are frequently associated with main-chain errors. In Fig. 6, temperature factors of *A*1 are plotted against the difference in C^{α} position from the final position. In the present case, a temperature-factor cutoff of approximately 40 Å^2 would have been most useful in finding errors but, as for RSCC, many errors are missed. As for the RSCC (Fig. 4), a less stringent cutoff would find more errors, but at the cost of inspecting a larger number of false negatives.

Residues with high RSCC values most frequently have low temperature factors and are associated with dependable regions of the structures. Where the RSCC and temperature factors are both high, the residues usually have weak density enveloping the side chain. (The most recent versions of *O* make it easier to inspect these cases by allowing the adjustment of the electron-density contour level in real time.) Most atoms with very high temperature factors also have poor RSCC values and weak density which does not match the side chain. In later models, residues with low RCCCs (0.7 for the 2.6 Å maps) represent a range of temperature factors, including some very low ones, and indeed most of these are not discernible errors. We conclude that either indicator is useful in finding problems, but that neither is foolproof. Poor values for either the RSCC or the temperature factors should make the user suspect problems in the model. The cutoff to be used in each case should be determined by inspection of the distribution of values, with the poorest 10% being a reasonable goal for routine inspection.

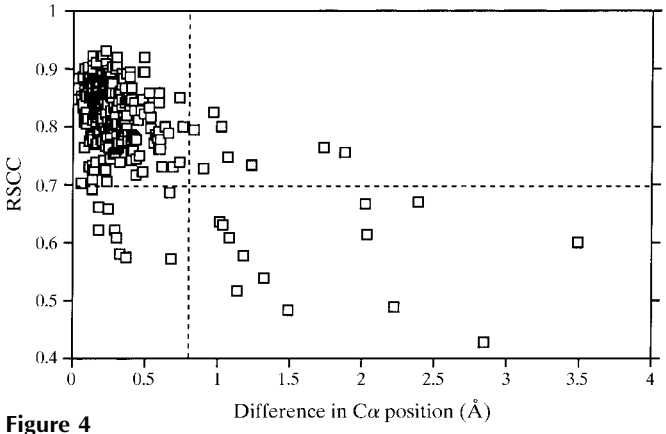

Figure 4
RSCC value, calculated using main-chain atoms (including C^{β}), plotted against the estimated error in C^{α} position for each residue of model *A*1. The values of RSCC (0.7) and the C^{α} difference (0.8 Å) which bound the main distribution are indicated by broken lines.

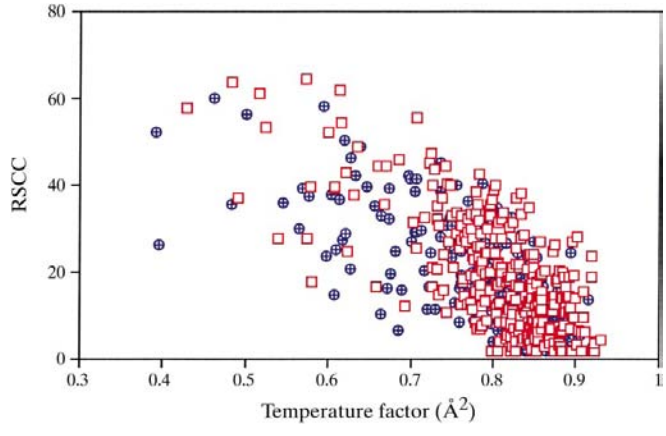

Figure 5
Plot of temperature factors *versus* RSCC for models *A*1 (red squares) and *C*1 (blue circles).

6. Analysis of main-chain structure

Ramachandran plots (Ramakrishnan & Ramachandran, 1965) have been used for many years to describe main-chain conformations and have recently (re)gained popularity as a indicator of structural quality. Useful discussions of the method may be found in Laskowski *et al.* (1993) and Kleywegt & Jones (1996*b*). The latter paper applies much tougher criteria in assigning allowed areas, but demonstrates that 98% of all non-glycine residues in high-resolution structures fall within the established boundaries. Ramachandran plots showed dramatic improvement during the three sets of refinements described here. The number of outliers [defined for the present purpose as all residues, including glycines, outside the region defined by Kleywegt & Jones (1996*b*)]

