Refinement of Purothionins Reveals Solute Particles Important for Lattice Formation and Toxicity. Part 1: α1-Purothionin Revisited

By Usha Rao, Boguslaw Stec and Martha M. Teeter*

Department of Chemistry, Boston College, Chestnut Hill, MA 02167, USA

Abstract

The three-dimensional structure of α1-purothionin (α1-PT), a wheat-germ protein and a basic lytic toxin, was previously solved by molecular-replacement methods using an energy-minimized predicted model and refined to an R factor of 21.6% [Teeter, Ma, Rao & Whitlow (1990). *Proteins Struct. Funct. Genet.* **8**, 118–132]. Some deficiencies of the model motivated us to revisit the structure and to continue the refinement. Here we report a significantly improved structure refined to an R factor of 15.5% with excellent geometry. The refinement of this relatively low resolution structure (~2.8 Å) is well suited to test the limitations of classical methods of refinement and to address the problem of overfitting. The final structure contains 434 atoms including 330 protein atoms, 70 waters, three acetates, two glycerols, one *sec*-butanol and one phosphate. The key solute molecules (acetate ion and phosphate ion) play a crucial role in the lattice formation. Phosphate and glycerol found in the structure may be important for biological activity of the toxins.

1. Introduction

α1-Purothionin (α1-PT) belongs to a family of low molecular weight highly basic toxins which are widespread in the plant kingdom. The members of the thionin family display 30–50% amino-acid sequence identity to crambin (Apel, Bohlmann & Reimann-Philip, 1990). These toxins show a broad range of toxic effects exerted on cell membranes of bacteria, fungi and higher organisms (Bohlman & Apel, 1991). We have attempted to understand the mechanism of toxicity through structural studies of several member toxins. As a part of a continuing program of homology-based structure prediction tested by solving the X-ray structure (Rao & Teeter, 1993), the three-dimensional structure of α1-PT has already been obtained in our laboratory (Teeter *et al.*, 1990). Moreover, we have recently solved the structure of β-purothionin (see accompanying paper, Stec, Rao & Teeter, 1995).

Because of the high homology of α1-PT to crambin, their structures maintain similarity in overall shape and architecture. Crucial differences, however, have been ob-

---

* To whom correspondence should be addressed.

Table 1. *Distribution of R factors and completeness of the data for α1-PT*

Total number of reflections used in the refinement was 1168 (10–2.5 Å and $F > 2\sigma$).

Resolution range (Å)	Number of reflections	Completeness Shell (%)	Sphere (%)	R factor (initial model) Shell (%)	Sphere (%)	R factor (final model) Shell (%)	Sphere (%)
4.83–10.0	248	95.4	95.4	29.9	29.9	21.3	21.3
3.91–4.83	232	94.3	94.9	19.5	24.0	14.5	16.5
3.44–3.91	196	84.9	91.7	18.0	21.8	13.7	15.8
3.13–3.44	187	78.3	88.5	18.9	21.0	14.0	15.5
2.91–3.13	151	63.8	83.7	20.7	20.9	13.6	15.3
2.75–2.91	88	39.9	76.7	22.2	20.9	12.4	15.1
2.61–2.75	31	13.2	68.1	23.4	21.1	18.3	15.2
2.50–2.61	35	15.3	61.5	27.9	21.6	21.8	15.5

served in the crystal packing and intermolecular contacts, due to the basic nature of α1-PT. Despite the relatively low resolution of the data (Table 1), the structure refined to an R factor of 21.6% (Teeter *et al.*, 1990) and revealed clear water molecules and other solute molecules.

This initially refined model was, however, not fully satisfactory and a revisitation to the structure refinement seemed necessary because: (*a*) the final difference map $(F_o - Fc)$ was not featureless, with a few peaks above 3σ in the asymmetric unit. (*b*) Several side chains had rare high-energy conformations. (*c*) The R factor (21.6%) was rather high, considering the maximal resolution of the data (2.5 Å, although only 15% of the data above 2σ were collected at 2.8–2.5 Å). (*d*) Certain side chains had conformations that resulted in unfavorable electrostatic interactions.

The low-resolution data presented a challenge for checking the limitations of classical refinement methods. We have asked whether it was possible to obtain a reliable model with relatively low resolution data. Additionally, what would happen when external information such as NMR data or information from a subsequent refinement of β-PT to higher resolution is compared and incorporated into further refinement. Hence, we undertook further refinement of the structure to test the limits of methods with the hope of obtaining a better model.

This paper presents a detailed description of the additional refinement of the α1-PT structure and emphasizes the process leading to a better model. This paper is organized into two major parts. In the first

---

part (*Methodology and progress of structure refinement*), various methods used on all stages of the structure refinement are covered. A few comments on *X-PLOR* refinement are also made. Finally the usefullness of the *R*-free method for the low-resolution structure is evaluated. In the second part (*Discussion*), we discuss the improvements in the stereochemistry following the refinement and the details of the crystal packing involving several solute molecules. We note these solutes and hypothesized about their role in the toxic function of these proteins, as is discussed further in the following paper.

2. Methods

2.1. α_1-PT crystals and the initial model

α_1-PT crystals were obtained by the vapor-diffusion method using 15% *sec*-butanol as precipitant in sodium cacodylate buffer at pH 5.9 or in Tris buffer at pH 8.0 (Teeter *et al.*, 1990). The protein crystallized in the space group *I*422 ($a = b = 53.59$, $c = 69.79$ Å) and the structure has been solved by molecular-replacement methods using the energy-minimized predicted model previously derived by Whitlow & Teeter (1985). The structure was initially refined with *PROLSQ* (Hendrickson & Konnert, 1980) to an *R* factor of 21.6% (Teeter *et al.*, 1990). It contained 330 protein atoms, 39 water molecules and four acetates. Refinement described in this paper begins with this structure.

2.2. Programs and computer devices

The initial model with *R* factor 21.6% was further refined with the restrained least-squares program *PRO-LSQ* (Hendrickson & Konnert, 1980) on a Microvax II computer and all model building was carried out on an Evans and Sutherland PS390 interactive graphics system using the program *FRODO* (Jones, 1985). *X-PLOR* refinement *Version* 3.1 (Brünger, 1992*a*) was also conducted but on the mini supercomputer Stellar GS1000 (Stardent Inc.).

