Acta Cryst. (1992). A**48**, 851-858

Model Bias in Macromolecular Crystal Structures

BY ALEC HODEL
The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, *USA*

SUNG-HOU KIM
Department of Chemistry and Lawrence Berkeley Laboratory, University of California, Berkeley, CA 94720, *USA*

AND AXEL T. BRÜNGER
The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, *USA*

(*Received* 10 *March* 1992; *accepted* 3 *June* 1992)

Abstract

Reduction of model bias in macromolecular crystallography through various omit-map techniques has been investigated. The two cases studied were the p21 protein complexed with GDP at 2.25 Å resolution and the AN02 Fab fragment of an anti-dinitrophenyl-spin-label murine monoclonal antibody complexed with its hapten at 2.9 Å resolution. In the former case, the correct model was compared to a partially incorrect model consisting of an exchanged pair of β strands along with rearrangement of the connecting loops whereas, in the latter case, the correct placement of an active-site tryptophan side chain was compared to an incorrect rotamer conformation. Partial structures were created by omission of spherical regions around the incorrect region. Omit maps without refinement of the partial structure showed a large degree of model bias. Model bias could be reduced significantly by refinement of the partial structure. Simulated-annealing refinement of the partial structure showed the best results, followed by conjugate-gradient minimization with or without prior randomization of the partial structure. To avoid compensation for missing atoms during simulated-annealing refinement of the partial structure, a suitable 'boundary' region was restrained to the starting coordinates. Model bias removal by iterative density modification was not successful in that it reduced density for both the correct and incorrect conformations.

Abbreviations

SA, simulated annealing; r.m.s., root mean square; MIR, multiple isomorphous replacement; RSR, real-space R factor.

1. Introduction

In the past decade, macromolecular crystallography has undergone major advances in crystallization (Wyckoff, Hirs & Timasheff, 1985), in data collection by synchrotron X-ray sources and area detectors (Hamlin, 1985; Fourme & Kahn, 1985) and in data analysis by high-performance computers and new computational techniques (Hendrickson, 1985; Brünger, Kuriyan & Karplus, 1987). In addition, recombinant gene technology (Goeddel, 1990) in many cases allows the expression of large amounts of protein. This has resulted in an unprecedented increase of the number of protein crystal structures elucidated. Despite these successes, the fundamental problem in X-ray crystallography, the phase problem (Hauptman, 1989), remains unchanged: that is, from a single-crystal monochromatic diffraction experiment it is possible to obtain the amplitudes but not the phases of the reflections; however, construction of the electron density by Fourier transformation requires both these components of the complex structure factors. While direct methods (Hauptman & Karle, 1953; Woolfson, 1987) have solved the phase problem for small molecules, so far they have failed for macromolecules. In the latter case, phase information has to be obtained through experimental procedures, most commonly, multiple isomorphous replacement (Green, Ingram & Perutz, 1954; Watenpaugh, 1985) or knowledge-based procedures, referred to as Patterson search (Hoppe, 1957) or molecular replacement (Rossmann & Blow, 1962; Rossmann, 1990). Phase information obtained through these techniques can be of limited accuracy and resolution, making it sometimes difficult to interpret electron-density maps in certain regions of the molecule. Furthermore, macromolecular crystals almost always diffract to less than atomic resolution, making the

0108-7673/92/060851-08$06.00

electron-density map prone to human misinterpretation. Once a model has been fit incorrectly to part of the map, most refinement methods reinforce the wrong features as well as the correct ones.

In order to improve the quality and resolution of the electron-density map, the observed phases are replaced or combined (Rossmann & Blow, 1961; Hendrickson & Lattman, 1970) with calculated phases as soon as an initial atomic model has been built. It is these combined electron-density maps that are used to improve and to refine the atomic model. The inclusion of calculated phase information implies the possibility of biasing the refinement process towards the current atomic model. This model bias can obscure the detection of errors in atomic models if sufficient phase information is unavailable. In fact, during the past decade several cases of incorrect or partly incorrect atomic models have been reported where model bias may have played a role (Bränden & Jones, 1990). Of course, most errors could have been detected either by refinement at high resolution, by comparison to homologous structures, or by well chosen site-directed mutagenesis and subsequent identification of the mutated sites as was done by Tong, Milburn, de Vos & Kim (1989). However, sometimes these options are not available and then the methods presented in this paper might be useful.

In this paper we have studied model bias in two cases. The first case is the p21 protein complexed with GDP at 2.25 Å resolution where the correct model was compared to a partially incorrect model consisting of an exchanged pair of β strands along with rearrangement of the connecting loops (de Vos *et al.*, 1988; Tong, de Vos, Milburn & Kim, 1991; Pai *et al.*, 1989). The second case is the Fab fragment of an anti-dinitrophenyl-spin-label murine monoclonal antibody (AN02) complexed with its hapten at 2.9 Å resolution (Leahy, Hynes, McConnell & Fox, 1988; Brünger, Leahy, Hynes & Fox, 1991) where the correct placement of the Trp 91 side chain near the hapten binding site was compared to an incorrect rotamer conformation. While using the same diffraction data that were employed in the original structure determinations and the corresponding atomic models containing incorrect regions or conformations, we have investigated the reduction of model bias by various omit-map techniques: ordinary σ_A-weighted omit maps, minimized omit maps with and without prior randomization, omit maps with density modification and a novel omit map that employs simulated-annealing (SA) refinement (Brünger *et al.*, 1987). Electron-density maps and real-space R factors (Jones, Zou, Cowan & Kjeldgaard, 1991) were used to assess model bias.

