Acta Cryst. (1995). D**51**, 663–681

Full-Matrix Refinement of the Protein Crambin at 0.83 Å and 130 K

BY BOGUSLAW STEC

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

RONGSHENG ZHOU

Materials Data Inc., Livermore, California, CA 94551-0791, *USA*

AND MARTHA M. TEETER

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

(*Received* 15 *January* 1994; *accepted* 12 *December* 1994)

Abstract

This paper describes the first successful full-matrix least-squares (FMLS) refinement of a protein structure. The example used is crambin which is a small hydrophobic protein (4.7 kDa, 46 residues). It proves the feasibility of refining such large molecules by this classic method, routinely applied to small molecules. The final structure with 381 non-H protein atoms (54 protein atoms in alternative positions), 367 H atoms, 162 water molecules (combined occupancy 93) and one disordered ethanol molecule converged to a standard unweighted crystallographic R factor of $R = 9.0\%$ when refined against F with reflections stronger than $F > 2\sigma(F)$ and $R = 9.5\%$ when refined against F^2. The programs *RFINE* [Finger & Prince (1975). *Natl Bur. Stand. (US) Tech. Note* 854. *A System of Fortran* IV *Computer Programs for Crystal Structure Computations*] and *SHELXL*93 [Sheldrick (1993). *SHELXL*93. *Program for Crystal Structure Refinement*, Univ. of Göttingen, Germany] were used for FMLS refinement with the high-resolution low-temperature (0.83 Å, 130 K) data set of a mixed-sequence form of crambin. A detailed analysis of the models obtained in FMLS and *PROLSQ* [restrained least squares or RLS; Teeter, Roe & Heo (1993). *J. Mol. Biol.* **230**, 292–311] refinements with the same data set is presented. The differences between the models obtained by both FMLS and RLS refinements are systematic but negligible and advantages and shortcomings of both methods are discussed. The final structure has very good geometry, fully comparable to the geometry of other structures in this resolution range. Ideal values used in *PROLSQ* and those by Engh & Huber [Engh & Huber (1991). *Acta Cryst.* A**47**, 392–400] differ significantly from this refinement and we recommend a new standard. FMLS refinement constitutes a sensitive tool to detect and model disorder in highly refined protein structures. We describe the modeling of temperature factors by the TLS method [Schomaker & Trueblood (1968). *Acta Cryst.* B**24**, 63–76]. Rigid body–TLS refinements led to a better understanding of different modes of vibrations of the molecule. Refinements using F^2 or F protocols converged and reached slightly different minima. Despite theoretical support for F^2-based refinement, we recommend refinement on structure factors.

Introduction

With the advent of protein crystallography, the use of restrained-refinement methods have become routine. In the refinement of proteins a relatively small number of observations (reflections) as compared to the number of free parameters requires the use of stereochemical restraints. This leads to the question of whether the restraints introduce systematic distortions to the refined molecular models. Even though the quality of refined protein structures have improved with time (Morris, MacArthur, Hutchinson & Thornton, 1992), the question of the validity of using restrained refinement remains.

PROLSQ refinement (Hendrickson & Konnert, 1980) uses restraints for ideal geometry as parameters in the refinement. Because of this, standard deviations (σ's) on atomic positions cannot be estimated from the diffraction data, as is performed in small-molecule structures. In order to estimate the accuracy of atomic positions the Luzzati plot (1952) of R *versus* resolution is most often used. Luzzati theory assumes that the errors in structure factors are produced by random errors in atomic positions. But there has not been a good system to test the validity of this assumption or to derive an experimental value of σ from diffraction data to test restraints.

The crystal structure of the small protein crambin provides an excellent opportunity to address these questions. Crambin crystals, despite sequence heterogeneity (the isolated protein is a mixture of two sequence isomers), form a well ordered lattice stabilized by strong hydrophobic contacts and diffract to at least 0.83 Å resolution. At 130 K as compared to 300 K, solvent density in crambin crystals is much stronger indicating more

localized solvent. Restrained least-squares refinement (*PROLSQ*, Teeter *et al.*, 1993) confirmed this expectation with almost the entire solvent structure resolved.

The high-resolution crambin data at 130 K and the small size of crambin (46 residues) should allow one to conduct full-matrix least-squares refinement with very limited use of restraints. Full-matrix refinement can answer important questions about protein geometry and bound solvent unprejudiced by restraints. Full-matrix least squares can also give a reliable estimate of σ based on the diffraction data to assess accuracy of determined parameters. Further, the validity and numerical accuracy of restraints used in *PROLSQ* and other programs can be tested to see how well they agree with the unrestrained geometry parameters. We have compared the restrained least-squares model (RLS, *PROLSQ*, Teeter *et al.*, 1993) with the models derived from full-matrix least squares (FMLS, primarily *RFINE* but also *SHELXL*93) for the same 130 K, 0.83 Å crambin data.

Issues other than accuracy of restraints and evaluation of σ were addressed by this study and can help to evaluate the potential of each method for refinement of proteins. In particular, the ability to model disorder accurately, to refine B values appropriately, to apply the TLS methods to vibrational factors (Schomaker & Trueblood, 1968), and the effect of F *versus* F^2 refinement protocols on the outcome, were explored.

From our previous experience with small-molecule and protein refinements, disorder is a major limiting factor in achieving better R factors and, in effect, better structures. Therefore, in both refinements, disorder has been explicitly modeled both for protein atoms and solvent molecules to the experimental limits of the method (*i.e.* stable refinement and negligible difference density). Sources of disorder in crambin crystals have been described elsewhere (Teeter *et al.*, 1993; Yamano & Teeter 1994).

The problem of disorder is closely related to the refinement of temperature factors. Vibrational ellipsoids reflect the natural thermal vibrations as well as any unmodeled disorder. Accurate modeling of vibrational anisotropy, using six parameters rather than the three which are used in *PROLSQ* refinement, should help to identify such unmodeled disorder. The rigid-body refinement or TLS method (Schomaker & Trueblood, 1968) was designed to model thermal vibrations for rigid systems while reducing the number of free parameters to refine. The effectiveness of this approach was tested by comparing the distribution of temperature factors from the full-matrix refinement with that from refinement with the TLS method.

A procedure is proposed to separate the external modes of vibrations from internal ones using TLS. By introducing different numbers of rigid bodies for the molecule and considering the common parts of this motion to be external, separation of parts of the motion can be achieved.

Several different refinement protocols were tested to judge their effect on the FMLS refinement. Harris & Moss (1992) recently concluded that there is no particular advantage in running the refinement on F^2 rather than on F, and both strategies have been tested here. Also, the influence of the number of reflections from different intensity and resolution cutoff levels on the stability and convergence of the refinement was analyzed.

In this paper, we begin by describing briefly the FMLS refinement methods used in this study. The results are subdivided into sections that describe major results and problems encountered in the FMLS refinement. The paper is concluded with a summary in which we present our major findings and recommendations.

