Proc. Nat. Acad. Sci. USA
Vol. 71, No. 9, pp. 3736–3740, September 1974

Refinement of the X-ray Structure of Rubredoxin by Conformational Energy Calculations

(theoretical calculations/energy minimization)

DANIEL RASSE, PAUL K. WARME, AND HAROLD A. SCHERAGA

Department of Chemistry, Cornell University, Ithaca, New York 14850

Contributed by Harold A. Scheraga, June 10, 1974

ABSTRACT **The x-ray structure of rubredoxin has been refined by energy minimization. The computed structure is constrained to have standard bond lengths, bond angles, and planar *trans* peptide groups. Also, since most of the steric overlaps have been relieved, it has a very low energy. As judged by the root mean square (RMS) deviation of the computed coordinates from those of the x-ray structure, the two are very similar. The reliability index *R* for the computed structure (determined from the structure factors for the calculated conformation) is 0.37, which is comparable to that for other proteins.**

The crystal structure of *Clostridium pasteurianum* rubredoxin has been determined at 1.5-Å resolution and refined to a high degree (1). However, on energetic grounds, it seems unlikely that the bond lengths and bond angles of the x-ray-refined structure should deviate from their equilibrium values as much as the examples shown in Fig. 1. Therefore, we have carried out an energetic refinement of the x-ray structure.

The refinement is carried out in three stages. In stage I (2), the dihedral angles of the backbone and side chains of the polypeptide chain are adjusted by a least-squares procedure to produce a structure conforming as closely as possible to the x-ray coordinates but, at the same time, one which is constrained to have fixed bond lengths, bond angles, and planar *trans* peptide groups that are characteristic of amino acids and small peptides*. In stage II (3), most of the large atomic overlaps are relieved by partial minimization of a portion of the total energy, namely, the nonbonded, hydrogen-bond, torsional, and "fitting potential" contributions. In stage III (4), the energy contributions of stage II, together with the electrostatic energy and a more refined hydrogen-bond potential (compatible with the electrostatic energy introduced at this stage), are minimized.

PROCEDURE

Energy minimization

The computational procedure was similar to that described previously (2–4), with the modifications noted below.

The tetrahedral iron complex of rubredoxin presented a new computational problem. During stage I and the first iteration of stage II, the positions of the cysteine sulfur atoms relative to the iron atom were not constrained to form a tetrahedral complex. Instead, reliance was placed on the fitting potential, E_{FP} of Eq. 1, to maintain the sulfur atoms in their proper positions. The fitting potential applied to all nonhydrogen atoms in the structure, including the cysteine sulfur atoms, is of the form

$$E_{FP} = W\sum_i D_i^2 \qquad [1]$$

where D_i is the distance between the calculated atom position and its corresponding x-ray position, and W is a weighting factor taken as 50 kcal/mole Å^2 in the first iteration of stage II (on segments 1–20, 18–37, and 35–54). During this initial application of the stage I and II procedures, the iron atom at the active site of rubredoxin was not included in the computations.

In a second iteration of stage II (carried out on segments 1–15, 13–29, 27–42, and 40–54), W was reduced to 10 kcal/mole Å^2, and a loop-closing energy, E_{loop}, was introduced to preserve the active site geometry. In this iteration, the iron atom at the active site was introduced into the computations, and the active site was treated by two alternative models in all subsequent energy calculations. In one series of calculations, the Fe–S bond lengths (including one very short Fe–S distance of 2.05 Å) and the S–Fe–S bond angles were maintained fixed by keeping the iron and the sulfur atoms at their positions observed by Watenpaugh *et al.* (1); in the alternative series of calculations, a slightly modified (but fixed) geometry around the iron atom (with the Fe–S_{42} bond lengthened from 2.05 Å to 2.20 Å, but with no other changes in bond lengths or bond angles around the iron atom) was examined. The loop-closing potentials were applied to the four C^β–S bonds, and calculated as

$$E_{\text{loop}} = (1/2)[K_l(l - l_0)^2] + K_{\nu_1}[1 - \cos(\nu_1 - \nu_1^0)] + K_{\nu_2}[1 - \cos(\nu_2 - \nu_2^0)] \qquad [2]$$

which is similar to the potential of Gibson and Scheraga (5) for closing disulfide loops. In Eq. 2, l is the distance between the C^β and S atoms and l_0 is the equilibrium bond length (1.82 Å*), ν_1 and ν_2 are the C^α–C^β–S and C^β–S–Fe bond angles and ν_1^0 and ν_2^0 are their equilibrium values [113°* and 104° (5), respectively], and the force constants were taken as K_l = 1000 kcal/mole Å^2, K_{ν_1} = 500 kcal/mole, and K_{ν_2} = 250 kcal/mole.