3. Methodology and progress of structure refinement

3.1. Guidelines for the further refinement

Certain expectations for the α_1-PT model which had not been met by initial refinement guided the course of further refinement. First, the torsional angles of the side chains should be close to the allowed rotamer conformations* that are statistically most preferred (Ponder & Richards, 1987). This rotamer assumption is supported by the observation of the narrow distribution(s) of side-chain conformation for oligopeptides (Ashida, Tsunogae, Tanaka & Yamane, 1987) and from the highly refined (full-matrix least squares) crambin structure as well (Stec, Zhou & Teeter, 1995). Secondly, since the protein is positively charged, there should be considerable

Table 2. *Summary of restrained least-squares refinement parameters*

	Initial refinement		Later refinement	
	Target σ	Actual r.m.s. deviation	Target σ	Actual r.m.s. deviation
Average ΔF		465.7		332.52
R factor*		0.216		0.155
R_{error}†		0.073		0.073
No. of structure factors		1168		1168
Structure-factor weight‡	250	373.9	150	252.6
R.m.s. deviations from ideal distances (Å)				
Bond distance	0.02	0.012	0.02	0.012
Angle distance	0.03	0.030	0.03	0.026
Planes 1–4 distance	0.04	0.029	0.04	0.028
R.m.s. deviation from planarity (Å)	0.015	0.005	0.015	0.005
R.m.s. deviation from ideal chirality (Å^3)	0.15	0.135	0.15	0.112
R.m.s. deviation from permitted contact distances (Å)				
For planar group (0 or 180)	2.0	0.9	3.0	1.0
For staggered group (60 or 180)	15.0	25.4	18.0	18.1
For orthogonal group (90)	20.0	19.9	20.0	9.5
R.m.s. deviation of the isotropic thermal factor differences (Å^2)				
For main chain bond	2.0	0.40	2.0	0.53
For main-chain angle	2.5	0.69	2.5	0.86

$* R = \sum_h ||F_o| - |F_c||/ \sum_h |F_o|.$

$\dagger R_{error} = \sum_h |\sigma(F_o)|/ \sum_h |F_o|.$

‡ The weight chosen for the structure-factor refinement, the 'target σ' of ΔF, was modeled by the function $\sigma = (1/\sigma)^2$ with $\sigma = 250 + [(-999)(\sin\theta/\lambda - 1/6)]$ for the initial refinement. For the later refinement, the σ value used was: $\sigma = 150 + [(-1100) \times (\sin\theta/\lambda - 1/6)]$.

negatively charged solute in the crystal as well. Acetate ions and phosphate ions, which had not been added to the crystallization dip but were left from the purification with ammonium acetate in phosphate buffer, may serve such a purpose. Thirdly, it was argued that the model should have more localized water molecules than it currently had. Fourthly, certain hydrogen bonds are expected to be conserved in the structures of this family of plant proteins. Based on these guidelines, several conformational changes were made on the model which vastly improved the quality of the structure

The refinement was initiated from the model (21.6%) presented in Teeter *et al.* (1990) and proceeded in ten stages. Each stage consisted of a graphics modeling session and several (usually ~20) cycles of least-squares refinement until convergence. More than 200 cycles of least-squares refinement were conducted which resulted in geometric improvement of the model and a significantly lower *R* factor. The first three stages concentrated mainly on improving the protein side-chain conformation with some effort made to model the waters and *sec*-butanol solvent. More attention was given to solute/solvent modeling in the next five stages of the refinement where molecules like phosphate, acetates and glycerols are successfully modeled. The following stage was a check procedure performed to test the authenticity

of waters in the model as will be described below [OMIT maps (Bhat, 1988; Brünger, 1992c)]. A comparative listing of restrained least-squares refinement parameters for both initial and final models is given in Table 2. Finally, a comprehensive study of R-free method (Brünger, 1992a,b) was carried out. The study was intended to check the correctness of the atomic contents of the model as well as to check the effectiveness of the R-free method when applied to this low-resolution high-symmetry structure.

Below are presented different methods we have used during the refinement to improve the structure and position individual atoms.

3.2. Sample structural changes

Below various structural modifications made are discussed under several subheadings and examples are presented.

3.2.1. *Ideal-geometry-guided modifications.* The statistical rotamer library of preferred conformations by Ponder & Richards (1987) was used to correct rare side-chain conformations. These types of changes were performed on the residues Lys1, Leu8, Cys29–Cys12 and Lys23. For example, Lys1 was changed from the +, t (+60°, 180°; see footnote) to the t, t conformation leading to a much better electron density (Figs. 1a and 1b). In Fig. 1, the side chains before and after such a change are superimposed on the appropriate density.

One of the four disulfides (Cys29–Cys12) was in a rare conformation (145, −86, −170, 64, −150) (Fig. 1c) which we had rationalized as allowed in the initial refinement for energetic reasons (Pullman & Pullman, 1974). The electron density for this disulfide was good and had persisted in OMIT maps (Bhat, 1988) calculated after ten cycles of refinement. However, by changing Cys29 from the t to the − conformation and a slight change in the Cys12 conformation, this disulfide attained the same conformation as Cys16–Cys25 (t,−90,−90,−90,−) and fit well into the density (Fig. 1d).

Lys23 had no continuous density for its side chain. It had been initially modeled into the density of its symmetry mate to avoid close contacts between like-charged groups (Fig. 1e). But once the side chains mentioned above were corrected, improved phases permitted determination of a new conformation for Lys23 (−,+ to −,t) (Fig. 1f).

3.2.2. *Hydrogen-bond guided modifications.* The requirement of a suitable geometry for efficient hydrogen bonding was used to improve the conformations of several side chains. The corrections were not obvious from density maps alone and required additional external information. In crambin Arg10 is responsible for pinning

down the N and C termini by hydrogen bonding to Thr2 OG1 and the carboxy terminus. This is also a key interaction in the toxins and necessary for their folding (Rao, Hassan & Hempel, 1993). In α_1-PT, by changing the χ_1 of Ser2 to the + conformation, a good hydrogen bond to Arg10 (see Fig. 2) was created.

Another hydrogen-bond guided modification involved residues Asn11 and Asn14. After flipping of the headgroup of the Asn14 side chain, the correct ND2···OD1 hydrogen bond is formed. The changes above improved the quality of the structure as is reflected in the lower R factor of 18.7%.