2. Methods

2.1. *Diffraction data and atomic models*

Human c-H-ras p21. The diffraction data on human c-H-*ras* p21 collected by de Vos *et al.* (1988) were used. The space group was $P6_522$ with unit-cell dimensions $a = b = 83.2$ and $c = 105.1$ Å. The diffraction data contained 46000 observations of 10400 unique reflections defining 87% of the theoretically observable data with a maximum resolution of 2.25 Å. As described by de Vos *et al.* (1988), the initial structure was solved to 2.7 Å by multiple isomorphous replacement and refined using *TNT* (Tronrud, Ten Eyck & Matthews, 1987) and several cycles of manual rebuilding to an R value of 26% at 2.25 Å resolution; subsequent *TNT* refinement with 100 ordered water molecules reduced the R value to 24% with a deviation of bond lengths and bond angles from ideality of 0.028 Å and 3.4°, respectively. The corrected structure included 40 solvent molecules and was refined to 2.2 Å resolution with an R value of 19% (Tong *et al.*, 1991). In the corrected structure, the deviations of bond lengths from ideality were 0.024 Å while the deviations for bond angles were 2.5°. For this study, we used the original data to 2.25 Å resolution and the partially incorrect and correct structures without water molecules. The Ramachandran plot (Ramachandran & Sasisekharan, 1968) was relatively poor for the regions containing errors.

AN02 Fab fragment. AN02 is the Fab fragment of a murine monoclonal anti-dinitrophenyl-spin-label antibody. Data collected from crystals of the hapten-bound Fab fragment were used (Leahy *et al.*, 1988; Brünger *et al.*, 1991). The space group was $P6_522$ with unit-cell dimensions $a = b = 73.23$, $c = 373.8$ Å. The diffraction data comprised 70064 observations of 12775 unique reflections with a maximum resolution of 2.9 Å defining 90% of the theoretically observable reflections. The structure was solved through generalized molecular replacement using PC refinement (Brünger, 1990). The structure was refined using *X-PLOR* (Brünger, 1992) to an R value of 19.5% at 2.9 Å resolution. The deviations of bond lengths and bond angles from ideality were 0.014 Å and 3.1°, respectively. The numbering scheme used in this paper refers to that of Kabat, Wu, Reid-Miller, Perry & Gottesman (1987).

2.2. *Methods for the removal of phase bias*

σ_A weighting. All maps presented here are $(2|F_o| - |F_c|)\exp(i\alpha_c)$ maps where the structure-factor amplitudes were weighted in order to reduce the model bias of an incomplete or partially incorrect structure. This modification was performed by the *SIGMAA* program written by Read (1986, 1990). The Fourier coefficients calculated by this program are given by

$$F_{\text{map}} = (2mF_o - DF_c)\exp(i\alpha_c), \qquad (1)$$

where m is the figure of merit and D, a measure of the error in the coordinates of the model, is defined as by Luzzati (1952). These coefficients may be calculated from F_o and F_c (Read, 1986). We will refer to maps computed from F_{map} as σ_A-weighted $2F_o - F_c$ maps. The coeffi-

cients m and D arise from the approximation put forward by Main (1979) that

$$mF_o \exp(i\alpha_c) \simeq \tfrac{1}{2}F_o + \tfrac{1}{2}F_c, \quad (2)$$

where F_c and α_c are calculated using an incomplete model. Thus, coefficients that reduce the model bias of a partial structure are given by $2m$. Read (1986) extended this derivation to reduce the model bias of a partial structure with errors, resulting in (1).

Ordinary σ_A-weighted omit map. In this case, atoms in the questionable region were removed from the calculation of F_c when constructing the weighted $2F_o - F_c$ maps (Bhat & Cohen, 1984) and no density modification or refinement was carried out.

Omit map with density modification. In analogy to the process of solvent flattening (Podjarny, Bhat & Zwick, 1987; Wang, 1985), we have attempted to reduce model bias in the questionable region through an iterative procedure:

1. A σ_A-weighted $2F_o - F_c$ omit map is calculated.
2. The region of the map which encompasses the area in question is set to zero density, removing all information from this area. Note that when these maps are calculated a constant is added so that F_{000}, or the average density, is set to zero. The region of the modification is defined by a mask surrounding the omitted atoms of the structure.
3. The structure factors that correspond to this modified map are calculated by Fourier transformation. In this way, any information about atomic positions in the questionable region should be reduced or 'flattened'.
4. The phases taken from these 'flattened' structure factors are now combined with $2F_o - F_c$ magnitudes calculated from the experimental data and the model structure.
5. A new electron-density map is then calculated using these combined structure factors.

This new map, which represents one cycle of density modification, can then be used to start the cycle over again at step 2. This process is repeated iteratively until convergence is reached. In the cases studied, we found it to converge within four to five cycles.

Minimized σ_A-weighted omit map. To reduce model bias originating from the atomic positions outside the omitted region, the structure *with the questionable region omitted* can be refined against the diffraction data. We utilized conjugate-gradient minimization (Powell, 1977) as an example for conventional crystallographic refinement, that is, a gradient-descent method. All references to the omitted atoms were removed from both the structure-factor calculation and the empirical energy function used in the refinement (Brünger, Karplus & Petsko, 1989).