Materials and methods

Crambin, a small (4.7 kDa) hydrophobic plant protein, is isolated by acetone extraction from defatted seeds of *Crambe abyssinica* (Van Etten, Nielsen & Peters, 1965). It has been crystallized by the vapor-diffusion method from 30 mg ml^{-1} solution of protein in 80% ethanol–water solution equilibrated against 60% ethanol–water (Teeter & Hendrickson, 1979). Crystals of approximate size 0.5 × 0.7 × 0.4 mm were grown in one month.

In order to eliminate errors due to absorption by the capillary and mother liquor, data have been collected by the method of Hope (*i.e.* flash cooling; Teeter & Hope, 1986) at 130 K on a four-circle diffractometer. A crystal was mounted on the tip of a glass fiber and frozen by sudden exposure to an N_2 stream at 130 K.

The protein crystallized in the monoclinic $P2_1$ space group with cell dimensions $a = 40.763\,(5)$, $b = 18.492\,(3)$, $c = 22.333\,(3)$ Å, $\beta = 90.61\,(1)°$ as established on measurements in the 44–62° range in 2θ (Hope, 1988). The Cu radiation data on a crystal 0.5 × 0.7 × 0.4 mm was collected with an appropriately sized collimator by an ω-scan to the resolution of 0.83 Å. More than 32 000 reflections were collected. Since no reduction in intensities was observed with time, no correction for radiation damage was required. Some high-resolution data which suffered from a systematic error, caused by backscattering from a low-temperature device, were eliminated from the data set. After the Lorentz and polarization correction, as described elsewhere (Teeter *et al.*, 1993), the final data set of 29 441 reflections has been used for the refinement. The resolution distribution of $F/\sigma(F)$ values is presented in Fig. 1. In the highest resolution shell, the signal-to-noise ratio $[F/\sigma(F)]$ is 6.

For full-matrix refinement, we used the program *RFINE*88, developed by Finger & Prince (1975). The code was optimized for a vector and parallel-processing environment by one of us (RZ). Modifications were necessary because such a long time was required to

run one cycle of least-squares refinement that successful refinement would have been prohibited. After changes which sped up the program by a factor of at least 40, one cycle of the refinement on the Stellar GS1000 (Stardent) minisupercomputer with 64 Mb memory and four parallel and one vector processor took around 8 h c.p.u. time.

The starting model of about 800 atoms (450 non-H atoms) was derived from an RLS model (*PROLSQ*) refined with all reflections above 2σ in the 10–0.83 Å resolution shell ($R = 15.8\%$). All the disordered atoms were removed except for a few clearly resolved water molecules with partial occupancies. Initial anisotropic U values were taken from the *PROLSQ* refinement.

Since the total number of atoms used in the refinement was above 800, refining all atoms anisotropically (ten parameters per atom) would result in an unfavorable ratio of free parameters to reflections and ill conditioned refinement. Thus, we limited the number of refined parameters to at least six observations for each parameter refined. To achieve this, we constrained H-atom positions to ride on the host atom and strongly restrained their B's to the value of the host atoms. We also used rigid-body constraints for disordered alternates of Phe13, Tyr29 and Asn43 to avoid deformations due to the close proximity of the disordered alternates. The refined unrestrained Tyr44 had a final aromatic ring geometry very similar, within experimental error, to that of the restrained Tyr29 and Phe13, giving credence to such a procedure. Isotropic B's were used for all H atoms and most of disordered atoms including low-occupancy water molecules. Small-molecule refinement frequently uses similar restraints on H atoms and when necessary on other molecular fragments.

In the first stage of the refinement, atoms were refined anisotropically if the diagonal elements of the β tensor were five times greater than their e.s.d. from the covariance matrix using a test least-squares cycle. An atom was converted to an isotropic if one of the diagonal elements of its β tensor became less than three times the e.s.d. or became unreasonably anisotropic (very large semiaxes ratio). During the course of the refinement isotropic atoms were repeatedly tested as anisotropic. If they failed one of the previously defined criteria, they were converted back to isotropic.

An atom was eliminated from the refinement if all of its β tensor diagonal elements refined below their e.s.d.'s or if its isotropic B was refined below its e.s.d. or was greater than 50. Initially, 14 water molecules were removed from the *PROLSQ* model using these criteria.

As the refinement progressed, the conversion criteria were relaxed, and we tried to maximize the number of anisotropic atoms. In the end, the refinement was well conditioned and for most of atoms the average $\beta_{ii}/\sigma(\beta_{ii})$ was 8.5. The final $2F_o - F_c$ electron-density map was very well defined even for low-occupancy sites. Well ordered sites could be distinguished from more dynamically disordered sites where the density was smeared out.

In first three cycles, we refined the protein and well ordered water molecules alternatively with the data in the 5.0–0.83 Å resolution range. Subsequently, we introduced the disordered water molecules and refined against the data from 10.0 to 1.2 Å. Next, we gradually increased the resolution and modeled the more disordered solvent. At 1 Å resolution more disorder in the side chains became apparent in difference electron-density maps. Disordered residues were refined constrained to have the occupancies for the alternates sum to 1 and the same occupancy for disordered atoms in each alternate. Occupancies were adjusted by hand to keep the temperature factors of alternates in the same range (for an example of the method see, Smith, Corfield & Hendrickson, 1988). Attempts at independent occupancy refinement failed presumably because most of the alternates are positioned close to their mates.

After around 200 cycles at 0.83 Å resolution, the R factor converged. The final model had 917 separate atom positions, of which 506 were for non-H atom positions. 27 waters had full occupancy, and 135 waters along with 54 protein atoms and one ethanol molecule had partial occupancies. 354 out of 506 non-H atoms were refined with fully anisotropic, six-parameter thermal ellipsoids.

The refinement was run either on F or F^2. Different resolution limits (27.0–0.83 and 10.0–0.83 Å) and intensity cutoffs [$F_o/\sigma(F_o) > 1$ and 2] were tried. The program's weighting scheme was,

$$w_F = 1/\sigma(F),\ w_F^2 = 1/\sigma(F^2) = 1/[2F\sigma(F)].$$

It may be partially responsible for different results obtained in F^2 and F refinements. The weights for outlier reflections were calculated from the formula,

$$w_F = \max\left[\left||F_o| - |F_c|\right|/A \text{ or } \left||F_o| - |F_c|\right|/\sigma(F)\right],$$

where A is an arbitrary parameter and F becomes F^2 in F^2 refinement. We changed A to be comparable to the average $\sigma(F)$ or to the average $\sigma(F^2)$ in refinement

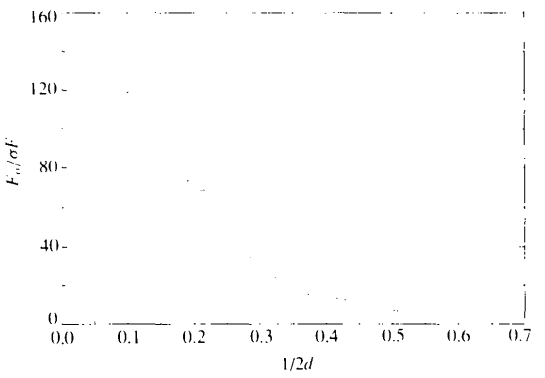

Fig. 1. Distribution of $F/\sigma(F)$ *versus* resolution ($1/2d$ or $\sin\theta/\lambda$).

on F or F^2, respectively, (2.0–10.0, for F refinement and 10.0–100.0 for F^2) but the R-factor change was small (less than 0.01) and overall refinement looked very similar regardless of the A value.