3.2.3. *Charge-neutralization guided modifications.* Several peaks above 3σ which persisted in the difference maps after the first stage of modeling were fit by
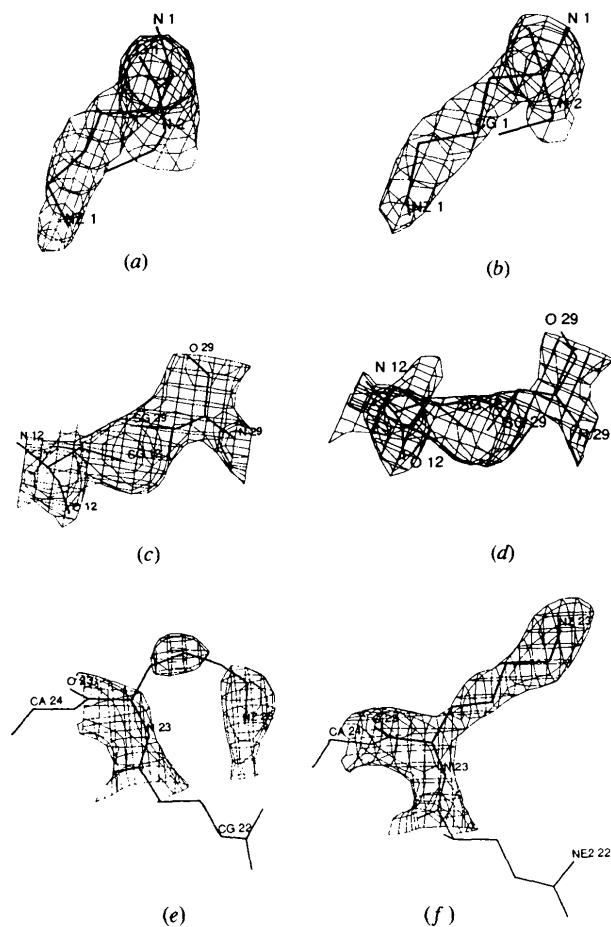

Fig. 1. Selected residues superimposed on corresponding $2F_o - F_c$ maps showing the initial incorrect conformations and the correct conformations after the later refinement. Electron-density map is contoured at the 2σ level. (a) Lys1 at the end of the initial refinement in +, t conformation. (b) Lys1 at the end of the later refinement in t, t conformation. (c) Cys29–Cys12 making a rare parallel bridge in the initial refined model. (d) Perpendicular Cys29–Cys12 bridge in the final model. (e) Rare conformation (−, +) of Lys23 poorly fit to the density in the initial refined model. (f) Lys23 in a new conformation (−, t) with a good density-fit after the later refinement.

---

*The notations used in this paper for the description of side-chain conformations are t = 180°, + = +60° and − = −60°. These definitions are the same as those used by Ponder & Richards (1987). The corresponding IUPAC convention is t = 180°, g^- = +60° and g^+ = −60°.

trial and error with different polar and negatively charged moieties. Careful interpretion of the positive and negative peaks of the difference maps and temperature factors of individual atoms after the refinement allowed us to limit the possibilities and model the best fitting moieties. The shape of the density at low resolution cannot be an ultimate guide in the decision making process. Final interpretation of the difference density came from comparison to the electron density in the similar structure of β-PT at much higher resolution (1.7 Å) and from other external sources such as NMR spectroscopy.

3.2.4. *Use of temperature factors and geometry of the refined moiety as a guide.* Two peaks above 3σ were positioned on a twofold axis. At those sites trial molecules were introduced. The correct moiety should match the electron density the best (temperature factors as well as the shape of the density). An incorrect trial moiety should refine with deviations from expected vibrational factors and geometry. Those deviations suggest more appropriate trial molecules.

Initially, we introduced water molecules at these two sites but the refinement to very low temperature factors and observed positive difference density peaks convinced us to model a heavier moiety. Subsequently, in both densities, an acetate ion was placed on a twofold axis with 0.5 occupancy such that its CH_3—C bond coincided with the twofold axis. Even though the acetate was not used in crystallization its presence can be a remnant from purification with ammonium acetate.

At the first site (close to Arg10), the two O atoms of the acetate which was introduced interacted with the NZ atoms of the symmetry-related Lys1 residues (Fig. 3a). This disordered acetate refined to a low temperature factor (~ 3.0 Å^2) compared to protein temperature factors of ~ 20 Å^2 and changed from the initial orientation.

At the first site, the refinement shifted two of the O atoms and the two methyl groups of the resulting disordered acetate ions so that the molecule had tetrahedral (T_d) rather than C_{2v} symmetry. This observation suggested that this density should be occupied by a tetrahedral molecule, like a cacodylate (CAC) which had been used as the crystallization buffer. When a CAC molecule was introduced, however, it had a very high temperature factor indicating that an atom lighter than the central As atom should be there instead.

Next a phosphate ion was introduced at the first site. Since phosphate was not used in the crystallization procedure, it had not been considered as a counterion before. But analysis of the protein's history indicated that phosphate and acetate were used in the purification. The phosphate group refined well at this special position (Fig. 3b). The presence of phosphate was confirmed by the results of three independent experiments. When the lyophilized protein was dissolved in the buffer without phosphate we detected a sharp peak in the ^{31}P NMR experiment which was shifted downfield from the phosphoric acid standard by 0.987 p.p.m. indicating a bound phosphate ion (unpublished observation)
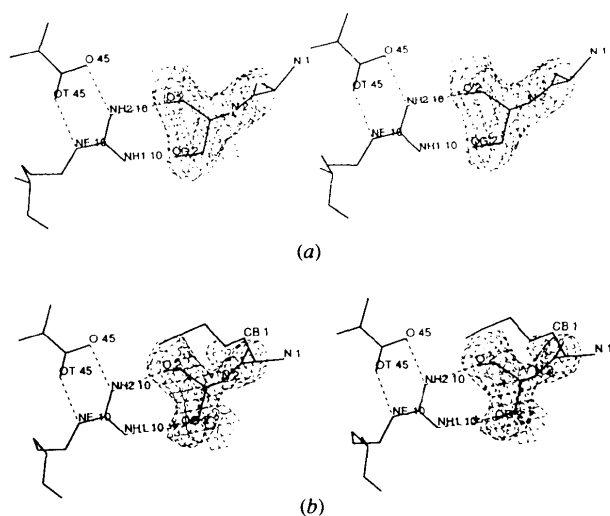

(a)

(b)