Simulated annealing σ_A-weighted omit map. The employment of SA in refinement has been shown to be useful in reaching more optimal structures than can be obtained through gradient-descent methods, such as least-squares optimization or conjugate-gradient minimization (Brünger *et al.*, 1987). The following procedure for obtaining a SA omit map was developed (see also Fig. 1):

1. The questionable region was omitted from the structure-factor calculation as well as from the empirical energy function in order to remove all information pertaining to the atoms' existence.
2. The refinement was restrained so as to prohibit the collapse of the surrounding atoms into the omitted region. Without the restraint, nearby atoms could move into the omitted region in an attempt to compensate for the lack of scattering in that region. In the studies presented here, the atoms of residues within 3 Å of the omitted region were harmonically restrained to their former positions throughout the refinement (Bruccoleri & Karplus, 1986).
3. SA refinement using the slow-cooling protocol by Brünger, Krukowski & Erickson (1990) was performed on the partial structure.
4. Using this new coordinate set, a σ_A-weighted $2F_o - F_c$ map was calculated with the questionable region still omitted.

It should be noted that our method represents an 'inverted' boundary condition. A 'normal' stochastic boundary condition consists of a reaction region whose atoms are treated explicitly, a layer of restrained atoms and the surroundings which are neglected (Brooks, Brünger & Karplus, 1985; Brünger, Brooks & Karplus, 1984).

2.3. *Real-space R factor*

To perform a quantitative evaluation of the fit of density maps to atomic models, the real-space R factor (RSR) was computed using the algorithm described by Jones, Zou, Cowan & Kjeldgaard (1991). This algorithm compares the density map to the electron density calculated directly from the position of the atoms in the model. The RSR for a particular residue is defined as

$$\text{RSR} = \frac{\sum |\rho_o - \rho_c|}{\sum |\rho_o + \rho_c|}, \quad (3)$$

where the summations are carried out over a finite grid in the neighborhood of the residue, ρ_o is the density map and ρ_c is the electron density calculated from the selected atomic positions of the residue.

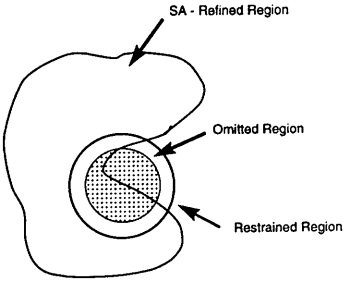


Fig. 1. A schematic representation of the simulated-annealing omit-map procedure. As described in *Methods*, the region in question is omitted from the structure. The atoms in a boundary region surrounding the omission are restrained to their starting coordinates. This restrained partial structure is then refined through simulated annealing to reduce the phase bias from the omitted region.

2.4. *Computational methods*

All computations were performed using the *X-PLOR* program by Brünger (1992), the *SIGMAA* program by Read (1986) and the *O* program by Jones *et al.* (1991).

3. Results and discussion

3.1. *p*21

The error in the partially incorrect structure of p21 originates in the misdirection of the backbone-chain trace resulting in the exchange of two out of six β strands (Fig. 2); one of the two strands is formed by residues 1 to 9, the other one by residues 54 through 61 (de Vos *et al.*, 1988; Tong *et al.*, 1989; Pai *et al.*, 1989). This misdirection in the chain trace yields three 'branch' points (residues 9–12, 48–54, 61–65), where, if one superimposed the correct structure on the incorrect one, the backbone-chain trace appears to have a choice between the correct and incorrect paths. We have focused on one of the branch points (residues 9–12) in the following. This had the strongest density of the three branch points. The other two branch points are partially disordered even in the correct structure. In fact, the error in assignment at branch point 9–12 occurred because if one followed the chain trace to the region of residues 61–65 some of the side chains in this disordered region appeared to fit better in the case of the incorrect connectivity.

The region that was deleted from the p21 structure in the omit-map calculations is indicated in Fig. 2. Specifically, all atoms belonging to residues within 5 Å of residue 10 were deleted from the incorrect model for all omit-map calculations and refinements of p21. It should be pointed out that, even with these omissions, the model used in the omit-map calculations and refinements still contained the incorrect β strands and connectivity.

Fig. 3 shows several omit maps for the p21 structure using the procedures described in *Methods*. To quantify the agreement between maps and models, RSR differences are reported in Fig. 4. As described in *Methods*, the RSR is the real-space analog to the R value used in crystallography; it measures the agreement between the electron density and the atomic models (Jones *et al.*, 1991). To quantify model bias, the RSR was calculated for the various omit maps using the backbone atoms of each residue in the loops of the correct and incorrect structures. Then the RSR value for each residue of the incorrect branch was

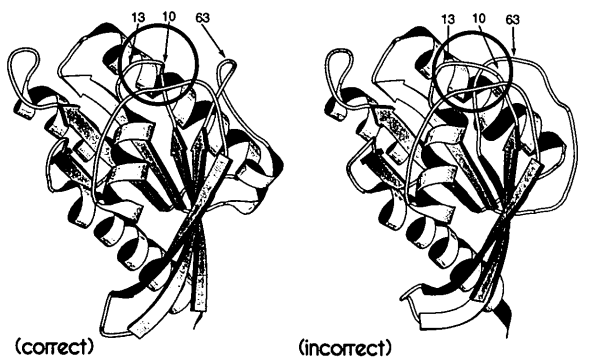


Fig. 2. Ribbon drawings of the correct and partially incorrect structures of p21. The dark circle is drawn around the region where the omit-map calculations were directed. Here, the β-strand crossover is apparent by the change in the conformation of the loops around residue 10 and residue 63. In the omit-map calculations, residues within 5 Å of residue 10 were removed from the incorrect structure.