The above weighting scheme relies heavily on estimated σ's, which for protein diffractometer data can be less reliable at high resolution. An alternative weighting scheme for full-matrix refinement, that deals with the inaccuracy of high-resolution σ's, has been implemented in *SHELXL*93 (Sheldrick, 1993). It is close to that proposed by Cruickshank (Pilling, Cruickshank, Bujosa, Lovell & Truter, 1961). The weighting scheme used in *PROLSQ* is $w = A + B(\sin\theta - 1/6)$, where A and B are estimated from the least-squares cycle to be about $F_o - F_c$.

After completion of the *RFINE*–FMLS refinement, *SHELXL*93 became available. The FMLS was run with this program in order to assess its efficacy and for comparison with the *PROLSQ* and *RFINE* refinement models. We began *SHELXL*93 refinement with the *RFINE*–FMLS model of crambin (x, y, z, B_{eq}) and ran 20 cycles of conjugated gradient (CGLS) with restraints and isotropic B's. Subsequently, the anisotropic B's were introduced and another ten cycles of CGLS with restraints were carried out. In the end, the restraints on heavy atoms were removed (H atoms were restrained to ride on heavy atoms in a manner similar to FMLS) and 15 cycles of full-matrix least squares were run. Even though the refinement was run on F^2 the R factor converged to 8.98% for the data $F > 2\sigma$. The R factor as well the geometry of the model is comparable to the *RFINE*–FMLS. Advantages of this program are the use of restraints to an average geometry for disordered atoms and a weighting scheme that may be more appropriate for proteins.

Results of the refinement. Differences between RLS (*PROLSQ*) and FMLS (*RFINE*)

In this section we describe the process of FMLS refinement focusing on major advantages and disadvantages of the method. Subsequently, we describe the final model of crambin and its comparison with the RLS-refined model. We also describe the statistical differences in distributions of bond distances and angles and assess the suitability of each refinement method for shaping different features of the models.

Small but significant differences between the models were observed in regions where disorder has either not been modeled at all or where the three parameter 'thermal ellipsoids' modeling used in *PROLSQ* was inadequate. We use quotes because, as indicated before (Dunitz, Schomaker & Trueblood, 1988), these ellipsoids describe the Gaussian distributions of atomic positions due to different factors including static and dynamic disorder averaged in time and space. A comprehensive treatment of disorder such as that carried out here highlights the problematic traditional interpretations of high vibrational parameters. In particular, we describe the following.

(1) Modeling of disorder.

(2) Average model stereochemistry *versus* commonly used ideal values.

(3) Differences between positions and temperature factors obtained in FMLS and RLS refinements.

(4) Analysis of individual temperature factors: shape and magnitude of motion.

(5) Influence of disorder on the modeling of thermal parameters.

(6) Differences in various stages of refinement: modeling disorder with two isotropic atoms *versus* one anisotropic atom (2×5 *versus* 10 atomic parameters models).

(7) Modeling temperature factors by TLS.

(8) Differences between F^2 and F refinement.

Disordered residues

The final FMLS–*RFINE* model (Fig. 2) contains 14 residues in alternative positions *versus* 13 in RLS (Asn12 not modeled in *PROLSQ*). These constitute 30% of the protein residues and include Thr1, Thr2, Ile7, Val8,

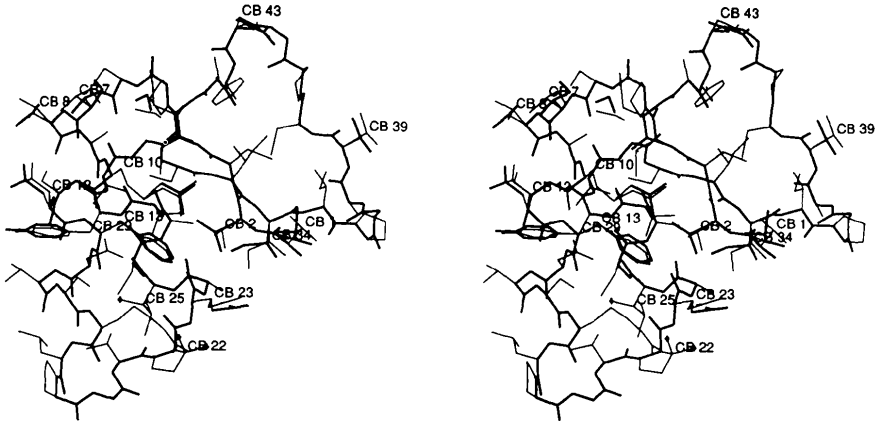

Fig. 2. Bond-and-stick model of crambin (heavy atoms only). Backbone bonds and disordered side chains shown as heavy lines. Disordered residues are labelled.

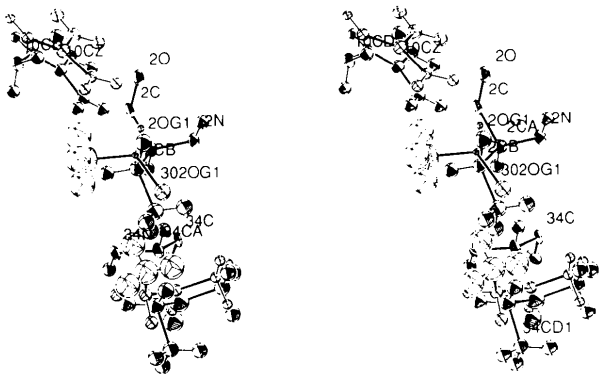

Fig. 3. FMLS model of the correlated disorder in Arg10, Thr2 and Ile34. Major positions represented by filled ellipsoids (20%) while the disordered alternatives have light ellipsoids (20%). The very small temperature factor of C_γ in residue Thr302 is an artifact of modeling a low-occupancy site in between the atoms of the major occupancy sites with distances smaller than the resolution limit of the data.