Fig. 2. Stereoview of conserved intramolecular hydrogen-bonding interaction between Ser2–Arg10 and the C terminal, with $2F_o - F_c$ map contoured at 2σ level superimposed on Ser2. (a) The map and the model from the initial refined model. (b) Those from the later refined model showing Ser2 in a favorable conformation to make a hydrogen bond with NH$_2$ of Arg10. Hydrogen-bond interactions are shown in broken lines.
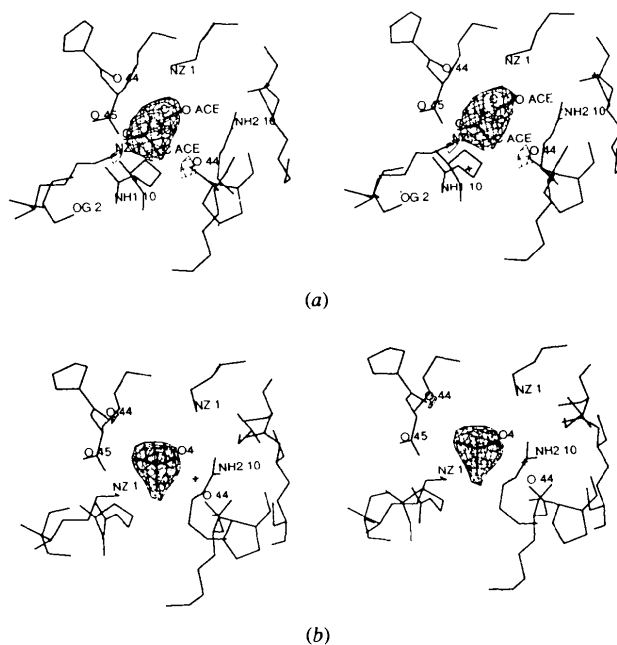

(a)

(b)

Fig. 3. The identification of a phosphate ion in the lattice. (a) The $2F_o - F_c$ map showing an acetate modeled on a dyad [110] in the initial refined model. The acetate and its symmetry-mate diverged to tetrahedral appearance. (b) The later refined model with a phosphate ion modeled at this position showing a better fit to the $2F_o - F_c$ density and also making a good interactions with Lys1 and Pro44. The view is the same as in (a).

which must have survived the intensive desalting steps. Secondly, after learning of the necessity of phosphate for crystallization, we added phosphate to a batch of β-purothionin from which we could not get crystals and readily obtained crystals. Thirdly, phosphate was subsequently modeled and refined well in a similar site in β-PT at 1.7 Å resolution (Stec, Rao & Teeter, 1994).

The fact that this phosphate group survived all the desalting steps during purification implies that it must have been bound very tightly and specifically. This fact has relevance for phospholipid binding to thionins and their toxic activity (Markman et al., 1993).

The second peak which was positioned on an intersection of two twofold symmetry axes was modeled in the same way (Fig. 4a) and it was concluded that an acetate ion with 0.25 occupancy was the molecule of choice at this site. The correctness of this modeled acetate is evident from the well behaved refinement and the improved density fit (Fig. 4b).

3.2.5. *Modifications assisted by the β-purothionin structure refinement.* The subsequent and parallel refinement of the β-purothionin structure at 2.25 Å and later at 1.7 Å resolution (Stec, Rao & Teeter, 1995) also provided an aid in recognizing and modeling the correct solvent/solute in the α_1-PT structure. One such example is shown in Fig. 5. Fig. 5(a) shows two acetates in a binding site formed by Ser2, Tyr13, Arg17 and Gln22. Although these acetates refined well, they were too close to each other with their methyl groups at covalent-bonding distances (~ 1.6 Å). This fact indicated the existence of a higher hydrocarbon, with a similar number of atoms to two acetates. The conditions used for crystallizing α_1-PT did not contain sugars or alcohol with this number of atoms, although the data were collected on the crystals obtained from sec-butanol, a larger alcohol. The structure of β-PT refined at higher resolution (2.25 Å) than α_1-PT showed clearly the existence of similar moiety at the same binding pocket. In β-PT, an MPD molecule (2-methyl-2,4-pentanediol) was used in the crystallization medium and was initially fit and refined in the β-PT structure. We modeled a comparable MPD molecule at the site of the two acetates in the α_1-PT.

Upon further refinement of the β-PT structure, it became clear from distortions of the tetrahedral group of the disubstituted C atom in MPD and elevated temperature factors that this assignement was incorrect. After careful examination of refinement results of α_1- and β-PT and knowledge of common lipid groups in plants (Voet & Voet, 1990), we decided to model a lighter hydrocarbon, a precursor of phospholipid synthesis – glycerol. Glycerol refined very well in both structures
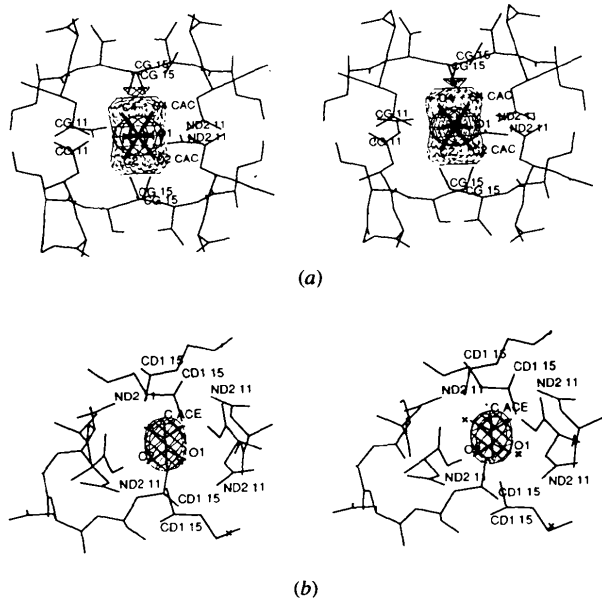

(a)

(b)