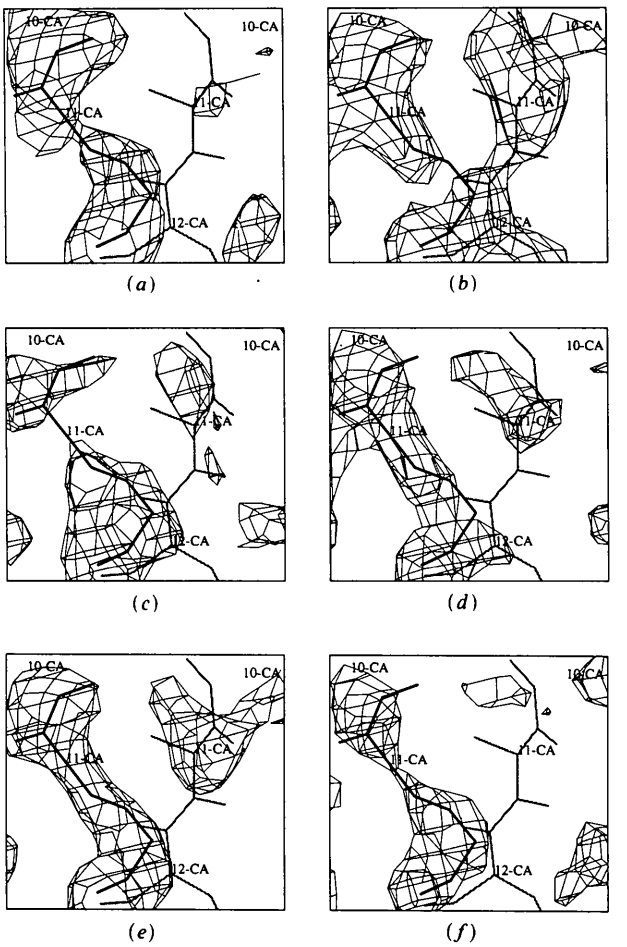


Fig. 3. The effects of various omit-map techniques around residues 10–12 in p21. The correct branch is shown in black, the incorrect branch is shown in gray. All maps are of the σ_A-weighted $2F_o - F_c$ type (see *Methods*) at 2.25 Å resolution. (*a*) Ordinary omit map of the correct structure shown at a contour level of 1.2σ. (*b*) Ordinary omit map of the incorrect structure shown at a contour level of 1.2σ. All residues within 5 Å of residue 10 were omitted in all calculations. (*c*) Omit map after four iterations of density modification shown at a contour level of 0.6σ. The density-modification region was defined by a 2.0 Å cushion around the omitted atoms. (*d*) Minimized omit map of the incorrect structure shown at a contour level of 1.0σ. The partial structure was refined through 120 cycles of conjugate-gradient minimization at 8.0–2.25 Å resolution. (*e*) Randomized and minimized omit map of the partially incorrect structure shown at a contour level of 1.0σ. (*f*) SA omit map of the partially incorrect structure at 1.0σ. The partial structure was refined using the SA slow-cooling protocol (Brünger *et al.*, 1990) with a starting temperature of 3000 K at 8.0–2.25 Å resolution.

subtracted from the value calculated for the corresponding residue of the correct branch. These RSR differences are plotted in Fig. 4 for residues 10 through 13, where a negative value corresponds to a better fit of the map to the correct structure than to the incorrect structure. Thus, the more negative the RSR difference is, the less the models are biased.

As shown in Fig. 3(*b*), the ordinary σ_A-weighted omit map of the incorrect structure shows density for both the correct and incorrect branches of the chain trace, creating an ambivalent situation. Relatively large model bias is indicated in Fig. 4 for residues 10, 11 and 12. The spurious density in this ordinary omit map originates from model bias contained in all the atomic coordinates of the model which *a posteriori* confirms the incorrect connectivity. Apparently, prior 'over'-refinement of the complete structure has produced a 'memory' for the incorrectly placed atoms in the rest of the atomic model. As a control, the σ_A-weighted omit map for the correct structure is shown in Fig. 3(*a*). Little density appears for the incorrect chain trace.

A σ_A-weighted omit map for the incorrect structure with four cycles of density modification is shown in Fig. 3(*c*). At each iteration, the overall density in the area was reduced, yet only small changes in the density map were observed when lowering the contour level appropriately. In fact, density was reduced for both the correct and incorrect chain trace. Thus, omit maps with density modification produced little reduction of model bias.

The next attempts at reducing the model bias focused on the modification of the structure itself. The partial structure was refined through 120 cycles of conjugate-gradient minimization against a hybrid energy function comprised of the crystallographic residual and an empirical energy function (Brünger *et al.*, 1989). The σ_A-weighted omit map calculated from this refined partial structure is shown in Fig. 3(*d*). The electron density along the incorrect backbone trace is now broken between residues 11 and 12, and the RSR differences indicate less model bias for residues 11 and 10. The broken density near the incorrect model and the increased density over the correct model indicate the direction of the proper backbone trace. However, residual model bias was still present manifesting itself as a patch of density over residue 11.