Arg10, Asn12, Phe13, Pro/Ser22, Glu23, Leu/Ile25, Tyr29, Ile34, Thr39, Asn43 plus a disordered ethanol molecule Eth66 (Table 1). All the disordered atoms in FMLS–*RFINE* had 300 added to the original sequence number. Because not all disordered side-chain atoms were modeled in RLS, the total number of disordered atoms in the FMLS model is larger. Specifically, residue Asn12 is disordered, and for Ile7 and Ile34, all four side-chain C atoms were modeled as disordered (only CD* atoms were modeled in RLS). However, atoms such as disordered CB in Phe13 and Tyr29 could not be modeled in FMLS because they were too close to the alternate. The same was true for Arg10 where in RLS two additional C atoms (CB and CG) were modeled in alternative positions. In both refinements, heterogeneous

* Atom designations used are from the Protein Data Bank. Greek letters are replaced by alphanumeric equivalents.

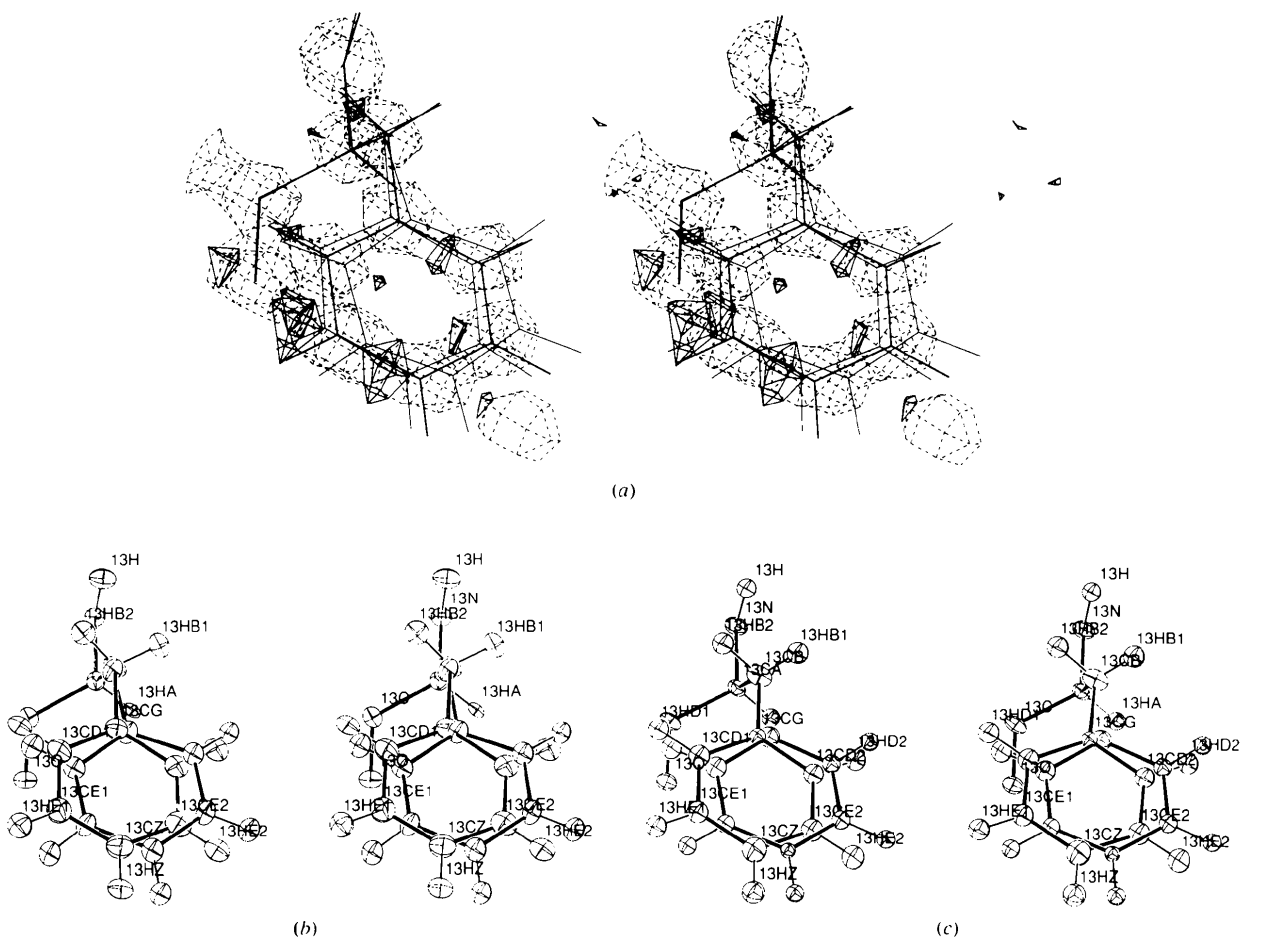

Fig. 4. Electron-density maps covering Phe13 before the alternative positions have been modeled. (*a*) A heavy line depicts the original model, thin lines two alternative positions of the side chain. $2F_o - F_c$ 'electron-density' map is shown as broken lines (1.6σ level) with the persistent difference map peaks as continuous lines (3.5σ level). (*b*) *PROLSQ* model of Phe13 in *ORTEP* representation, (*c*) model derived from FMLS, both models on 20% ellipsoid level.

Table 1. *Side-chain torsions of disordered alternates for the FMLS-RFINE model*

For Arg10 only the angles χ_2, χ_3, χ_4 as indicated by square brackets [χ_2, χ_3, χ_4] are given.

	χ_1, [χ_2] (°) orig, alter.	χ_2, [χ_3] (°) orig, alter.	χ_3, [χ_4] (°) orig, alter.
Thr1	−63, −164		
Thr2	57, −67		
Ile7	−65, −82	176, −61	
Val8	−77, −165		
Arg10	64, 64	62, 74	−176, 177
Asn12	−76, −62	−17, −26	
Phe13	176, −174	−84, −84	
Pro/Ser22	−35, −176, 78	41	
Glu23	−73, −59	−173, 177	−17, −36
Leu/Ile25	−77, −77, −77	−70, −70, 167	
Tyr29	161, −171	68, 50	
Ile34	−59, −38	172, −70	
Thr39	−63, 45		
Asp43	53, 47	−15, −9	

residues (22 and 25) were modeled similarly with the OG of Ser22 and CD1 of Ile25 modeled in two positions while the CG1 in Ile25 was modeled in one position. Examples of the observed discrete disorder are given in Figs. 3–7.

Some disorder appears to be clearly correlated among spatially adjacent residues (Yamano & Teeter 1994). Fig. 3 shows an example of such a disordered path found for the residues 10-2-34. Other disordered pathways involve residues 7-8-12-12′ (symmetry-related), 1-23-44′ and 39-Eth′-7′, where Eth′ is a symmetry-related ethanol. For Thr39, the backbone is apparently disordered (but shifts are too small to model), and the modeled disorder involves rotation of the side chain around the C—C_α bond. Together, these create the appearance of translational disorder. For both alternates, the O_γ forms a hydrogen bond to a disordered ethanol molecule. The commonly seen inclusion of C_β with side-chain disorder (residues 1, 2, 7, 8, 12, 34 and 39) suggests backbone atoms are also disordered and do contribute to structural substates. However, no backbone disorder has been modeled.