After these two iterations of stage II, it was observed that the calculated coordinates of the first several residues differed from the x-ray coordinates significantly more than did those of the remaining residues, probably because the fixed coordinate system for the whole molecule was initially based on the N, C^α, and C^β atoms of the first residue; unfortunately,

Abbreviation: RMS, root mean square.

* F. A. Momany, R. F. McGuire, and H. A. Scheraga, manuscript to be submitted.

the coordinates of this N-terminal residue were not very accurate, judging from the B factors in Fig. 2. Therefore, the coordinates of the first several residues were readjusted, by a slight modification of the stage I and II procedures, in order to improve their fit to the x-ray coordinates and also to prevent propagation of these errors in the N-terminal region to other parts of the molecule (4). This resulting structure was used as a starting point for the final (stage III) refinement.

In stage III, the total energy that was minimized was $E_{NB} + E_{ROT} + E_{loop} + E_{FP} + E_{EL} + E_{HB}$ (4). The FeS_4 moiety was taken as neutral (assuming that the net charge of -1 is balanced by a counterion from the external medium, there being no lysine side chains in the immediate neighborhood of this moiety). The partial charges on the backbone and side-chain atoms are the same as those used by Warme and Scheraga (4). A value of 2 (6) was taken for the dielectric constant. An important difference between our earlier use of the stage III procedure (4) and that applied here to rubredoxin is that the energy of the *whole* molecule (rather than segments thereof) was minimized with respect to the backbone and side-chain dihedral angles, using the Fletcher minimization procedure (7, 8). As the energy decreased, the weighting factor W in E_{FP} was reduced progressively from 10 to 5 to 1 kcal/mole Å^2, and 250 iterations in total (i.e., 60, 40, and 150, respectively, for the three values of W) were used in the minimization. The energy decreased by 0.2 kcal/mole or less per iteration during the last 20 iterations, indicating that the energy was probably close to its minimum value; no significant changes of conformation were observed during the last 20 iterations, none of the dihedral angles being altered by more than 0.9°.

Structure factors

The crystallographic structure factors of the conformation obtained after stage III were calculated with a computer program kindly supplied by Dr. G. N. Reeke, Jr. of Rockefeller University. The 425 heavy atoms of the molecule (omitting hydrogen atoms) and the oxygens of the 130 water molecules at the positions reported by Watenpaugh *et al.* (1) were included in the calculation. The effect of anomalous scattering by the iron and sulfur atoms was also taken into account. Using the amplitudes of the calculated structure factors (F_{calc}) for all the 5005 reflections reported in the x-ray diffraction study (1), and the observed amplitudes (F_{obs}), the overall R index of our refined structure was calculated as

$$R = \frac{\sum_{5005} ||F_{calc}| - |F_{obs}||}{\sum_{5005} |F_{obs}|} \quad [3]$$

TABLE 1. *Energy contributions at each stage of refinement of rubredoxin*

Energy (kcal/mole)	Stage I	Stage II (1st iteration)	Stage II (2nd iteration)	Stage III*	Stage III†
E_{NB}	2.77×10^8	1227‡	183‡	−143‡	−145‡
E_{HB}	7.6×10^3	164‡	70‡	−27‡	−27‡
E_{EL}	−471	−511	−498	−515‡	−515‡
E_{ROT}	332	313‡	300‡	206‡	202‡
E_{TOT} (without E_{loop})	2.77×10^8	1193	55	−479	−485
E_{loop}	375	459	114‡	20‡	17‡
E_{TOT} (with E_{loop} included)	2.77×10^8	1652	169	−459	−468

* Using the x-ray geometry for the active site.

† Using the modified geometry for the active site.

‡ Contribution to the energy minimized during this stage of the refinement.