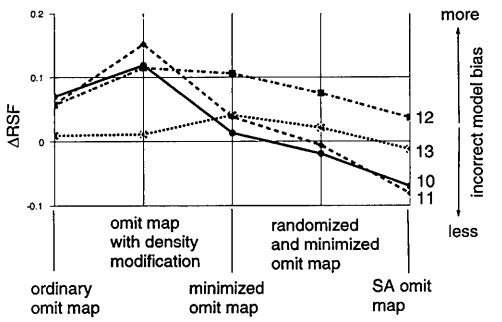


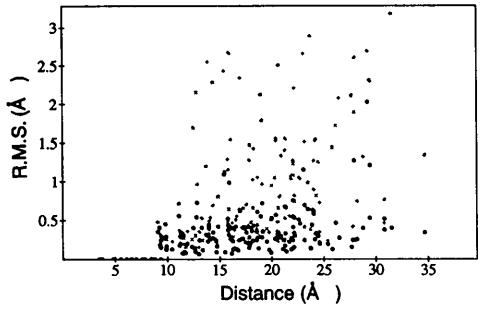
Fig. 4. The difference between the RSR calculated for the correct structure and that of the incorrect structure (Jones *et al.*, 1991). The RSR values for all backbone atoms of each residue of the incorrect branch site were subtracted from the values calculated for the corresponding residue in the correct structure. These values are plotted for the various σ_A-weighted $2F_o - F_c$ omit maps *computed for the incorrect structure*. The RSR is analogous to the *R* value used in crystallography. Thus, a small value signals a better fit. For the difference plotted here, a negative value denotes a better fit between the map and the correct structure than between the map and the same residue in the incorrect structure, *i.e.* less model bias towards the incorrect structure.

The origin of this residual model bias is caused by the coherent placement of all atoms in the structure that occurs during refinement of the complete model, that is, the structure is 'trapped' in a local minimum of the target function used for refinement. Conjugate-gradient minimization appears to be insufficient to escape from this local minimum even when omitting the incorrect atoms. It is then conceivable that this coherent configuration which causes the model bias could be reduced by adding some random number to the coordinates of the structure followed by refinement. To this end, random numbers were added to the coordinates of the partial structure. The random numbers were taken from a Gaussian distribution centered around zero with a standard deviation of 0.5 Å. This randomized model was refined using 120 cycles of conjugate-gradient minimization. The resulting map, calculated from the randomized and minimized model, showed slight improvement over the minimized omit map (Figs. 3*e* and 4).

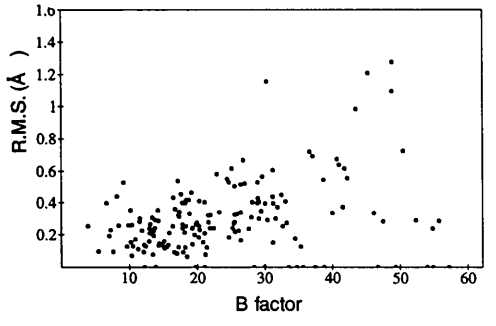
The next step was to use the SA-refinement method (Brünger *et al.*, 1987), which has a larger radius of convergence. Boundary restraints around the omitted region were employed as described in *Methods*. The resulting coordinates were used to calculate a σ_A-weighted $2F_o - F_c$ omit map (Fig. 3*f*). This map provided the clearest indication of which branch of the backbone trace is correct. The density along the wrong branch disappears entirely while the correct density shows only little reduction. Fig. 4 (last column) shows that model bias is now minimal for residues 10 through 13.

The changes in the p21 structure that took place as a result of the SA omit refinement, and hence are responsible for the removal of model bias, are analyzed in Fig. 5. The root-mean-square (r.m.s.) shifts in coordinates between the atoms of the incorrect model and the coordinates after the SA omit refinement show little correlation with the spatial distance to residue 10 (Fig. 5*a*). The correlation coefficient is 0.28% for backbone atoms and 0.45% for side-chain atoms. Note that the atoms in the restrained shell around the omitted region were left out of the correlation-coefficient computation. Fig. 5(*b*) shows that the r.m.s. shifts of the incorrect structure are more correlated with the refined atomic temperature factors (the correlation coefficient is 0.53%). Clearly, the distribution of r.m.s. shifts is not purely random: it has to account for the image of the spurious density in the omitted region. It is interesting to note that the magnitude of the changes in the backbone coordinates that have taken place is of the same order as the error in the coordinates determined by a

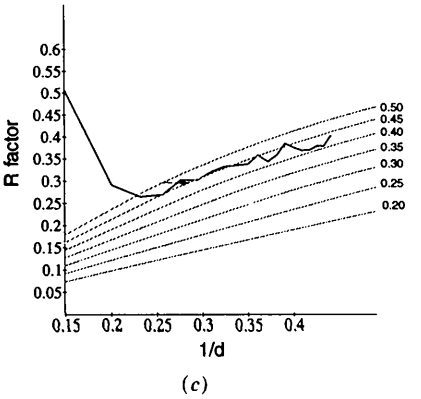
Luzatti plot at 2.25 Å (Luzatti, 1952) for the correct structure (Fig. 5*c*). From this plot, one can estimate the error in the coordinates to be between 0.4 and 0.45 Å. The average value of the r.m.s. shifts of the backbone atoms after the SA omit refinement is 0.34 Å. Thus, the r.m.s. shifts that are responsible for producing the model bias are 'hidden' in the expected coordinate error of the model.