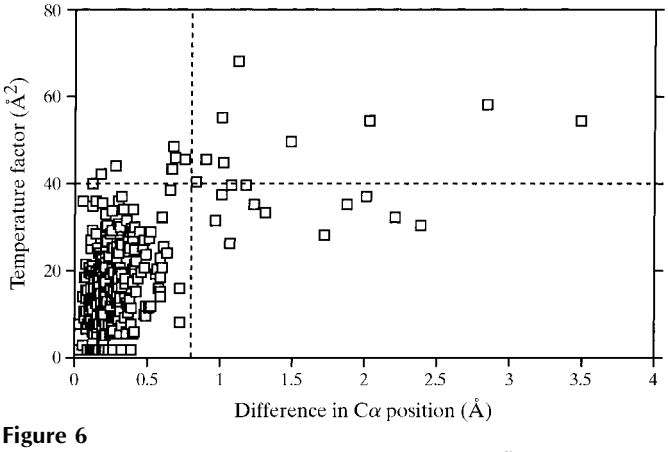

Figure 6
Plot of temperature factors *versus* estimated error in C^{α} position for each residue of model *A*1. The values of the temperature factor (40 Å^2) and the C^{α} difference (0.8 Å) which bound the main distribution are indicated by broken lines.

Table 2
Correlation of Ramachandran outliers with various other properties of the models.

For each property, the percentage of the total number of residues with that characteristic is also given.

Model	Number of outliers† (% of total)	Number of outliers with C^{α} > 0.8 Å from 1.84 Å structure (% of all large differences)	Number of outliers with low RSCC‡ (% of all low RSCCs)	Number of outliers with pepflip > 2.0 (% of all high pepflips)	Number of outliers with RSC > 1.0 (% of all high RSCs)
*A*0	57 (19)	22 (16)	12 (21)	23 (68)	6 (25)
*A*1	32 (10)	15 (58)	15 (75)	15 (68)	6 (14)
*A*2	28 (9)	6 (40)	9 (60)	12 (75)	5 (14)
*A*3	26 (8)	2 (29)	7 (41)	9 (64)	2 (6)
*B*0	111 (37)	51 (34)	45 (39)	28 (58)	35 (39)
*B*1	39 (14)	13 (31)	9 (39)	17 (55)	9 (18)
*B*2	46 (15)	20 (56)	17 (57)	17 (63)	12 (22)
*B*3	22 (7)	0 (0)	7 (50)	6 (46)	5 (11)
*C*0	91 (31)	36 (31)	34 (44)	28 (58)	11 (37)
*C*1	42 (15)	14 (41)	16 (50)	14 (64)	9 (24)
*C*2	22 (7)	2 (20)	3 (30)	7 (54)	3 (7)
*C*3	22 (7)	0 (0)	1 (20)	7 (47)	2 (5)
Final, 1.84 Å	15 (5)	—	0 (0)	6 (46)	2 (7)

† The numbers of outliers includes glycine residues, since errors at these residues are also often outside the 'allowed' regions. ‡ RSCC values, calculated using main-chain atoms, below cutoffs of <0.7 for the $2F_o - F_c$ maps and 0.5 for the MIR map.

decreased sharply after the first few cycles, and approached but did not reach the number observed for the high-resolution structure (Table 2). Much of the latter effect is a consequence of the difference in resolution alone, since higher resolution structures are generally closer to the stereochemical ideal (Kleywegt & Jones, 1996*b*).

Ramachandran outliers were found to be highly correlated with main-chain errors (defined as C^{α} positions greater than 0.8 Å from the final structure), as well as with regions of poor electron-density fit (Table 2). The number of residues with significant differences in C^{α} position or low RSCC values which are also Ramachandran outliers is much greater than would be expected by chance (for example, compare the percentage of values appearing in column 2 of Table 2 with those in columns 3 and 4). Residues in ill-defined regions of the electron density according to a visual inspection were often found among the outliers.

Main-chain characteristics may also be compared using differences in φ and ψ angles. The φ/ψ differences between the 2.6 Å models and the final high-resolution structure decreased during the series of refinements (Table 1). The $\langle|\Delta\varphi|\rangle$ and $\langle|\Delta\psi|\rangle$ values relating the various model 3 structures to the final model are in the range 8–10°, the median value found for 'identical' molecules related by non-crystallographic symmetry (Kleywegt, 1996). The differences between the various final 2.6 Å models are comparable (9–11°), further indicating that these structures had converged to each other.