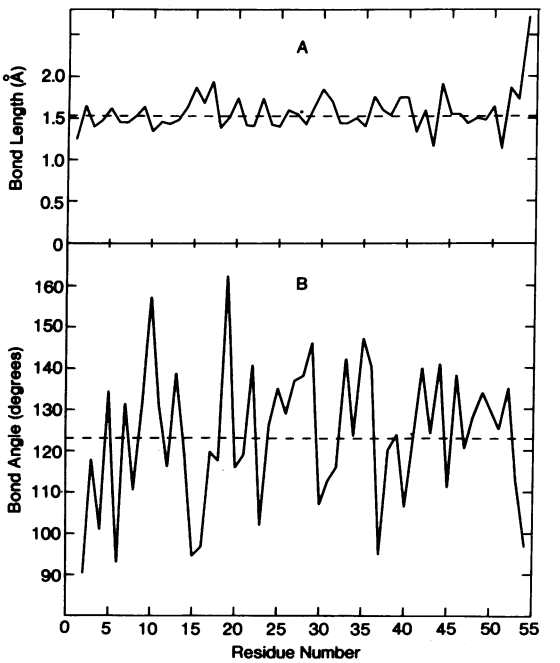

FIG. 1. (*A*) Deviations of the C^α–C′ bond lengths [calculated from the x-ray coordinates of rubredoxin (1)] from the "standard" value of 1.53 Å (shown as a horizontal line). (*B*) Similar plot for deviations of the C′–N–C^α bond angles from the "standard" value of 123° (shown as a horizontal line).

RESULTS

A. Energy contributions at various stages of refinement

The magnitudes of the total energy and its various contributions at each stage of refinement are summarized in Table 1.

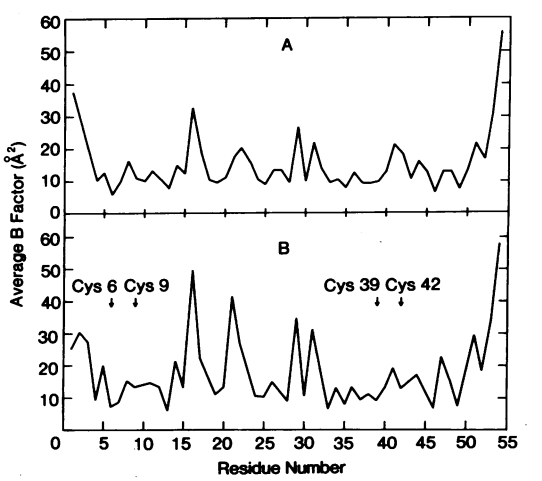

FIG. 2. B (temperature) factors (1) for backbone (*A*) and side-chain (*B*) heavy atoms.

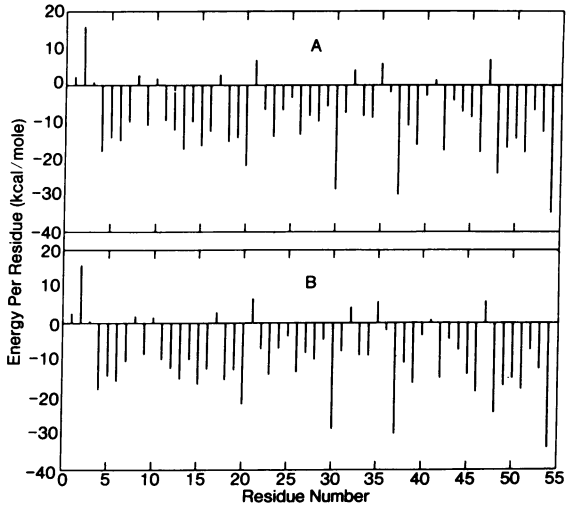

FIG. 3. Plots of total energy ($E_{NB} + E_{HB} + E_{ROT} + E_{EL}$) of each residue of rubredoxin after stage III refinement, calculated as one-half of the sum of the energies of interaction with all other atoms of the entire protein which are within the normal cut-off distances (4). *A*, for x-ray and *B*, for modified active-site geometry.

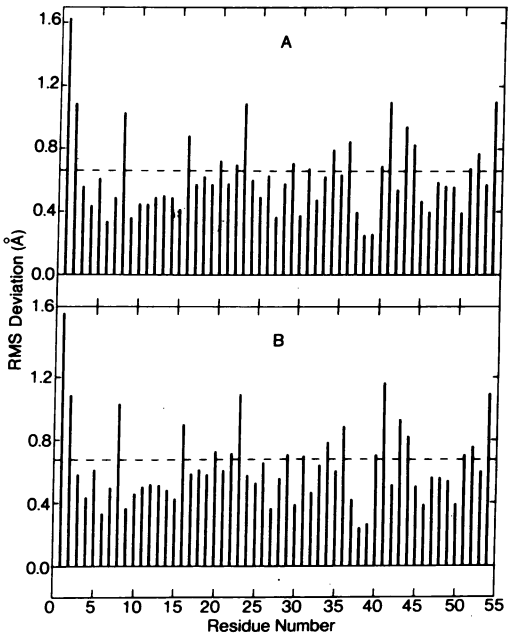

FIG. 4. Plot of RMS deviation of each residue of rubredoxin after stage III refinement. The overall RMS deviation is shown as a broken line. *A*, for x-ray and *B*, for modified active-site geometry.