The methodology we adopted in modeling disordered atoms was as follows. In cases where the persistent difference electron-density map (> 2σ) was present in the vicinity of an anisotropically refined atom for several cy-

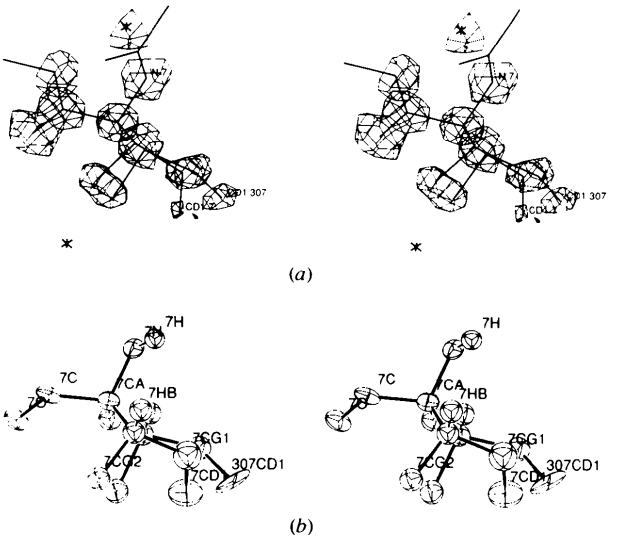

Fig. 6. (*a*) $2F_o - F_c$ electron density (1.6σ level) for Ile7. (*b*) *ORTEP* drawing of the final FMLS model (ellipsoids on 20% level).

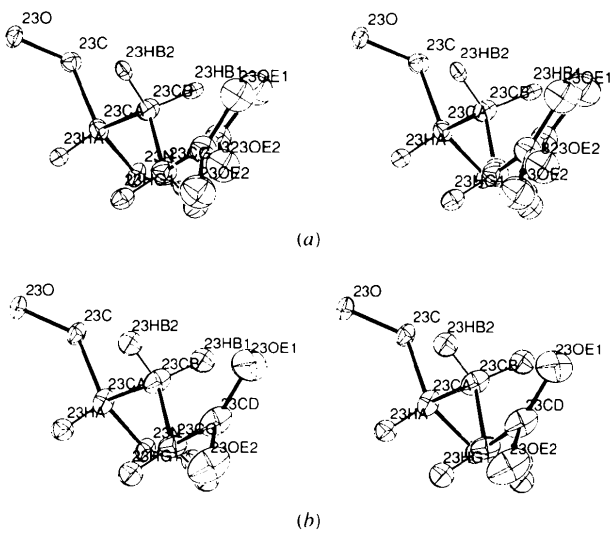

Fig. 5. *ORTEP* representation of Glu23, ellipsoids are at the 20% level. (*a*) Model derived from *PROLSQ*, (*b*) FMLS model before modeling disorder.

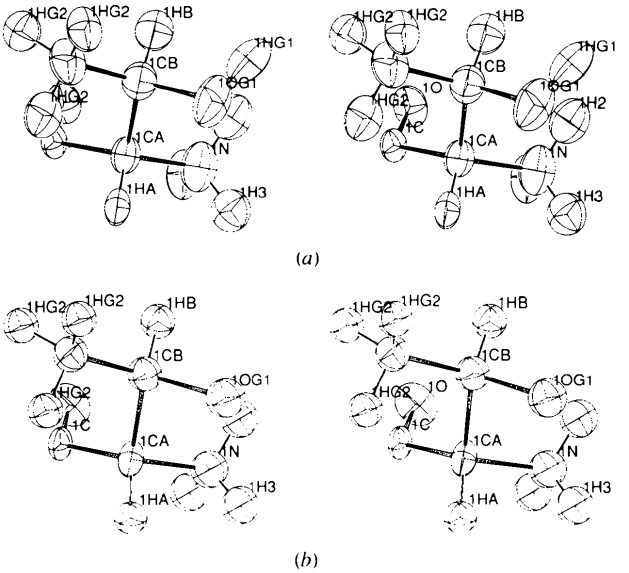

Fig. 7. *ORTEP* drawing of Thr1 model for major occupancy sites with ellipsoids at the 20% probability level. (*a*) Model from *PROLSQ* refinement. (*b*) Model from FMLS refinement

Table 2. *The differences in positions, average temperature factors and occupancies between alternative positions for disordered atoms in both PROLSQ (upper rows) and in RFINE (lower rows) models*

The last row for each model is the number of atoms modeled in two positions. Alternates are denoted by subscripts 1 and 2 (*e.g.* Q_1, Q_2, B_1, B_2)

	T1	T2	I7	V8	R10	N12	F13	P/S 22	E23	L/I 25	Y29	I34	T39	D43
PROLSQ Δr^* (Å)	1.55	1.82	1.87	1.58	0.56	NM†	0.47	0.19	0.59	NA‡	1.56	2.73	0.82	0.79
Q_1	0.8	0.8	0.7	0.6	0.8		0.6	0.6	0.8	0.6	0.6	0.8	0.7	0.7
Q_2	0.2	0.2	0.3	0.4	0.2		0.4	0.4	0.2	0.4	0.4	0.2	0.3	0.3
B_1	6.15	6.72	10.78	6.49	3.98		4.01	4.10	5.83		5.27	8.54	6.76	4.38
B_2 (Å^2)	12.55	5.13	7.47	4.47	4.51		3.20	3.91	6.12		7.34	5.71	8.19	6.30
No. of atoms	5	3	1	3	6	0	7	3	4	1	8	1	3	4
FMLS Δr (Å)	1.62	1.93	1.10	1.44	0.71	0.63	0.53	NA‡	0.67	NA‡	1.71	1.73	0.90	0.68
Q_1	0.8	0.74	0.55	0.55	0.75	0.5	0.55	0.55	0.8	0.55	0.55	0.75	0.65	0.8
Q_2	0.2	0.25	0.45	0.45	0.25	0.5	0.45	0.45	0.2	0.45	0.45	0.25	0.35	0.2
B_1	6.15	5.56	6.36	6.00	3.75	2.82	3.71	4.10	4.90	4.62	5.26	4.36	5.57	4.98
B_2 (Å^2)	6.94	5.31	5.55	6.16	3.30	3.08	2.87	5.15	3.65	4.54	10.4	7.54	6.01	5.64
No. of atoms	5	3	4	3	4	4	6	2	4	1	7	4	3	4

* $\Delta r = [\sum(x_1 - x_2)^2]^{\frac{1}{2}}/n$.
† NM, not modeled.
‡ NA, not applicable – different atom types.

cles indicating two centers rather than large anisotropic motion, we tried to model two atoms in alternative positions (Fig. 4). Usually the thermal ellipsoid of the atom in question was not excessively elongated (the ratio of semiaxes around 2–3, see Fig. 5). When disorder was modeled, the two atoms in different positions were refined isotropically. In most such cases, the disordered atoms refined well.