(*a*)



(*b*)

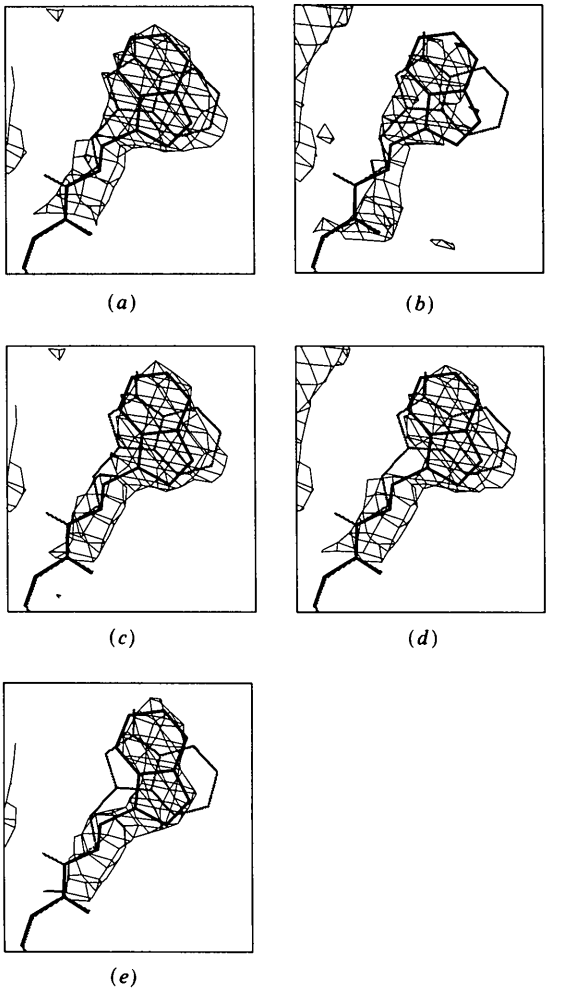


(*c*)

Fig. 5. The effects of the SA omit refinement on the coordinates of the incorrect model of p21. (*a*) Shown are the r.m.s. shifts in the coordinates of the backbone atoms (black dots) and side-chain atoms (gray dots) before and after the SA omit refinement of the residues of p21 *versus* the distance of that residue from residue 10 in the omitted region. (*b*) R.m.s. shifts in the coordinates of the backbone atoms before and after the SA omit refinement for each residue are plotted *versus* the average temperature factor of each residue's backbone atoms. (*c*) A Luzatti (1952) plot of the p21 data using the correct structure at 2.25 Å resolution.

3.2. *The binding pocket of the AN02 Fab fragment*

If an atomic model is more or less correct, model bias can still manifest itself in a subtle manner, such as obscuring the conformation of a side chain. Although this form of model bias generally has little effect on the R



(*a*) (*b*) (*c*) (*d*) (*e*)

Fig. 6. The effects of various omit-map techniques around Trp 91 located in the light chain of the AN02 Fab fragment. The correct conformation of Trp 91 is shown in black, the initial incorrect conformation is shown in gray. All maps are of the σ_A-weighted $2F_o - F_c$ type (see *Methods*) at 2.8 Å resolution. Residues 89–97 in the light chain, 93–102 from the heavy chain, and the hapten molecule were omitted for all map calculations and refinements. (*a*) Ordinary omit map of the initial incorrect structure shown at a contour level of 1.2σ. (*b*) The fourth iteration of density modification on the initial incorrect structure shown at a contour level of 0.7σ. The region of modification was defined by a 2.0 Å cushion around the omitted region. (*c*) Minimized omit map of the initial structure shown at a contour level of 1.2σ. The partial structure was refined through 120 cycles of conjugate-gradient minimization at 8.0–2.8 Å resolution. (*d*) Randomized and minimized omit map of the initial structure shown at a contour level of 1.2σ. (*e*) SA omit map of the initial structure shown at a contour level of 1.2σ. The partial structure was refined using a SA slow-cooling protocol (Brünger *et al.*, 1990) with a starting temperature of 3000 K at 8.0–2.8 Å resolution.

value or the average phase accuracy of the structure, it can cause misinterpretation of structure/function relationships when it occurs near biologically important sites of a macromolecule. A case in point is Trp 91 of the light chain of the AN02 Fab fragment, which is positioned in the hapten binding pocket by stacking against the dinitrophenyl group of the hapten (Brünger *et al.*, 1991). During the course of refinement of AN02, an ordinary omit map of the binding pocket suggested the initial incorrect placement of the side chain of Trp 91 (Fig. 6*a*). Only the use of a SA omit map revealed the correct rotamer conformation of Trp 91 (Brünger *et al.*, 1991). Accurate placement of this residue was crucial for proper interpretation of biophysical studies on this system. In the following we have analyzed the reduction of model bias through the various omit-map techniques in more detail.