Fig. 4. (a) A disordered CAC molecule modeled at special position ($\frac{1}{2}$,0,$\frac{1}{4}$) superimposed on the corresponding $2F_o - F_c$ density (solid lines). Also superimposed is the negative difference Fourier map (contoured at 2σ level, in broken lines), indicating the absence of methyl plane. View shown is the projection down the [110] axis. (b) A 90° rotated view of the similar map from the later refined model, showing the disordered acetates at the special position making good interaction with four symmetry-related Asn11 residues.
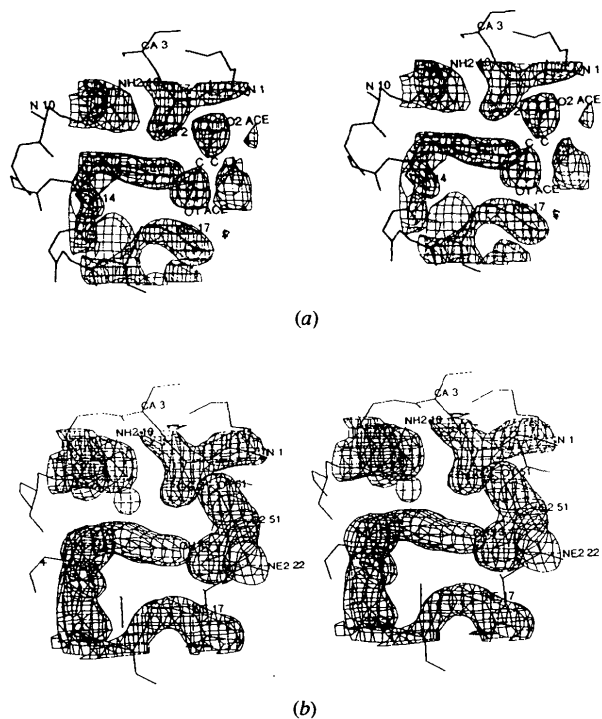

(a)

(b)

Fig. 5. (a) A section of the $2F_o - F_c$ map in stereo showing two acetates at a special position which refined to a covalent bonding distance from each other (from the model G). (b) A similar section from the final model after this density is correctly modeled as glycerol which interacts with Tyr13, Arg17 and Ser2. Electron density is contoured at the 2σ level.

($B \simeq 35$ Å^2 in α_1-PT and $B \simeq 50$ Å^2 in β-PT), eliminating all the previously described problems. Subsequently, we collected higher resolution data for β-PT, and the refinement at 1.7 Å clearly confirmed our modeling.

The glycerol shows an excellent fit to the electron density, which covers all the atoms (see Fig. 5b). We cannot account for the possible source of glycerol in crystallization solution although the subsequent NMR studies on binding of glycerol-3-phosphate in our laboratory showed relatively high affinity for glycerol-like molecules. However, we can associate glycerol with the reported activation of phospholipases D and A$_2$ (Evans, Wang, Shaw & Vernon, 1989). Besides of being a precursor of phospholipid synthesis, the glycerol could have been obtained by cleavage of phosphatidyl glycerol.

3.2.6. *Use of OMIT maps. OMIT* maps were used several times during the course of the refinement in order to check or establish appropriate positions for the extended side chains (Arg19, Fig. 6a,b, and Lys23). Nevertheless the real importance of the method became apparent when the water structure was investigated.

As the refinement proceeded, water molecules were added to the model when density warranted it and the refinement of the positions was monitored. However, the resultant model had 433 atoms being refined against only 1168 structure factors. That means that the ratio of observations to parameters (four per atom) is considerably less than one. Although use of the restraints should improve the ratio, water molecules are not restrained. The drop in

the R factor could be an artifact of adding several waters and may not reflect the correctness of the model.

Therefore, in order to establish the credibility of water positions, all 70 waters were deleted (16% of the atoms in the structure) and the remaining model refined. Protein side chains to which these water were bound could not be resonably omitted because of this protein's small size. After ten cycles of refinement (R factor = 25.6%), *OMIT* maps were constructed. 85% of the earlier water positions were readily located in the difference Fourier ($F_o - F_c$) map.

Fig. 7(a) shows an example of water electron-density maps around Lys41. Water positions which were deleted are shown by crosses. The peaks in the $F_o - F_c$ map fall on these water positions. Fig. 6(b) shows the corresponding map from the later-refined model showing that these waters when added again refined well. Note that density for the side chain of Lys41 disappeared in the *OMIT* map, but, when the waters were reintroduced, it reappeared again (see Fig. 7b).

These observations establish unequivocally that in small proteins like α_1-PT the solvent contributes enor-
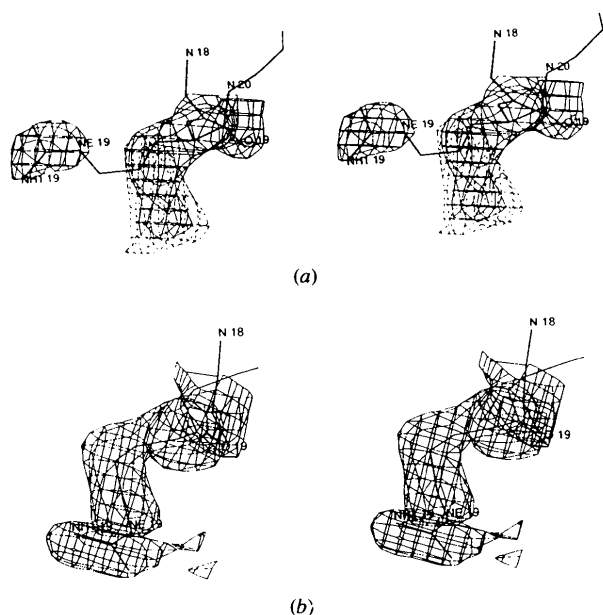

(a)

(b)

Fig. 6. Use of an *OMIT* map for Arg19. (a) $2F_o - F_c$ map (solid lines) around Arg19 before the *OMIT* map, showing the poor density for the side chain. Both $2F_o - F_c$ map and $F_o - F_c$ map (broken lines) with the side chain omitted are indicative of a new conformation for this residue. (b) $2F_o - F_c$ map superimposed on Arg19 modeled in a new conformation in the final model.
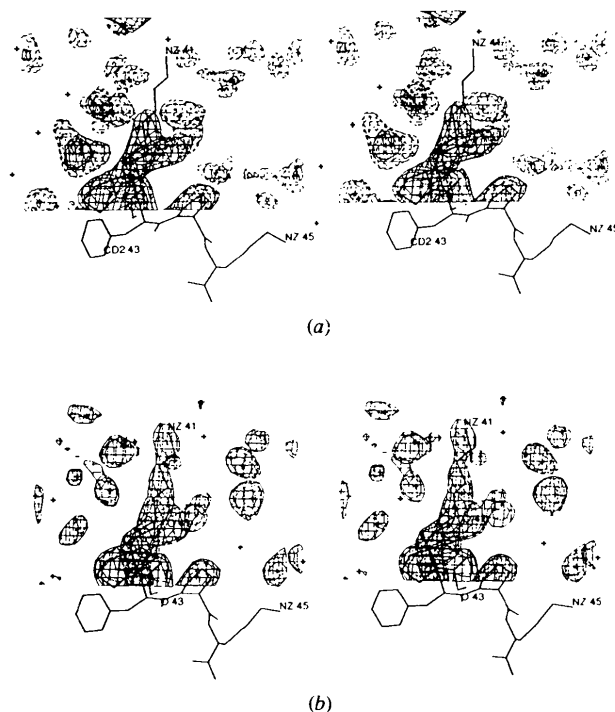