The last two columns correspond to the x-ray and modified geometry, respectively, around the active site. At the end of stage I, the structure (constrained to have standard bond lengths and bond angles, with a close fit to the x-ray structure) has severe overlaps, as can be seen from the high values of E_{NB} and E_{HB}. The value of E_{loop} at the end of stage I reflects the departure of the C^{β}–S bond lengths and the C^{α}–C^{β}–S and C^{β}–S–Fe bond angles from their equilibrium values (see Table I of documentation).

The high energy obtained at the end of stage I is largely decreased during the first iteration of the stage II procedure. No loop closing potential was used during the first iteration of stage II, and thus it is not surprising that E_{loop} increases in this step. However, during the second application of the stage II procedure, in which most of the strong overlaps are relieved, the inclusion of E_{loop} leads to better closure of the loops, i.e., the C^{β}–S bond lengths and the C^{α}–C^{β}–S and C^{β}–S–Fe bond angles approach their equilibrium values.

In stage III, the energy is lowered considerably, for both choices of the active-site geometry (last two columns of Table 1). Both active-site geometries lead to similar energies (Table 1) and similar dihedral angles (see Tables II and III of documentation). The nonbonded and hydrogen-bond energies decrease to negative values in stage III. However, 11 pairs of atoms still are in fairly close contact, contributing nonbonded energies greater than 1 kcal (see Table IV of documentation). The highest contribution from such contacts is 3.33 kcal, arising from a contact between C^{α} of Gly 43 and the carbonyl oxygen of Cys 39. The electrostatic energy, which was minimized only in stage III, has decreased slightly in this stage. However, E_{ROT} decreases significantly in stage III.

The total energy ($E_{NB} + E_{HB} + E_{ROT} + E_{EL}$) is plotted in Fig. 3. The net energy for most of the residues is negative; the highest energy occurs for Lys-2, and arises mainly from an unfavorable rotational contribution from χ_1.

B. Deviations from the x-ray coordinates

A protein structure, obtained by refinement from x-ray diffraction data, is not only expected to have a low energy, but also to remain close to the original x-ray structure. The root mean square (RMS) deviations from the x-ray co-

TABLE 2. *RMS deviations from x-ray coordinates and average deviations of dihedral angles from their x-ray values at the consecutive stages of the refinement*

	Stage I	Stage II (1st iteration)	Stage II (2nd iteration)	Stage III*	Stage III†
Backbone RMS dev., Å	0.53	0.59	0.55	0.61	0.61
Side chain RMS dev., Å	0.80	0.78	0.77	0.75	0.76
Overall RMS dev., Å	0.63	0.66	0.64	0.66	0.67
Average backbone dihedral angle deviation	20°	20°	20°	18°	18°
Average side-chain dihedral angle deviation	38°	38°	32°	34°	33°

* Using x-ray geometry for the active site.
† Using the modified geometry for the active site.

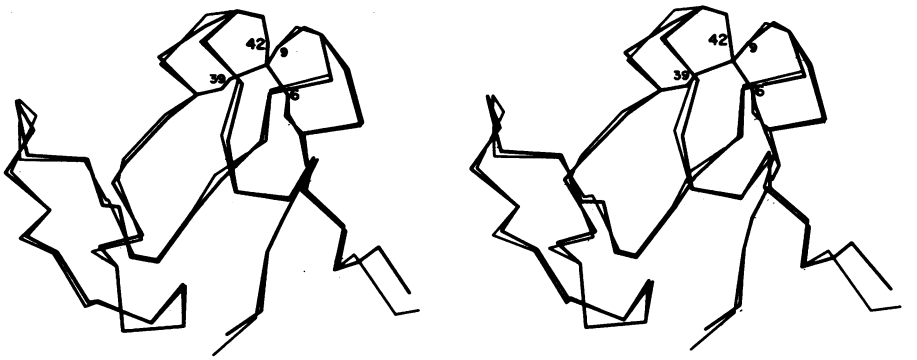

FIG. 5. Stereoscopic view of rubredoxin, showing the refined structure (heavy lines) superimposed on the x-ray structure (light lines). Only the C^{α}, Fe, and cysteine side-chain atoms are shown. The numbers correspond to the sulfur atoms of the four cysteine residues.

ordinates for the backbone atoms, side-chain atoms and the RMS deviation for all atoms, calculated after each stage of the refinement are listed in Table 2.