Both the number of Ramachandran outliers and the average φ/ψ differences started high for model 0 (Tables 1 and 2) and then dropped sharply during the first refinement cycle. Most of this resulted from the process by which the initial models were 'autobuilt' using pentapeptide fragments from the library of main-chain conformations. In this method, only the central three residues of the pentapeptide are actually used for building. Therefore, the (φ,ψ) values within the central residues are the same as in the structure in the database, but the (φ, ψ) values at the ends will be affected by the linkage to the abutting segments. The discontinuities cause deviations from the preferred regions, but because the errors in atomic positions are random and small, they are rapidly fixed by the first round of crystallographic refinement. The effect was more dramatic in *B* and *C*'s models, because *A* explicitly checked that main-chain hydrogen bonding in the initial model was reasonable, rebuilding where necessary prior to refinement. *B* and *C* generally took the 'autobuild' results without question. When residues were repositioned to fix side-chain density at this point, the main-chain conformation was also not rechecked by these builders. In the final analysis, either approach worked.

The *pep_flip* option of *O* (Jones *et al.*, 1991) allows the same library of main-chain conformations to be used in locating unusual (and possibly incorrect) peptide orientations during rebuilding sessions. Pepflip scores vary from close to 0.0 (corresponding to a peptide orientation which is frequently observed) to roughly 3.5 Å (where the peptide O atom points in the opposite direction for all similar main-chain segments in the database). Average pepflip scores decreased during the refinements here, rapidly approaching the value found for the final structure (Table 1). The final mean pepflip score to be expected will vary with the protein. For all α-helical structures, for example, the score will be very low. For structures with little secondary structure, the score should be much higher.

A more informative view of the pepflip results is shown in Fig. 7, where the difference in position of each main-chain carbonyl O atom from that in the final structure is plotted against the pepflip score for that residue. Analysis of a number of distributions suggested that differences in the position of

Table 3
Correlation of residues with pepflip scores >2.0 Å with differences in position of the main-chain O atom from that in the final 1.84 Å structure.

Model	Number of pepflip scores	Number of high pepflip scores	Number of high O-diff., high pepflip (errors found)	Number of high O-diff., low pepflip (errors missed)	Number of low O-diff., high pepflip (false positives)	% of high pepflip scores which indicate errors	% of errors found by pepflip
*A*1	303	20	11	14	9	55	44
*A*3	303	14	5	7	9	36	42
*B*1	272	32	22	45	10	69	32
*B*3	293	13	2	6	11	15	25
*C*1	271	18	11	23	7	61	32
*C*3	287	15	1	7	14	7	14
Final, 1.84 Å	303	13	—	—	—	—	—

the main-chain O atom greater than 0.8 Å are significant. Residues found within the main distribution of pepflip scores (*i.e.* at values less than 2.0 Å) are very likely to have correct main-chain conformation. Although residues with larger pepflip scores in early models are likely to represent errors, a high score does not always indicate that a residue is wrong. Even in a well refined structure, a significant number of such residues will remain (∼4% for ribokinase; ∼3% for the unrelated glucose/galactose-binding protein; Zou & Mowbray, 1994) and they will often represent interesting structural features. For whatever reason, they deserve inspection. A few problematic residues have low pepflip scores, and some high pepflip scores represent residues that should indeed have high scores, but in association with a different main-chain conformation. In the example shown in Fig. 7, five high pepflip scores in the *A*3 model represent true problems, while seven are correct as they stand. In addition, six residues have low scores, but in fact represent differences from the high-resolution structure and so represent problems not located by this method.

The above cutoffs were used to evaluate the various 2.6 Å models (Table 3). During refinement, the number of residues with pepflip scores above the 2.0 Å cutoff decreases. The majority of high pepflip scores in the earliest models indicate main-chain errors (55–69%) and many such main-chain errors can in fact be located in this way (32–44%). High pepflip scores still find an appreciable fraction of the problem areas in the third-round models. For example, high pepflip scores effectively spotlighted *A*'s struggles with a *cis*-peptide and an incorrect *X-PLOR* refinement dictionary, as well as other errors which would otherwise only be found by an analysis of hydrogen bonding. Poor scores in any model were very often associated with poor electron density. However, the omission of poorer portions of the structure in *B* and *C*'s models prevented calculation of pepflip values for adjoining residues (because the first two and last two residues of any segment will not receive a score). The number of residues without pepflip scores becomes more significant as the number of omitted segments increases, and this in itself can be a problem, since a significant number of findable/fixable errors may be found in these neighbouring regions. It should be repeated that the pepflip analysis was not actually used during the 2.6 Å series and the structures would certainly have been improved if it had.