(a)

(b)

Fig. 7. (a) A section of the *OMIT* map produced by deleting all the waters in the model. A section near Lys41 is shown. The $2F_o - F_c$ map contoured at the 2σ level is in solid lines while $F_o - F_c$ map contoured at the same level is in broken lines. Waters which were deleted are shown by plus (+) sign. About 85% of the waters reappeared in the *OMIT* map. It should be noted that the density for Lys41 side chain disappeared in the *OMIT* map altough side-chain atoms were not removed. (b) Similar section of $2F_o - F_c$ map from the later refined model, after reintroducing the waters found in the *OMIT* map (a). Notice that Lys41 now has good density fit.

mously to the phasing and that the modeling of solvent is very important. Solvent contribution to R-free calculations in the high-symmetry space group are described below.

At the end of this extended refinement, the model now has 330 protein atoms, 70 waters, two acetates, one phosphate, two *sec*-butanols and one glycerol molecule with an R factor of 15.5%.

3.3. *X-PLOR refinement of the initial model with R factor = 21.6%*

In order to see whether *X-PLOR* refinement (Brünger, Kuriyan & Karplus, 1987) could replace tedious modeling steps and speed up the convergence, we ran several rounds of refinement. In all the runs, no charges were used on the side chains. The final R factor of *X-PLOR* refinement on the initial refined model without solvent was 25% while the model with solvent usually converged to ~19%. Fig. 8 shows a comparison of the initial refined model and one of the *X-PLOR*-refined models which included all the solvent.

The level of success depended on the temperature used in the simulated-annealing (SA) portion of the *X-PLOR* runs. We tried different temperatures at 1000–3000 K and different weights (wa) for the X-ray term, which value ranged from 50 to 100% as determined by the *CHECK* procedure. Usually, the procedure corrected some conformations but could never improve the geometry of certain side chains and even when corrected manually they departed from the correct positions. The level of divergence depended almost linearly on increase in temperature. In our case, 1000 K produced minimal divergence from the manually corrected conformations and similar R factors to that obtained in *PROLSQ*. Thus, despite a decrease in the R factor, *X-PLOR* refinement can produce incorrect features, as noted by others (Brünger, 1988). Partial responsibility for the departures can be assigned to

the low-resolution data set which in the α_1-PT case involved only 1200 reflections.

3.4. *Evaluation of the structure by the R-free method*

In order to evaluate the correctness of the model and especially the newly found solute molecules, we performed a series of R-free SA calculations (Brünger, 1992*a,b*). In the case of α_1-PT, the process was not entirely straightforward because of the high symmetry of the lattice (*I*422) and the large number of solute molecules positioned on the symmetry elements. Secondly, there were only around 1200 reflections used in the refinement. This meant that only a small number of reflections (~100) were partitioned into the test set (90–10%), which brings into question the statistical validity of the procedure.

In close cooperation with Axel Brünger, we have overcome the technical difficulties and carried out the R-free evaluation. First, we established that the data should be partitioned into the main and test data sets of 85 and 15%, respectively. Also, we had to limit the interactions and fix the atoms positioned on the symmetry elements. We checked a series of models including the protein alone, the protein plus solute molecules positioned on the symmetry elements, plus all the solutes, and, in the end, we gradually increased the number of water molecules, starting from those with the lowest temperature factors. Finally, we checked the whole model with the water molecules fixed, and subsequently released but harmonically restrained with 20, 10, and 5 kcal. All the calculations were carried out with the initial temperature of 1000 K in the SA cycle.

In summary, it is difficult to answer unequivocally the question concerning overfitting. The difficulty is associated with the limitation of applicability of the method in a case where there are a small number of reflections to be used. As judged by the R-free results for other protein structures, all the solute molecules and a very limited number of water molecules (less than 20) significantly improve the model and have to be recognized as correct (R = 20.3%, R-free = 29.6%). The rest of the water molecules would have to be considered redundant as seen in R-free calculations with entire model R = 14.0%, R-free = 33.1%.

Because the water molecules were clearly visible on the *OMIT* as well as on $2F_o - F_c$ maps we decided to model more water molecules regardless of the R-free results. We think that gradual fitting of the water molecules which have temperature factors comparable to protein and conservative refinement with *PROLSQ* allowed us to achieve a high level of model accuracy for the α_1-PT despite relatively poor model determination by the X-ray data. Nevertheless, when all 70 water molecules were fixed and R-free calculated the differences were small (R = 15.3%, R-free = 21.8%) showing that water mobility is responsible for the difference. The R-free calculations proved to have limited utility in this refinement.
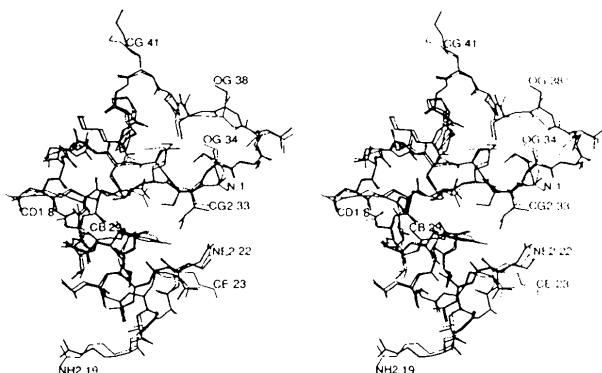

Fig. 8. Comparison of the *X-PLOR*-refined and *PROLSQ*-refined models. *X-PLOR*-refined model (light lines) superimposed on the initial model (dark lines).

Table 3. *Comparison of side-chain parameters of α_1-purothionin before and after the present refinement showing improvement*

Parameters were calculated using *PROCHECK* which is a protein analysis program. See Laskowski *et al.* (1993) and Morris *et al.* (1992) for the description of the parameters.