When we started the second stage of refinement, we became aware that the amino-acid sequence derived by Watenpaugh *et al.* (1, 9) from the x-ray structure differed in seven places from the sequence obtained by McCarthy (10) by chemical analysis. Ile 8, Ile 24, and Ile 44 all became valines, and we simply ignored the C^{δ} x-ray coordinates. For the other changes (Val 12 became Ile; Glu 21 became Asp; Val 28 became Thr; and Ile 41 became Leu) we have used the existing x-ray coordinates for calculating the fitting potential and RMS deviations for all atoms whose deviations from the calculated positions did not exceed 0.80 Å after the stage I procedure was reapplied to these side chains.

In Fig. 4, the RMS deviations are plotted against residue number. It is apparent that the atoms of about 12 residues have rather large RMS deviations and contribute significantly to the overall RMS deviations. As can be seen in Fig. 2 from the average temperature factors per residue plotted against residue number, the x-ray coordinates of most of these residues are associated with rather high average temperature factors and thus are probably less accurate because of thermal motion.

C. Variations of dihedral angles

The backbone and side-chain dihedral angles obtained at the end of the refinement of rubredoxin are listed in Tables II and III of the documentation. The average deviations of the backbone and side-chain dihedral angles from the dihedral angles computed from the x-ray data have been calculated after each stage of the refinement and are listed in Table 2. The deviations for the side-chain dihedral angles are about twice as large as the deviations of the backbone dihedral angles. The decrease of the backbone dihedral angles deviation obtained at the end of the third stage of refinement is a consequence of the reorientation of the computed structure compared to the x-ray structure of rubredoxin, i.e., by changing the coordinate system. By that procedure, the deviations of the backbone dihedral angles of Lys 3 were reduced significantly.

The R index, calculated for the computed structure using the modified active-site geometry, is 0.37, which may be compared to the value of 0.126 of Watenpaugh *et al.* (1). Our computed value of R may also be compared with values of 0.44 for subtilisin BPN′ (11), 0.53 for cytochrome b_5 (12), and a range of 0.32–0.52 for tosyl-α-chymotrypsin (13). Thus, it appears that our refined *low-energy* structure of rubredoxin yields the observed structure factors with a reliability in the range of that obtained for other proteins.

A stereoscopic view (14), comparing the refined and observed backbone and active-site structures, is shown in Fig. 5. Further documentary material, including the cartesian coordinates and dihedral angles of the computed structure of rubredoxin, are available as document No. NAPS-02422 of the ASIS National Auxiliary Publication Service, c/o Microfiche Publications, 305 E. 46th street, New York, N.Y. 10017. A copy may be secured by citing the document number and by remitting $1.50 for microfiche or $5.00 for photocopies. Advance payment is required. Make checks or money orders payable to Microfiche Publications.

DISCUSSION

The computed value of R and the results shown in Tables 1 and 2 indicate that our objective of refining the x-ray structure of *Clostridium pasteurianum* rubredoxin has been achieved. Our final structure has the characteristic bond lengths and bond angles derived from crystals of its constituent amino acids, exhibits a very low conformational energy, and yet remains very similar to the x-ray structure.