High pepflip scores are correlated with Ramachandran outliers at any stage of the refinement (Table 2). In the high-resolution structure, almost half of all residues with high pepflip scores are also Ramachandran outliers and about half of the Ramachandran outliers also have high pepflip scores. These are not errors in that structure, but are well supported by electron density and the demands of the surrounding protein.

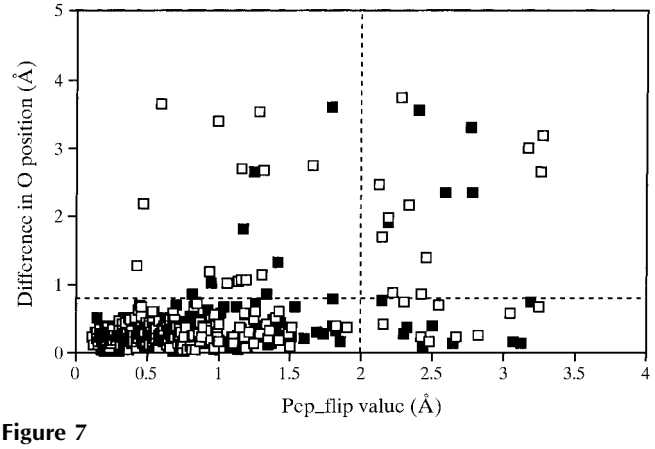

Figure 7
Difference in position of each main-chain carbonyl O atom from that in the final 1.84 Å structure, plotted *versus* the pepflip score for the same residue (open squares for *A*1 and closed squares for *A*3).

7. Analysis of side-chain structure

The average r.m.s. differences between side-chain coordinates of pairs of model 3 structures ranged from 1.2–1.4 Å, with *A* and *C*'s structures being the most similar. The range was nearly identical (1.1–1.4 Å) for the comparison of each model 3 to the high-resolution structure. Both sets of numbers, however, are somewhat biased by the fact that different numbers of atoms are compared in each case (Table 1). Differences between structures may also be assessed using side-chain χ angles. The calculated χ_1 variance between different pairs of models 3 was 0.057–0.091 and the χ_2 variance was 0.136–0.152, again indicating a high degree of similarity. Similar numbers resulted from comparisons of models 3 to the final structure (Table 1).

Table 4
Correlation of residues with RSC values >1.0 Å with r.m.s. differences to the side-chain atoms of the final 1.84 Å structure.

Model	Number of high RSC scores	Number of high r.m.s.d.	Number of high r.m.s.d., high RSC (errors found)	Number of high r.m.s.d., low RSC (errors missed)	Number of high RSC, low r.m.s.d. (false positives)	% of high RSC which indicate errors	% errors found by RSC
*A*3	33	46	17	29	16	52	37
*B*3	44	53	24	29	20	55	45
*C*3	36	45	20	25	16	56	44

Distributions of side-chain r.m.s. differences from the final structure (calculated per residue) are more informative than the overall numbers (Fig. 8). The main peak in each distribution includes r.m.s. differences less than 1.0 Å and primarily represents residues for which both electron density and C^{α} position are good. Plots of side-chain r.m.s. differences *versus* differences in C^{α} position from the final structure confirm that a 1.0 Å cutoff works well as a general criterion of accurate side-chain position. Residues with side-chain r.m.s. differences greater than 2.0 Å were most often associated with errors in the C^{α} position and with low RSCC values. There is a distinct peak for r.m.s. difference values in the 1–2 Å range, particularly for *B* and *C*'s models. These residues are not usually associated with large C^{α} differences and their RSCCs are almost uniformly good. Closer inspection showed that most atoms of these residues are correctly placed, but some are very wrong. In particular, cases where the terminal χ angle of leucine, valine, histidine, asparagine, glutamine and, occasionally, serine residues is 180° away from the correct angle result in an r.m.s. difference in this range. Since the resulting conformation often matched the electron density well, that criterion alone was not adequate to find the problems. Inspection of hydrogen bonding would generally resolve the errors involving polar residues.