Stereochemical parameter	No. of data points		Parameter value		Typical value	Band width	No. of band widths from mean	
	Present	Initial	Present	Initial			Present	Initial
χ_1 g(−) st. dev	4	5	10.8	20.8	23.6	6.5	−2.0	−0.4
χ_1 trans st. dev	12	13	28.1	30.0	23.4	5.3	0.9	1.2
χ_1 g(+) st. dev	19	17	14.8	13.9	22.0	4.9	−1.5	−1.7
χ_1 pooled st. dev	35	35	19.4	22.0	22.8	4.8	−0.7	−0.2
χ_2 trans st. dev	10	10	22.5	32.6	23.6	5.0	−0.2	1.8

4. Discussion

4.1. Quality of the structure

As listed in the last two columns of Table 2, the r.m.s. deviations in bond length and angle parameters for the model are 0.012 Å and 1.2°, respectively. Fig. 9 shows a typical electron-density section. Superposition of the *R*-factor curve with the theoretical curves for different random mean position errors shows the r.m.s. error of 0.25 Å in the atomic coordinates (Luzzati, 1952) (Fig. 10) *versus* 0.64 Å reported earlier (Teeter *et al.*, 1990).

The backbone torsion angles of α_1-PT did not have major changes and the corresponding Ramachandran plot (Ramachandran, Ramakrishnan & Sasisekharan, 1963) is almost identical to that given in the previous publication (Teeter *et al.*, 1990). Arg30 has a left-handed helical conformation. It is situated in a turn at a crystal contact and has well defined electron density in the final Fourier map. Three of the residues, Arg10, Arg17 and Arg19 attain less frequently observed side-chain conformations. Arg10 and Arg17 have the same conformation as in crambin. Arg17 and Arg19 are involved in a very tight fourfold-related intermolecular contact.

The stereochemistry of the refined models (initial and final) was analyzed by the program *PROCHECK* (Laskowski, McArthur, Moss & Thornton, 1993). This program assumes that a well refined structure should have stereochemical parameters such as bond lengths, angles, torsional angles, hydrogen-bond energies, *etc.*

comparable to the mean values obtained from the analysis of the structures deposited in Protein Data Bank (Morris, MacArthur, Hutchison & Thornton, 1992) (Table 3). The number of standard deviations from the mean (in band widths), presented in last two columns, represents the goodness of the model where a positive number means that the model is worse and a negative number better than mean. The analysis shows significant improvement over the initial model (Teeter *et al.*, 1990). This is especially true for χ_1 (*trans*) and χ_2 (*trans*) conformers which lay outside the allowed standard deviation in the initial model.

4.2. Description of the structure

α_1-PT structure resembles that of crambin and its general architecture was described earlier (Teeter *et al.*, 1990). There is little change in overall structure from that published before except for tightening of the helices and making good backbone hydrogen bonds. The detailed description of the structure and details about new solute molecules would be described in the accompanying paper. In the following we describe briefly the new elements found in the structure.
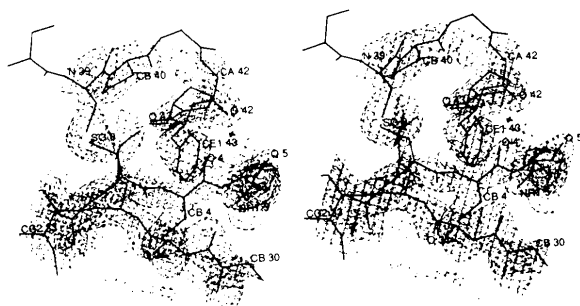

Fig. 9. Stereoscopic view of a typical electron density. Superimposed are the backbone atoms from residues 2–5, 30–33, 39–40 and 42–44. Side chains of residues 3, 4, 5, 31, 32, 39, 40, 43, and 44 are also shown.
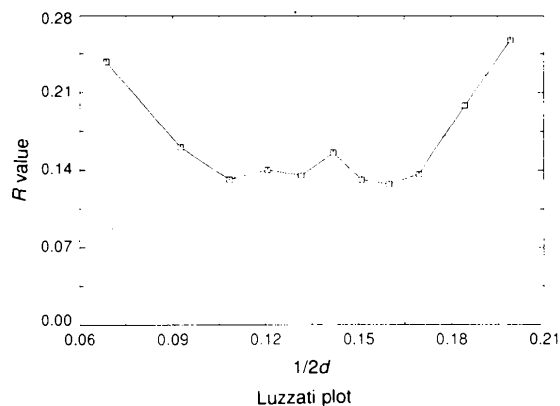

Fig. 10. Luzzati plot of the *R* factor as a function of resolution ($\sin\theta/\lambda$). The broken lines (from bottom to top: $\Delta r = 0.20, 0.25, 0.30$) show the theoretical vatiation in *R* for non-centric data (10–2.5 Å) when only coordinate errors of the model contribute to the difference between the observed and the calculated quantities (Luzzati, 1952).

4.3. *New solute molecules essential for the lattice formation*

The new features of the structure that were not seen earlier are provided by the solute molecules found in the structure. All of these solute molecules are tightly bound through several contacts, the majority of them being electrostatic as well as polar. The fact that these solutes and solvents, especially phosphate and glycerol, which are located in the lattice even though they were absent in the crystallization medium, may have biological significance. Their presence may reflect the existence of a phospholipid-binding site which is important for the toxic effect exerted on the membrane.

4.3.1. *Phosphate-binding site.* In α_1-PT, the phosphate is bound at the D interface (Teeter *et al.*, 1990). The polar dimer anchored by Asn11 and Asn14 results in a relatively high positive electrostatic potential formed at the center of the polar dimer due to the closeness of two Arg10 residues, the two Lys1 residues and two N-terminal amino groups. The phosphate ion binds to the symmetry-related pairs of NZ atoms of Lys1 (2.8 Å) and carbonyl O atoms of Pro44 (3.2 Å) (Fig. 3b).

4.3.2. *Glycerol-binding site.* In addition to phosphate, another interesting group found in α_1-PT is glycerol bound close to the phosphate ion making interactions with the residues that outline the groove between the helical stem and the β-sheet arm. O3 of the glycerol molecule is bound to the hydroxyl O atom of Tyr13 (3.3 Å), the NE Arg17 (3.0 Å), and O2 is bonded to OE1 of Gln22 (2.6 Å), whereas O1 is in van der Waals contact with Ser2 OG (3.4 Å).

4.3.3. *Acetate-binding site.* Disordered acetate is found in the four-helical bundle formed by four symmetry-related α_1-PT molecules. The presence of the acetate is reminiscent of the four-helix bundle of cytochrome *b* in which the heme group is found in the middle. Each disordered acetate forms hydrogen bonds to two symmetry-related Asn11 residues protruding toward the center (N Asn11 to O Ace, 2.7 Å). In such a location, it helps to neutralize this extremely positively charged molecule (+9) and contributes to the stability of the lattice.