The rules used for selecting disordered atoms were stricter than rules used in small-molecule refinement, where the ellipsoids for individual atoms are frequently very elongated. In small-molecule structures, we believe that unmodeled disorder for certain atoms can be detected when high B_{eq} corresponds to very elongated thermal ellipsoid (high semiaxes ratio).

For closely spaced disordered alternates, dynamic disorder cannot be ruled out *a priori*. However, the lack of difference density between alternates and overall improvement in the quality of the refinement (lower R factor and improvement of clarity of the $2F_o - F_c$ maps) leads one to believe that discrete disorder is predominant. The extension of the side-chain disorder to correlated alternate water networks confirms this discrete disorder. For example, even though the alternates for Asn12 or Ile7 are closer than the resolution limit and separate peaks are difficult to distinguish, waters adopt distinguishable networks of disorder and support two distinct alternative models for both side chains. Successful refinement of two positions for most disordered residues indicates that molecule is present in two different conformational substates. Temperature factors and occupancies for the disordered residues are presented in Table 2.

One can argue that dynamical disorder can create an impression of static disorder with two distinct positions. However, the electron density and temperature factors in combination with a total occupancy that sums to 1 indicate that the majority of residues are discretely disordered. This observation supports the notion that below 200 K (Frauenfelder, Parak & Young, 1988) a protein is frozen into conformational substates with significantly diminished mobility, regardless of the origin of the observed disorder (packing, sequence heterogeneity *etc.*).

Average stereochemistry of the models, PROLSQ idealities

Based on the FMLS low positional σ'*s* unbiased by restraints (all atom $\sigma_{pos} = 0.022$ Å) and the low R factor (9%), the crambin FMLS model represents a very well refined structure, comparable to small-molecule structures. From the average backbone $\sigma_{pos} = 0.0089$ Å, we have calculated the backbone $\sigma_{bond} = 0.0126$ Å and backbone $\sigma_{ang} = 0.89°$. These values were used to estimate the accuracy of the refinement and to assess the stereochemistry of our model. We have calculated the backbone bonds and angles and checked their distributions with statistical t and χ^2 tests.

By using a second program for FMLS refinement [the newly developed program *SHELXL*93 (Sheldrick, 1993)], we can hopefully identify program-dependent influences on backbone geometry. Values of σ_{pos} for this model are very similar to the FMLS–*RFINE* model.

The comparison of the refinement parameters for *PROLSQ*, *SHELXL*93 and *RFINE* is presented in Table 3. The stereochemistry for the *PROLSQ*, *SHELXL*93 and *RFINE* models was compared statistically (Table 4) to the *PROLSQ* ideal values and to the newly developed protein geometric parameters (Engh & Huber, 1991).

The results presented in Table 4 indicate some statistically significant differences between the refined models and ideal backbone bond-length values from *PROLSQ* and from Engh & Huber (Engh & Huber, 1991). While many backbone bond lengths are close to the Engh & Huber values, the value for the C—O bond, which has a normal distribution according to the χ^2 test, was

Table 3. *The comparison of crucial statistics for different refinement modes*

GOF $= [\sum w(|F_o| - |F_c|)^2/(n-m)]^{1/2}$, $R = \sum(||F_o| - |F_c||)/\sum_2 |F_o|$, $R_w = \sum w(|F_o|^2 - |F_c|^2)^2/\sum w|F_o|^4$ for refinement on F^2, and $R_w = \sum w(|F_o| - |F_c|)^2/\sum w|F_o|^2$ for refinement on F.

	No. of reflections	Mode of refinement, low-resolution cutoff (Å)	Cutoff on F_o/σ	R (%)	R_w (%)	*RFINE* weight *A* *PROLSQ* *A* and *B*	Average shift/ e.s.d.	GOF (s.d. of the observation)
PROLSQ	23786	F, 10	2	10.3	11.1	8.0, −5.0	0.0005	
PROLSQ	28686	F, 27	1	11.3	10.9	5.4, −3.4	0.0005	
FMLS 1	29338	F, 10	1	11.4	7.4	8.0	0.05	2.12
FMLS 2	29364	F, 27	1	11.5	7.4	8.0	0.05	2.12
FMLS 3	23759	F, 10	2	9.0	6.2	3.0	0.06	2.03
FMLS 4	23773	F, 27	2	9.1	6.7	3.0	0.15	2.12
FMLS 5	23759	F^2, 10	2	9.5	14.7	90	0.28	2.42
FMLS 6	23773	F^2, 27	2	9.6	14.8	90	0.40	2.62
TLS1	29358	F, 27	1	13.8	8.4	3.0	0.32	2.27
TLS2	29358	F, 27	1	13.5	8.2	3.0	0.10	2.23
TLS3	29358	F, 27	1	13.3	8.1	3.0	0.12	2.20
TLS21	23773	F, 27	2	12.2	8.2	3.0	0.04	2.49
TLS22	23773	F, 27	2	12.0	8.1	3.0	0.10	2.45
TLS23	23773	F, 27	2	11.7	7.9	3.0	0.04	2.40
*SHELX*1	28726	F^2, 10	1	10.4	23.7	0.16, 0.0	0.10	0.968
*SHELX*2	23209	F^2, 10	2	8.96	22.51	0.16, 0.0	0.27	1.031

Table 4. *Stereochemical comparison of the FMLS, PROLSQ and SHELXL93 (SHELX93) models with the PROLSQ idealities and Engh & Huber set of atomic parameters*

The sample size is given in square brackets. The standard deviations are in parentheses.