Since the final structure is obviously not available during a structure solution or refinement, it is necessary to use methods which can be applied to any coordinate set. The side chains observed in highly refined structures in fact show a high preference for certain conformations, the same ones that would be predicted from energy considerations (Janin *et al.*, 1978). These common side-chain conformations (rotamers) were used here in structure building and rebuilding; the present *O* database includes only those rotamers found with a frequency of at least 5% for that residue type (Kleywegt & Jones, 1998). It has been shown (*e.g.* Zou & Mowbray, 1994) that the use of such conformations in (re)building improves the quality of the resulting models. The side-chain conformations in a given model can also be compared quantitatively to those of the common rotamers using the *rsc_fit* option of *O* (Jones *et al.*, 1991). This RSC (rotamer side-chain) analysis was not used during refinement of any of the models described here. The average RSC score (which includes the zeroes which result for glycine and alanine residues) decreases during refinement (Table 1). High RSC scores are not correlated with Ramachandran outliers (Table 2), except to a small degree in models 0, where it is merely echoing the generally poor stereochemistry.

A typical relationship between the side-chain r.m.s. differences (to the final structure) described above and the RSC score of individual residues is shown in Fig. 9. Residues with RSC scores greater than 1.0 Å are much more frequently associated with significant errors in side-chain conformations than those with smaller RSC scores. Leucine residues with χ_2 180° away from the correct value, for example, usually fall in the 1–1.5 Å RSC range.

Using a cutoff of 1.0 Å for both RSC and side-chain r.m.s. differences, it is possible to estimate the efficiency of the RSC analysis in finding errors (Table 4). RSC scores greater than

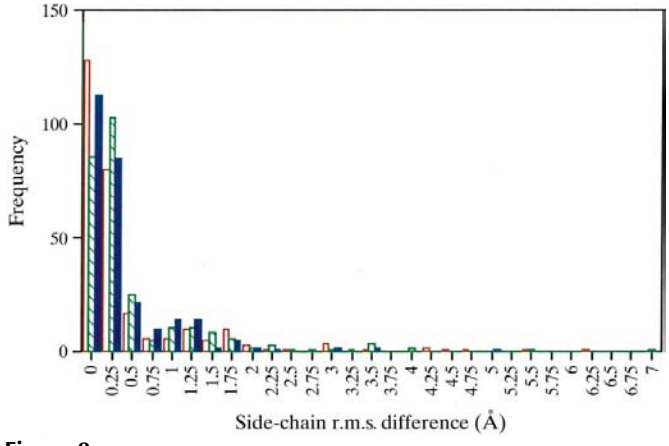

Figure 8
Distributions of side-chain r.m.s. differences for *A*3 (open red bars), *B*3 (striped green bars) and *C*3 (filled blue bars).

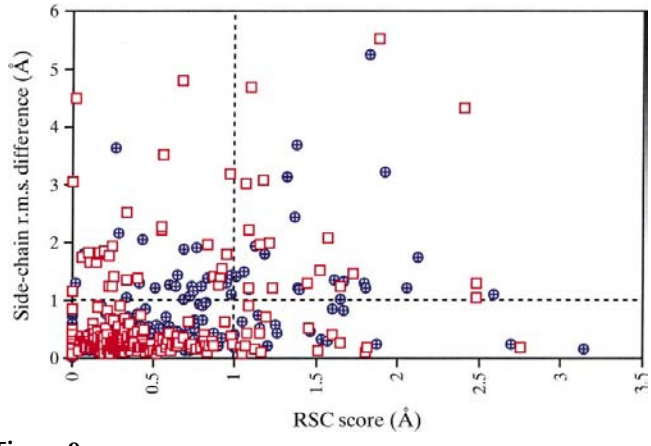

Figure 9
Side-chain r.m.s. differences *versus* RSC scores, shown for *A*3 (red squares) and *C*3 (blue circles).

1.0 Å would find an average of 42% of the side-chain errors in the last 2.6 Å models, and an average of 54% of the high RSC scores are actually errors. When it is considered that these models had been through three cycles of refinement/rebuilding and that rotamers were used in the rebuilding process, the performance of this analysis is particularly impressive. Most problems with leucine and valine side chains would be located, among others; the majority of the remaining errors could have been located using hydrogen-bonding and electron-density fit criteria.