In considering the reliability index R, it should be pointed out that Watenpaugh *et al.* (1) achieved a very low value of R by determining the structure of rubredoxin using a technique generally applied to *small* molecules; i.e., Watenpaugh *et al.* (1) treated the coordinates of the atoms and the temperature factors as parameters to be adjusted in their refinement. When this technique is applied to small molecules, the ratio of the number of reflections to the number of parameters to be adjusted is at least 10:1, thereby increasing the reliability of the reported atomic positions, and the R index is very low for a reliable structure. In the case of a protein, the ratio is smaller than 10:1 (and is 2.2:1 in the case of rubredoxin), and the crystallographer usually circumvents this problem by reducing the number of parameters to be adjusted, i.e., by assuming values for the bond lengths, bond angles, and planarity of the peptide group (usually with the aid of Kendrew models) in fitting the electron density maps. Thus, R values generally reported for proteins are in the range quoted in the *Results* section. With the procedure used for rubredoxin (1), a low R value was obtained but, as indicated in Fig. 1, the structure shows many large departures of the bond lengths and bond angles from their usually accepted values. It remains to be seen whether this situation arises because of the relatively small number of reflections measured in the determination of the structure of rubredoxin. Any refinement which

imposes constraints (2–4, 15, 16) will move the atoms from the positions reported in the x-ray study to produce a conformation having normal bond lengths and bond angles, and consequently will increase the value of the R index (as well as the reported RMS deviations). Thus, our value of $R = 0.37$, compared to that of 0.126 of Watenpaugh *et al.* (1) is not unexpected, and lies in the range of R values obtained for other proteins.

The major contribution to the RMS deviations of our refined coordinates from the x-ray coordinates comes from those areas of the molecule that undergo substantial thermal motion in the crystal (compare Figs. 2 and 4). Indeed, we observe larger deviations for the atoms belonging to the N- and C-terminal residues as well as for the side chains lying on the surface of the molecule. The lower accuracy of the x-ray coordinates of those atoms induces large deviations right at the start of the computations (stage I) when we constrain the structure to adopt a "standard" geometry. Nevertheless, the total energy obtained at the end of stage III is very low.

The geometry around the active site has been preserved without giving rise to larger-than-average deviations from their x-ray positions of the atoms of the cysteine residues. The energies observed for the refined structures corresponding to two alternative geometries for the active site are very close to each other, and do not allow us to select one of them over the other.

We are indebted to Dr. L. H. Jensen for providing the x-ray coordinates and structure factors in advance of publication; to Dr. G. N. Reeke, Jr., for providing the computer program to compute the structure factors and R index (and for help in its use); and to Drs. D. Gabel and P. K. Ponnuswamy for helpful discussions of the energy calculations. This work was supported by research grants from the National Science Foundation and from the National Institutes of Health.

1. Watenpaugh, K. D., Sieker, L. C., Herriott, J. R. & Jensen, L. H. (1973) *Acta Crystallogr.* **B29**, 943–956.
2. Warme, P. K., Go, N. & Scheraga, H. A. (1972) *J. Comput. Phys.* **9**, 303–317.
3. Warme, P. K. & Scheraga, H. A. (1973) *J. Comput. Phys.* **12**, 49–64.
4. Warme, P. K. & Scheraga, H. A. (1974) *Biochemistry* **13**, 757–767.
5. Gibson, K. D. & Scheraga, H. A. (1967) *Proc. Nat. Acad. Sci. USA* **58**, 1317–1323.
6. McGuire, R. F., Momany, F. A. & Scheraga, H. A. (1972) *J. Phys. Chem.* **76**, 375–393.
7. Fletcher, R. (1970) *Comput. J.* **13**, 317–322.
8. Fletcher, R. (1971) *IMA/NPL Symposium on Numerical Methods for Unconstrained Optimization, Jan. 1971.*
9. Watenpaugh, K. D., Sieker, L. C., Herriott, J. R. & Jensen, L. H. (1971) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 359–367.
10. McCarthy, K. F. (1972) Ph.D. Dissertation, George Washington Univ., quoted by Eaton, W. A. & Lovenberg, W. (1973) in *Iron-Sulfur Proteins*, ed. Lovenberg, W. (Academic Press, New York), pp. 131–162.
11. Alden, R. A., Birktoft, J. J., Kraut, J., Robertus, J. D. & Wright, C. S. (1971) *Biochem. Biophys. Res. Commun.* **45**, 337–344.
12. Matthews, F. S., Levine, M. & Argos, P. (1972) *J. Mol. Biol.* **64**, 449–464.
13. Birktoft, J. J. & Blow, D. M. (1972) *J. Mol. Biol.* **68**, 187–240.
14. Warme, P. K., Tuttle, R. W. & Scheraga, H. A. (1972) *Comput. Progr. Biomed.* **2**, 248–256.
15. Levitt, M. & Lifson, S. (1969) *J. Mol. Biol.* **46**, 269–279.
16. Diamond, R. (1971) *Acta Crystallogr.* **A27**, 436–452.