Average distances (Å)	N—CA* [46]	CA—C* [46]	C—O* [46]	N—C* [45]
Engh & Huber	1.458 (19)	1.525 (21)	1.231 (20)	1.329 (14)
PROLSQ ideal	1.470	1.530	1.240	1.320
PROLSQ	1.459 (13) $^{R}_{A}$ **A**	1.522 (18) $^{R}_{A}$ **R**	1.240 (14) $^{A}_{R}$**A**	1.337 (18) $^{R}_{R}$ **R**
PROLSQ 2σrej†	1.460 (12)	1.521 (13)	1.239 (13)	1.335 (14)
*SHELX*93	1.459 (13) $^{R}_{A}$ **A**	1.523 (17) $^{R}_{A}$ **R1**	1.238 (15) $^{A}_{R}$ **A**	1.337 (18) $^{R}_{R}$ **R**
*SHELX*93 2σrej†	1.459 (10)	1.524 (14)	1.238 (13)	1.337 (12)
FMLS	1.463 (24) $^{R1}_{A}$ **R**	1.524 (29) $^{A}_{A}$ **R**	1.239 (19) $^{A}_{R1}$ **R**	1.345 (24) $^{R}_{R}$ **R**
FMLS 2σrej†	1.460 (22)	1.520 (25)	1.235 (16)	1.344 (23)
FMLS nd‡	1.460 (23) $^{R1}_{A}$ **R**	1.521 (26) $^{R2}_{A}$ **R**	1.240 (20) $^{A}_{R1}$ **R**	1.343 (27) $^{R}_{A}$ **R**

Average angles (°)	N—CA—C* [46]	CA—C—O* [46]	O—C—N* [45]	C—N—CA* [45]
Engh & Huber	111.2 (2.8)	120.8 (1.7)	123.0 (1.6)	121.7 (1.8)
PROLSQ ideal	109.66 (2.452 Å‡)	120.96 (2.414 Å‡)	124.96 (2.271 Å‡)	123.03 (2.453 Å‡)
PROLSQ	111.81 (2.61) $^{R}_{A}$ **R**	120.23 (1.63) $^{R}_{R1}$ **R**	123.02 (1.49) $^{R}_{A}$ **R**	120.92 (1.98) $^{R}_{R1}$ **R**
PROLSQ 2σrej†	111.70 (2.05)	120.42 (1.08)	122.89 (1.29)	120.94 (1.51)
*SHELX*93	111.32 (2.43) $^{R}_{A}$ **R**	120.01 (1.76) $^{R}_{R}$ **R**	122.71 (1.37) $^{R}_{A}$ **R**	120.90 (1.78) $^{R}_{R}$ **R**
*SHELX*93 2σrej†	111.29 (1.53)	120.03 (1.36)	122.71 (1.37)	120.72 (1.58)
FMLS	111.47 (2.44) $^{R}_{A}$ **R**	120.33 (2.06) $^{R1}_{A}$ **R**	122.85 (2.11) $^{R}_{A}$ **R**	120.51 (2.40) $^{R}_{R}$ **R**
FMLS2σrej†	111.53 (2.17)	120.19 (1.79)	122.65 (1.92)	120.12 (1.92)
FMLS nd§	11.89 (2.40) $^{R}_{A}$ **R**	120.23 (2.17) $^{R2}_{A}$ **R**	122.95 (2.11) $^{R}_{A}$ **R**	121.24 (2.60) $^{R}_{A}$ **R**

* Statistical *t*-test code is at the right-hand side of an entry. Upper symbol represents sampling against *PROLSQ* idealities and the lower symbol against Engh & Huber parameters. The last (bold) symbol denotes the χ^2 test for obtained distributions of bonds or angles. A rejection (**R**) means that it is unlikely that the distribution is normal and has only one mean. Criteria for acceptance and rejection of the null hypothesis are as follows: **A** accepted; **R**2, reject at 10% and accept at 5% level; **R**1, reject at 5% and accept at 1%; **R**, rejected at 1%.

† *PROLSQ*2σrej, *SHELX*932σrej and FMLS2σrej denote the averages with outliers in the sample (bond or angles > 2σ) omitted.

‡ Hendrickson's 1–3 distances, corresponding to planar restraints.

§ FMLS nd denotes the model with disordered residues omitted: 1, 2, 7, 8, 12, 13, 22, 25, 29, 34, 39, 43.

closer to the *PROLSQ* ideal value than that derived by Engh & Huber. The differences are more significant for FMLS–*SHELX* refinement than for *RFINE* refinement because the σ's are generally larger in the FMLS–*RFINE* refined models.

Results for backbone angles show non-normal distributions and make the results difficult to validate statistically. However, this may result from the very low σ_{ang} (0.89°) in the FMLS–*RFINE* and –*SHELX*93 refined model. Although we cannot recommend ideal values for backbone angles based on parametric statistics at this time, the average angle values for the models are remarkably similar despite the different methods used. The angle trends reported here are independent of the sample size (outliers omitted) or the presence/absence of disordered residues in the sample (data not shown).

The fact that Table 4 shows that the means for both backbone bonds and angles are remarkably refinement independent gives us hope that we will be able to define the distributions for the angles more precisely with further analysis. Another interesting observation is a secondary-structure dependence of the average backbone values as reported earlier (Teeter, 1985). This property is also difficult to verify statistically for the small sample sizes belonging to the particular secondary-structure category. More detailed analysis is underway.

Nonetheless, some conclusions can be drawn about backbone bond distances and angles at this time. We suggest more accurate values for backbone distances and angles can be adopted as target for protein refinement. These are the values in the rows obtained with *SHELXL*93 in Table 4. Those values are very representative for other refinements and have relatively low σ's.

The distribution of ω-angles (Fig. 8) as well as torsional angles φ and ψ (Fig. 9) agree with those reported recently in the literature (Ashida, Tsunogae, Tanaka & Yamane, 1987; Morris *et al.*, 1992). However, the crambin distribution for ω is wider than usually assumed in protein refinement restraints ($\sigma = 3°$) with standard deviation around 180° about 4.9°. It is, however, smaller than reported by the Thornton's group (6°) (Morris *et al.*, 1992). The torsional side-chain angle ($\chi_{1,2}$) distribution is much narrower than often seen (Bhat, Sasiseskharan & Vijayan, 1979) and centered around three staggered states 60, –60, 180° (Fig. 9). For each residue type, it also falls close to the rotamer values for that residue type (Ponder & Richards, 1987).

The backbone torsional angles for helix and sheet fall within expected ranges. Helices (residues 8–18 and 23–28) are short but regular with average values for $\varphi = 63.1\,(6.5)°$ and $\psi = 37.9\,(8.5)°$. These values indicate more 'open' hydrophilic helices (Blundell, Barlow, Borkakoti & Thornton, 1983). Backbone hydration is described in Teeter *et al.* (1993). The β-sheet region (residues 1–4 and 32–35) has the averaged torsional angles for the backbone $\varphi = -122.85\,(12.5)°$ and $\psi = 148.7\,(11.5)°$. These values agree well with *PROLSQ* refinement ideal values, although backbone torsions were not restrained. Standard deviations are given in parentheses.