8. General lessons

The refinements described here were essentially straightforward and so provide a reasonably simple case for analysis. Having three builders with varying experience offers a unique opportunity to isolate the types of problems faced by a novice. It was reassuring that all of the 2.6 Å refinements converged to a similar structure, which was in turn close to the final one. The different types of errors found reflected accurately the different situations involved. Previous experience in building/refinement may be described as $A >> B \simeq C$, while the time spent on the building and refinement process was $C > B >>$ A. Most quality indicators suggest that the final 2.6 Å structures should be ranked as $C > A > B$ (except for side-chain conformation, where $A > C >$ B, and the fact that the *C* model was less complete than the others). Builder *A* had an early lead in the quality race because of experience, but was unable to maintain it through a lack of available time. *B* and *C*'s problems mostly reflected innocence (mistakes in rotamers and hydrogen bonding being most noteworthy). For an inexperienced person, it seemed to be more efficient to leave questionable segments out in early cycles and let the maps resolve the problems in later cycles. 'Educated guesses' are a poorer risk when education is limited. It was clear that $R_{\rm free}$ is a good indicator of even minor problems (witness the difference between the refinements of *B* and *C* where two small segments were wrong in the former until cycle 3, but were corrected earlier in the *C* models). Doing the best job of fitting the electron density is clearly a good policy, provided other structural criteria are met. *C*'s models were excellent, except for some side-chain conformations, and the fact that more of the structure was omitted. *B* would almost certainly have performed better if allowed more time for reflection. Prior to being told the results of this study, builder *B* had gone on to build a model at 1.8 Å resolution which contained few detectable errors, without using the analytical tools we describe. The remaining problems were resolvable using hydrogen-bonding criteria. Where resolution was limited, the difficulties were clearly greater, and the proper use of the analysis tools would have been of real value. Thus, the end results of any given refinement are determined by a combination of factors which will be familiar to all crystallographers.

It is obviously a matter for debate whether it is better to have some estimate of position for the residues of mobile loops, even knowing that they are significantly lower in quality than the rest. In the case of ribokinase, the problem is of real interest; many of the residues in question were located in the most important loop of a novel nucleotide-binding fold. Builders *A* and *B* realised this, and were, therefore, more reluctant to omit those residues. Reasonable electron density was in fact present for the ligands in all three final 2.6 Å maps, and it seems likely that the problem could have been resolved with further refinement at that resolution. It should be noted further that the positions where coordinate error was large in *A* and *B* models (and omitted from *C* models) coincided with regions where the temperature factors for the main chain were higher than average (35–50 Å^2) in the high-resolution structure, despite good fit to clear, but often weaker, electron density.

9. Which tools give the best value?

This study provided an ideal opportunity for evaluation of different tools, because of the multiple models and the subsequent availability of the high-resolution structure. The fact that none of our builders made use of such tools during rebuilding becomes an asset in this context, since the evaluations of the methods themselves were not biased in that way. However, we believe that the problem of not using the available tools is a universal one, and that the tools must become easier to use and immediately available at the graphics workstation or they will not be applied. It might appear obvious, but it is worth repeating, that poor fit to electron density is the single most dependable indicator of error. While opinions vary as to whether F_o or $2F_o - F_c$ maps are most informative, it is clear that poor electron-density fit will help find errors and that structural information helps in deciding how to fix them. Our results indicate that either RSCC or temperature factors (checking atoms/residues with values outside the main distributions), and preferably both, should be used to help assess fit to electron density. Residues with values beyond the chosen cutoff can easily be selected in *O* using the graphing utilities and then coloured with the *paint_property* command. For main-chain conformations, both a Ramachandran plot and a pepflip analysis (checking residues with scores greater than 2.0 Å) should be used, since they offer somewhat different types of information. We demonstrated above that Ramachandran outliers are correlated with C^α errors, high pepflip scores and low RSCC values, but not with RSC scores. The correlation of pepflip and Ramachandran results remains even for well refined structures, but it does not lessen the value of either approach, especially for earlier models. For side-chain structure, residues with an RSC score above 1.0 Å or poor RSCC/temperature factors should be inspected with particular care.