Fig. 6 shows omit maps using the coordinates of the structure that had been refined with Trp 91 in the incorrect conformation. The two loops bounding the hapten molecule (residues 89–97 in the light chain and 93–102 from the heavy chain) were deleted from the structure and all residues within 3 Å of these loops were harmonically restrained to their former positions. This partial structure was then used for all omit maps and refinements in Fig. 6. The ordinary omit map (Fig. 6*a*) clearly creates an ambivalent situation where the incorrect rotamer appears to fit the density somewhat better than the correct rotamer. As in the p21 case, density modification reduces density for both the correct and incorrect rotamer conformation (Fig. 6*b*). Refinement of the partial structure reduces model bias, where SA refinement clearly shows the least model bias (Figs. 6*c* – *e*). The SA omit map also suggests that Trp 91 assumes only one rotamer conformation or that any alternative rotamer conformation must have a very low occupancy.

4. Concluding remarks

In evaluating the quality of an X-ray crystal structure, comparison of the model to various electron-density maps is most fundamental. However, the inaccuracy and limited resolution of observed phases (*e.g.* from MIR) usually necessitates phase combination with model phases. The resulting model bias can be severe enough to *a posteriori* confirm the chain trace of an (partially) incorrect structure or at least to fail to indicate the incorrectness of the structure. In the two cases studied, relatively small changes in atomic backbone position (< 0.6 Å) can create spurious density. In fact, most of these coordinate displacements are within the expected coordinate error of the structures as estimated from a Luzatti plot. Thus, the coordinate differences that are responsible for producing the spurious density or model bias are 'hidden' in the expected coordinate error of the refined model. Ordinary omit maps without prior refinement of the partial structure appeared as ineffective in removing the model bias regardless of the weighting employed. Thus, ordinary omit maps should be avoided for macromolecular structure determinations. Density modification was found to reduce the density of both correct and incorrect conformations, thus it did not reduce model bias either.

We investigated the reduction of model bias by refining the coordinates of the partial model against a hybrid energy function used in crystallographic refinement (Brünger *et al.*, 1987) with the questionable region omitted. We found that conjugate-gradient minimization significantly reduces model bias where the model has been previously refined *with the questionable region included*. Thus, it appears that conjugate-gradient minimization in many cases is sufficient to detect incorrect positioning of atoms. However, we found that this approach is not always sufficient to remove model bias completely. The refinement could be trapped in a local minimum and the radius of convergence for gradient-descent-refinement techniques could be too small to escape from this minimum although the erroneously placed atoms are omitted. Even randomization of the coordinates before minimizing does not let the structure escape from the local minimum. This failure is not specific to our method of minimization. Conjugate-gradient minimization has been shown to be at least as efficient as least-squares refinement (Brünger, 1988).

The refinement of the partial structure by simulated annealing (Brünger *et al.*, 1987, 1990) showed the best results of all the methods attempted. One could argue that this is a result of the use of SA to refine the complete model in the first place: SA refinement of the incorrect structure brought the model even deeper into the incorrect local minimum so clearly only SA would be able to reverse the process. However, in the case of H-*ras* p21, the partially incorrect model was refined through several cycles of rebuilding and least-squares (*i.e.* gradient-descent) refinement and yet SA refinement of the partial region reduced the model bias further as compared to conjugate-gradient minimization. Thus, we conclude that the SA omit-map method optimally reduces model bias for structures that were refined by conventional (*i.e.* least-squares) or SA methods. Clearly, this method requires substantial computational resources.

The success of any omit-map technique depends heavily on the quality of the experimental data. A clear image of the molecule cannot be produced in a region where the data do not sufficiently define the electron density. In fact, two of the three mistraced loops in the p21 structure around residues 47–54 and residues 61–65 produced omit maps which had very little information at all (not shown). However, the corresponding omit maps computed from the correct structure also do not clearly define the chain trace in those two loop regions. This is also corroborated by the high temperature factors in residues 61–65 (Tong *et al.*, 1989). This implies that these loops are disordered or that the original diffraction data used do not adequately describe these regions in any detail at the observed resolution. The correct connectivity around residue 59 was tested by site-directed mutagenesis (Tong *et al.*, 1991); a

difference electron-density map using the phases from the partially incorrect model revealed the correct position of residue 59, which in turn resolved the ambiguity around branch point 9–12. As we have shown here, the omit-map techniques with refinement of the partial structure could have provided an alternative means of resolving the ambiguity around this branch point.

In this study we purposely restricted ourselves to using only the observed amplitude information and the (partially) incorrect structures. Inclusion of MIR phases could potentially improve the quality of the maps (Stuart & Artymiuk, 1985). However, such independent phase information is not available when using molecular-replacement techniques (Hoppe, 1957; Rossmann & Blow, 1962) or *ab initio* phasing methods should the latter become feasible for macromolecules.

We are grateful to Michael Rossmann for suggesting the application of density modification to the omit-map procedure, and to Robert Fox, Dan Leahy and Tomas Hynes for providing the AN02 diffraction data and coordinates. AH is a Howard Hughes Medical Institute Predoctoral Fellow. Support through the Howard Hughes Medical Institute (ATB), the Pittsburgh Supercomputer Center (Grant No. DIR9021975) and a loaner grant to ATB of a SGI 4D/220 GTX from Silicon Graphics Inc. is gratefully acknowledged.