We may conclude that without those negatively charged solute molecules (phosphate and acetate) placed on the symmetry elements it would be impossible to form the α_1-PT crystals.

5. Concluding remarks

The challenge of the refinement of a low-resolution structure raises problems of overfitting that need independent evaluation. The *R*-free test turned out to be a useful but limited tool in deciding which elements are redundant. We believe that careful refinement using *PROLSQ* allowed some discrimination between artifacts

of phasing and real structural elements. Further confirmation from the β-PT structure at 1.7 Å (following article) and NMR spectroscopy validate the results.

The results of this refinement show that it was possible to refine well a structure against a relatively low resolution data set and obtain a much better model. This was possible because of the external information not pertinent to the X-ray data which allowed the location of individual atoms. The low *R*-factor value (15.5%) reflects the correctness of the atomic model. This low value can be understood by analysis of the Luzzati plot. Usually, it has a dip at medium resolution (5–3 Å) for a high-resolution structure. So when a high-resolution structure is fit against lower resolution data the resulting *R* factor is low.

The remarkable density fit of α_1-PT model even at a medium resolution (~2.8 Å) and the location of a phosphate-binding site after the further refinement are clear testimonials to the benefits of analyzing in detail problems encountered during the initial stages of refinement. It cannot be underscored enough that the appropriate use of the classical tools such as a rotamer library of statistically preferred conformations (Ponder & Richards, 1987), optimized hydrogen-bond geometry and careful placing of solute and solvent molecules to match the charge, symmetry and density requirements at certain positions leads to much better structures. This is especially important in the case of small proteins like α_1-PT where even a small number of non-protein solute molecules contribute considerably to the scattering of the crystal. The use of various trial models in the refinement proved to be a helpful tool in defining non-protein parts of the model.

Despite the wide use and applicability of *X-PLOR* refinement that is seen in the literature, the present study indicates that *X-PLOR* as an automatic refinement tool for low-resolution structures needs to be used with great care. Usually it gives some improvement in the overall structure, but there can certainly be some departure of isolated side chains from correct conformations. In order to use it as a refinement program, the temperature of simulated annealing (SA) should be first optimized for the particular data set to avoid departure for manually fitted elements. It is important especially in the case of small proteins such as α_1-PT where solute molecules contribute significantly to the phasing.

The remarkable finding of acetate, glycerol and phosphate groups in the lattice, even though they were not used during crystallization, shows how tightly they have bound to α_1-PT molecules. The existence of the same molecules at nearly the same locations in the β-PT structure (see the following paper) may have important biological implications for this family. Preliminary NMR experiments carried out in our laboratory showed that small phospholipid analogs replaced inorganic phosphate at the active site of α_1-PT and β-PT. This suggests that

the phosphate and glycerol molecule may occupy the phospholipid-binding site, and they suggest a possible mechanism of toxicity for those toxins.*

---

* Atomic coordinates have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 2PLH). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: GR0392). At the request of the authors, the atomic coordinates will remain privileged until 15 March 1996.

References

APEL, K., BOHLMANN, H. & REIMANN-PHILIPP, U. (1990). *Physiol. Plant.* **80**, 315–321.
ASHIDA, T., TSUNOGAE, Y., TANAKA, I. & YAMANE, T. (1987). *Acta Cryst.* **B43**, 212–218.
BHAT, T. N. (1988). *J. Appl. Cryst.* **21**, 279–281.
BOHLMANN, H. & APEL, K. (1991). *Ann. Rev. Plant Physiol.* **42**, 227–240.
BRÜNGER, A. T. (1988). *J. Mol. Biol.* **203**, 803–816.
BRÜNGER, A. T. (1992a), *X-PLOR Version 3.1, A System for X-ray Crystallography and NMR*. New Haven and London: Yale Univ. Press.
BRÜNGER, A. T. (1992b). *Nature (London)*, **355**, 472–475.
BRÜNGER, A. T. (1992c). *Acta Cryst.* **A48**, 851–859.
BRÜNGER, A. T., KURIYAN, J. & KARPLUS, M. (1987). *Science*, **235**, 458–460.
EVANS, J. E., WANG, Y., SHAW, K. P. & VERNON, L. P. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 5849–5853.
HENDRICKSON, W. A. & KONNERT, J. H. (1980). *Incorporation of Stereochemical Information into Crystallographic Refinement*. In *Computing in Crystallography* edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN, pp. 13.01–13.23. Bangalore: Indian Academy of Sciences.
LASKOWSKI, R. A., MACARTHUR, M. W., MOSS, D. S. & THORNTON, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.
MARKMAN, O., STEC, B., RAO, U., HEFFRON, G., LEWIS, K. A. & TEETER, M. M. (1993). *Mode of phospholipid binding to the membrane active plant toxin phoratoxin-A*. In *New Developments in Lipid Protein Interactions and Receptor Function*, NATO ASI series, edited by J. A. GUSTAFSSON & K. W. A. WIRTZ, pp. 263–274. London: Plenum Press.
MORRIS, A. L., MACARTHUR, M. W., HUTCHINSON, E. G. & THORNTON, J. M. (1992). *Proteins Struct. Funct. Genet.* **12**, 345–364.
PONDER, J. W. & RICHARDS, F. M. (1987). *J. Mol. Biol.* **193**, 775–791.
PULLMAN, B. & PULLMAN, A. (1974). *Adv. Protein Chem.* **28**, 347–526.
RAMACHANDRAN, G. N., RAMAKRISHNAN, C. & SASISEKHARAN, V. (1963). *J. Mol. Biol.* **7**, 95–99.
RAO, A. G., HASSAN, M. & HEMPEL, J. (1994). *Protein Eng.* **7**, 1485–1493.
RAO, U. & TEETER, M. M. (1993). *Protein Eng.* **6**, 837–847
STEC, B., RAO, U. & TEETER, M. M. (1995). *Acta Cryst.* D51, 914–924.
STEC, B., ZHOU, R. S. & TEETER, M. M. (1995). *Acta Cryst.* D51, 663–681.
TEETER, M. M., MA, X.-QI, RAO, U. & WHITLOW, M. (1990). *Proteins Struct. Funct. Genet.* **8**, 118–132.
VOET, D. & VOET, J. G. (1990). *Biochemistry*. New York: John Wiley.
WHITLOW, M. & TEETER, M. M. (1985). *J. Biomol. Str. Dynam.* **2**, 831–848.