Plotting two quality indicators together can provide a particularly potent strategy: residues with two strikes against them are more likely to represent errors, so why not deal with those first? Obviously, the plots of interest will be those that are based on information actually available during refinement. The on-the-fly graphing features of the most recent versions of *O* (Jones & Kjeldgaard, 1997*a*) make this approach especially convenient.

For finding main-chain errors, looking at the outliers in a Ramachandran plot is an excellent place to start. These plots can now be viewed with a simple pull-down menu command in *O*. By default, the symbol colouring is taken from that of the C^{α} atoms; if the residues have been coloured according to high pepflip scores, low RSCC values or any other property, those colours will appear in the symbols on the plot. An internal macro centres at each residue as its symbol's location on the plot is picked. Of course, the user can easily create macros to recalculate properties, or recolour residues that have been checked or fixed. Plot files can be generated, or screen dumps may be used to record the results for future reference. In the final Ramachandran plots, there will always be some small number of outliers remaining. Roughly half of them will have high pepflip scores as well, but none should have poor fit to the electron density.

Side-chain problems are efficiently located using a scatter plot of RSCC value *versus* RSC score for each residue (Zou & Mowbray, 1994). In Fig. 10, a series of structures illustrates the effects of improved resolution and structural quality. In a well refined structure, most residues should be clustered in the upper-left area, that is, should have a good fit to the electron density as well as to the rotamer database. The graphing commands of *O* now make these plots useful as an interactive tool in rebuilding. Again, colouring for each symbol makes it possible to view other properties simultaneously, and the centring macro is initiated by picking the individual symbols. Where residues do not match the rotamer library within the 1.0 Å cutoff, they should be checked to see whether a rotamer would be as good or better a fit to the electron density. In the final structure, no residues should have a high RSC score and poor fit to the electron density; where electron density is poor, a rotamer should always be used in preference to a non-rotamer. As mentioned above, some well supported non-rotamer conformations will always remain. As before, user-written macros may be used to recalculate values or recolour during the rebuilding session and plot files can be obtained for a permanent record.

10. Goals for the future

Experience should make it possible to obtain an initial model which is closer to the final structure, and the resulting refinement should then converge faster toward the final solution. However, it seems certain that even better initial models can actually be built, both in terms of stereochemistry and fit to the electron density. In other studies (for example, Mowbray & Cole, 1992), we have noticed that the final refined structure fits better to the original MIR map than the original model did. In the present study, we extend this to note that the 1.8 Å resolution structure is a better fit than any of the final 2.6 Å structures to the original experimental map (Fig. 11). These observations suggest that the goodness-of-fit to the MIR map can be used more effectively in building initial models. That electron-density fit is not the entire story is demonstrated by the fact that *A*'s initial model had the poorest fit to the experimental map, but its refinement progressed more rapidly. We believe this is a consequence of the nature of the errors found in the first models. *A*'s initial model had fairly large 'random errors' arising, for example, from the rough positioning of C^{α} atoms. However, the model had significantly fewer 'non-random errors' such as wrong peptide orientations or side-chain conformations (Table 3). Both were actually the result of experience, that is, of knowing what to expect from the refinement programs, and of attention to various aspects of protein structure.

Although improvements in computer performance have increased the speed of refinement greatly, the manual intervention and rebuilding portions of the average refinement macro-cycle continue to be rate-limiting. It is clear that crystallographers do not wish to spend a lot of time on manual checks which can be made more quickly and consistently by a computer, but that at present people must make most of the judgement calls. The current challenge is to determine the best ways to take the tedium, and the mistakes, out of the jobs. It seems likely that this will be performed by including more of the secondary and tertiary considerations,

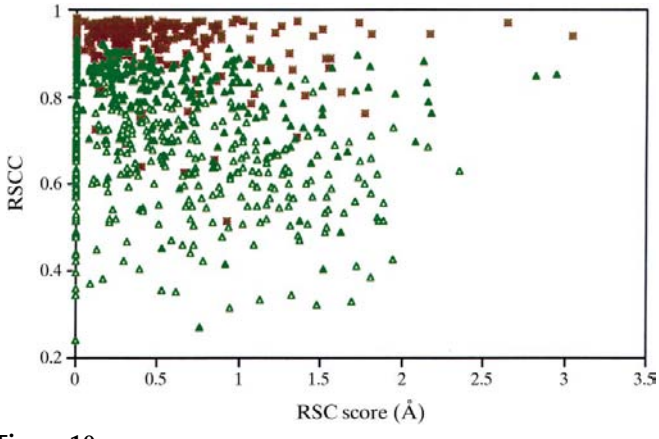

Figure 10
Scatter plot of RSCC values *versus* RSC scores, shown for *B*0 (open green triangles), *B*3 (filled green triangles) and the final 1.84 Å model (dark red stars).