References

BHAT, T. N. & COHEN, G. H. (1984). *J. Appl. Cryst.* **17**, 244–248.
BRÄNDEN, C. I. & JONES, A. (1990). *Nature (London)*, **343**, 687–689.
BROOKS, C. L., BRÜNGER, A. T. & KARPLUS, M. (1985). *Biopolymers*, **24**, 843–865.
BRUCCOLERI, R. E. & KARPLUS, M. (1986). *J. Comput. Chem.* **7**, 175–185.
BRÜNGER, A. T. (1988). *J. Mol. Biol.* **203**, 803–816.
BRÜNGER, A. T. (1990). *Acta Cryst.* A**46**, 46–57.
BRÜNGER, A. T. (1992). *X-PLOR Manual*, Version 3.0. Yale University, New Haven, CT, USA.
BRÜNGER, A. T., BROOKS, C. L. & KARPLUS, M. (1984). *Chem. Phys. Lett.* **105**, 495–500.
BRÜNGER, A. T., KARPLUS, M. & PETSKO, G. A. (1989). *Acta Cryst.* A**45**, 50–61.
BRÜNGER, A. T., KRUKOWSKI, A. & ERICKSON, J. W. (1990). *Acta Cryst.* A**46**, 585–593.
BRÜNGER, A. T., KURIYAN, J. & KARPLUS, M. (1987). *Science*, **235**, 458–460.
BRÜNGER, A. T., LEAHY, D. J., HYNES, T. R. & FOX, R. O. (1991). *J. Mol. Biol.* **221**, 239–256.
FOURME, R. & KAHN, R. (1985). *Methods Enzymol.* **114**, 281–299.
GOEDDEL, D. V. (1990). Editor. *Methods in Enzymology*, Vol. 185. New York: Academic Press.
GREEN, D. W., INGRAM, V. M. & PERUTZ, M. F. (1954). *Proc. R. Soc. London Ser. A*, **225**, 287–307.
HAMLIN, R. (1985). *Methods Enzymol.* **114**, 452–471.
HAUPTMAN, H. A. (1989). *Phys. Today*, **42**, 24–29.
HAUPTMAN, H. A. & KARLE, J. (1953). *Solution of the Phase Problem. 1. Centrosymmetric Crystal.* ACA Monograph 3. Polycrystal Book Service, Pittsburg, PA, USA.
HENDRICKSON, W. A. (1985). *Methods Enzymol.* **115**, 252–270.
HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst.* B**26**, 136–143.
HOPPE, W. (1957). *Acta Cryst.* **10**, 750–751.
JONES, T. A., ZOU, J.-Y., COWAN, S. W. & KJELDGAARD, M. (1991). *Acta Cryst.* A**47**, 110–119.
KABAT, E. A., WU, T. T., REID-MILLER, M., PERRY, H. M. & GOTTESMAN, K. S. (1987). *Sequences of Proteins of Immunological Interest*, 4th ed. National Institutes of Health, Bethesda, MD, USA.
LEAHY, D. J., HYNES, T. R., MCCONNELL, H. M. & FOX, R. O. (1988). *J. Mol. Biol.* **203**, 829–830.
LUZATTI, V. (1952). *Acta Cryst.* **5**, 802–810.
MAIN, P. (1979). *Acta Cryst.* A**35**, 779–785.
PAI, E. F., KABSCH, W., KRENGEL, U., HOLMES, K. C., JOHN, J. & WITTINGHOFER, A. (1989). *Nature (London)*, **341**, 209–214.
PODJARNY, A. D., BHAT, T. N. & ZWICK, M. (1987). *Annu. Rev. Biophys. Biophys. Chem.* **16**, 351–371.
POWELL, M. J. D. (1977). *Math. Prog.* **12**, 241–254.
RAMACHANDRAN, G. N. & SASISEKHARAN, V. (1968). *Adv. Prot. Chem.* **23**, 283–437.
READ, R. (1986). *Acta Cryst.* A**42**, 140–149.
READ, R. (1990). *Acta Cryst.* A**46**, 900–912.
ROSSMANN, M. G. (1990). *Acta Cryst.* A**46**, 73–82.
ROSSMANN, M. G. & BLOW, D. M. (1961). *Acta Cryst.* **14**, 641–647.
ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24–31.
STUART, D. & ARTYMIUK, P. (1985). *Acta Cryst.* A**40**, 713–716.
TONG, L., MILBURN, M. V., DE VOS, A. M. & KIM, S.-H. (1989). *Science*, **245**, 244.
TONG, L., DE VOS, A. M., MILBURN, M. V. & KIM, S.-H. (1991). *J. Mol. Biol.* **217**, 503–516.
TRONRUD, D. E., TEN EYCK, L. F. & MATTHEWS, B. W. (1987). *Acta Cryst.* A**43**, 489–501.
VOS, A. M. DE, TONG, L., MILBURN, M. V., MATIAS, P. M., JANCARIK, J., NOGUCHI, S., NISHIMURA, S., MIURA, K., OHTSUKA, E. & KIM, S.-H. (1988). *Science*, **239**, 888–893.
WANG, B. C. (1985). *Methods Enzymol.* **115**, 90–142.
WATENPAUGH, K. D. (1985). *Methods Enzymol.* **115**, 3–14.
WOOLFSON, M. M. (1987). *Acta Cryst.* A**43**, 593–612.
WYCKOFF, H. W., HIRS, C. H. W. & TIMASHEFF, S. N. (1985). *Methods Enzymol.* **114**, 49–198.