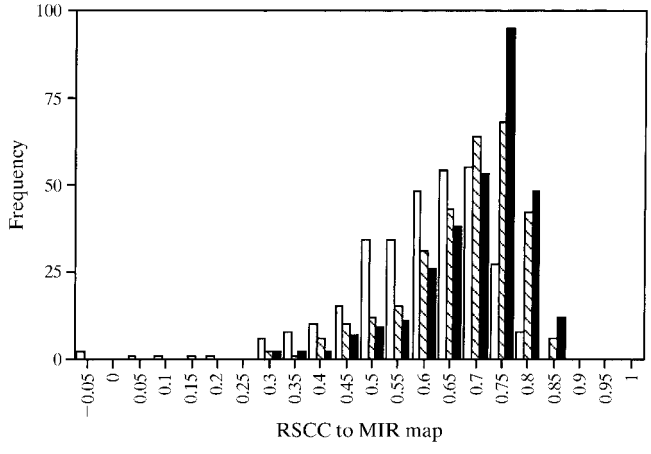

Figure 11
Histogram showing the frequency of different ranges of RSCC values for models *B*0 (open bars), *B*3 (striped bars) and the final model (solid bars), each calculated using the MIR map.

such as hydrogen-bonding checks and recognition of structural patterns, which would normally be part of the bias of the experienced model builder, while simultaneously aiming for better fit to the electron-density maps. The greater the number of errors which can be located and fixed during each manual interlude, the fewer macro-cycles will ultimately be required to obtain structures of higher quality. Then proportionately more time can be spent on the interesting aspects of structure and function which were the original purpose of the study.

References

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Brändén, C.-I. & Jones, T. A. (1990). *Nature (London)*, **343**, 687–689.

Brünger, A. T. (1992). *X-PLOR Version* 3.1. *A System for X-ray Crystallography and NMR.* New Haven: Yale University Press.

Deisenhofer, J. & Steigemann, W. (1975). *Acta Cryst.* B**31**, 238–250.

Hooft, R. W., Vriend, G., Sander, C. & Abola, E. E. (1996). *Nature (London)*, **381**(6580), 272.

Janin, J., Wodak, S., Levitt, M. & Maigret, B. (1978). *J. Mol. Biol.* **125**, 357–386.

Jones, T. A. & Kjeldgaard, M. O. (1997*a*). *Methods Enzymol.* **277**, 173–208.

Jones, T. A. & Kjeldgaard, M. O. (1997*b*). *O: the Program, Version* 6.1.1. Uppsala, Sweden.

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.

Kleywegt, G. J. (1995). *CCP4/EACBM Newslett. Protein Crystallogr.* **31**, 45–50.

Kleywegt, G. J. (1996). *Acta Cryst.* D**52**, 842–857.

Kleywegt, G. J. & Jones, T. A. (1996*a*). *Acta Cryst.* D**52**, 829–832.

Kleywegt, G. J. & Jones, T. A. (1996*b*). *Structure*, **4**, 1395–1400.

Kleywegt, G. J. & Jones, T. A. (1997*a*). *Methods Enzymol.* **277**, 208–230.

Kleywegt, G. J. & Jones, T. A. (1997*b*). *Methods Enzymol.* **277**, 525–545.

Kleywegt, G. J. & Jones, T. A. (1998). *Acta Cryst.* D**54**, 1119–1131.

Laskowski, R. A., MacArthur, M. V., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Mowbray, S. L. & Cole, L. B. (1992). *J. Mol. Biol.* **225**, 155–175.

Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.

Sigrell, J. A., Cameron, A. D., Jones, T. A. & Mowbray, S. L. (1997). *Protein Sci.* **6**, 2474–2476.

Sigrell, J. A., Cameron, A. D., Jones, T. A. & Mowbray, S. L. (1998). *Structure*, **6**, 183–193.

Zou, J. & Mowbray, S. L. (1994). *Acta Cryst.* D**50**, 237–249.