Proline residues in crambin form two distinct classes distinguished by more or less puckered rings. The puckering seem to be correlated with backbone conformation, *i.e.* with the distances between two neighboring C_α atoms. Neighboring $C_{\alpha n-1}$ and $C_{\alpha n+1}$ in tight turns and α-helices are closer together than those in extended conformation. If the proline is found in a reverse turn (Pro19 or Pro36), it has a less puckered ring with average values of $\chi_1 = 18°$, $\chi_2 = -25°$, $\chi_3 = 21°$. When it is located in a stretch of residues in a more extended conformation (Pro5, 22, 41) then the values become larger: $\chi_1 = |34°|$, $\chi_2 = |42°|$, $\chi_3 = |33°|$ with both puckering states represented (+, –, + and –, +, –). Electron-density maps as well as thermal ellipsoids give no indication of disorder for either class. In the more puckered prolines, C_γ has usually a temperature factor similar to other atoms (*i.e.* $\langle B_{CG} - B_{CA} \rangle = 0.7$ Å^2) indicating low mobility and a very stable position. For less puckered proline residues $C\gamma$ has slightly elevated temperature factor (*i.e.* $\langle B_{CG} - B_{CA} \rangle = 3.2$ Å^2) showing less stable conformation. This magnitude of difference is observed despite a range of average B values for its neighboring residues. The key values for torsion angles and B factors for the prolines are reported in Table 5. The puckering state of proline ring appears, then, to be governed by the dihedral angle C_{n-1}—N—C_α—C_β or, equivalently, by the direction of the CO group. For each puckering, the torsion angle χ_1 has the same sign as the C_{n-1}—N—C_α—C_β. It is possible that all the prolines in high-resolution proteins may have similar properties.

Another interesting feature of this highly refined model is that we can see many more water molecules in pentagonal ring structures, than reported previously (Teeter, 1984). Even partially occupied sites forming alternating networks have the tendency to form pentagonal and hexagonal rings. We have identified several new rings extending already identified polyhedron water

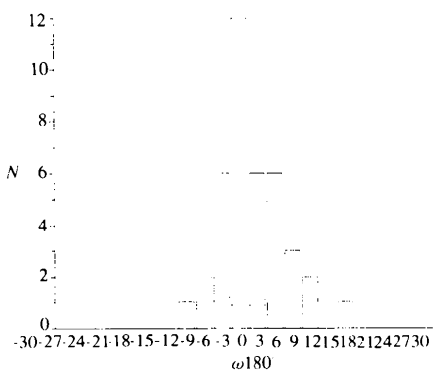

Fig. 8. Distribution of the backbone ω torsional angles.

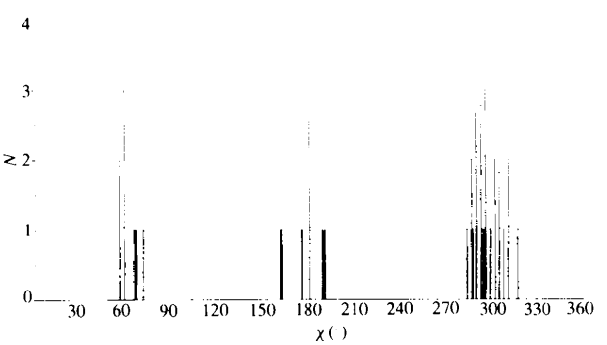

Fig. 9. Distribution of the side-chain χ torsional angles. Open bars, χ_1; full bars, χ_2.

Table 5. *Selected conformational data for proline residues*

Note differences in torsion angles between class I (Pro19 and 36) and class II (Pro5, 22 and 41) residues as described in the text. Two last columns represent B equivalent with the e.s.d. calculated from the FMLS in parentheses.

	Residue	$d(C_{\alpha_{n-1}}, C_{\alpha_{n-1}})$ (Å)	$C_{n-1}-N-C_\alpha-C_\beta$ (°)	χ_1 (°)	χ_2 (°)	χ_3 (°)	B_{C_α} (Å^2)	B_{C_γ} (Å^2)
Class I	Pro19	5.42	164	17	−27	25	2.9 (0.4)	5.8 (0.7)
	Pro36	5.51	160	19	−24	18	5.4 (0.7)	8.8 (1.0)
Class II	Pro5	5.62	161	33	−42	36	3.2 (0.5)	3.6 (0.5)
	Pro22	6.44	−170	−35	42	−32	3.1 (0.5)	3.1 (1.0)
	Pro41	6.73	158	33	−42	32	6.8 (0.9)	8.6 (1.0)

networks. An example of such behavior is presented in Fig. 10, where we present new ring F attached to rings B and C. The details of the water structure will be published separately.

The final model has 162 water molecules with occupancies ranging from 1.0 to 0.15. The total occupancy of water molecules adds up to 93 fully occupied sites, which is in agreement with previous estimates (Teeter *et al.*, 1993). Water molecules which are close to the protein surface have clear electron density. The water peaks get weaker away from the surfaces, a fact reflected in occupancy of the sites. The second shell of water molecules is generally separable into two distinct networks with occupancies 65 and 35%. Far from the surface of the protein, water networks become less apparent, reflecting the more dynamic nature of the solvent here.

In the middle of cavity between the symmetry-related protein molecules, the water networks break down. At this location we have seen persistent, slightly elevated but uninterpretable difference density ($2.5\sigma \simeq 0.48$ e Å^{-3}), suggesting dynamically disordered water. Because such regions are distributed in a periodic manner along the cell this might indicate that the water positions do not obey the crystal symmetry.

Differences between positions and temperature factors obtained in FMLS and RLS refinements

The FMLS–*RFINE* model has very well defined positions and temperature factors as shown by the e.s.d.'s obtained from the full-matrix refinement. Average positional e.s.d.'s (σ_x, σ_y, σ_z) are 0.022 Å with parameter to e.s.d. ratios (x/σ_x, y/σ_y, z/σ_z) of around 1000 and average e.s.d. for B_{eq} around 0.8 with $B/\sigma_B \simeq 8.5$. The overall structure refined by FMLS–*RFINE* method (Fig. 2) looks almost identical to that of refined by RLS ($R = 10.5\%$, Teeter *et al.*, 1993), as expected. The differences in the backbone φ, ψ angles are small (Fig. 11) and the Ramachandran plot (Fig. 12) looks very similar to that presented before (Teeter *et al.*, 1993). The biggest changes in a backbone conformation were correlated with either disorder or high temperature factors (Fig. 13). The average r.m.s. differences between positions of the backbone atoms is 0.029 Å (comparable to FMLS e.s.d.'s of 0.022 Å) and for side chains, 0.39 Å (including disordered atoms).

The differences in atomic positions of disordered atoms for *PROLSQ* refinement compared to *RFINE* are very small. Small but more pronounced differences are observed in temperature factors as described below.

The analysis of individual temperature factors shows relatively small differences between the *PROLSQ* and *RFINE* refined models. The *RFINE* model has, on average, lower B factors with the difference $(\Sigma|B_{RLS} - B_{FMLS}|/n)$ on backbone and side-chain atoms of 0.35 and 1.14, respectively. The differences in positions between disordered atoms were smaller on average in the RLS than in the FMLS–*RFINE* refinement although some of the atoms differed more. The temperature factors of atoms in alternative positions were always lower and usually much closer in value in

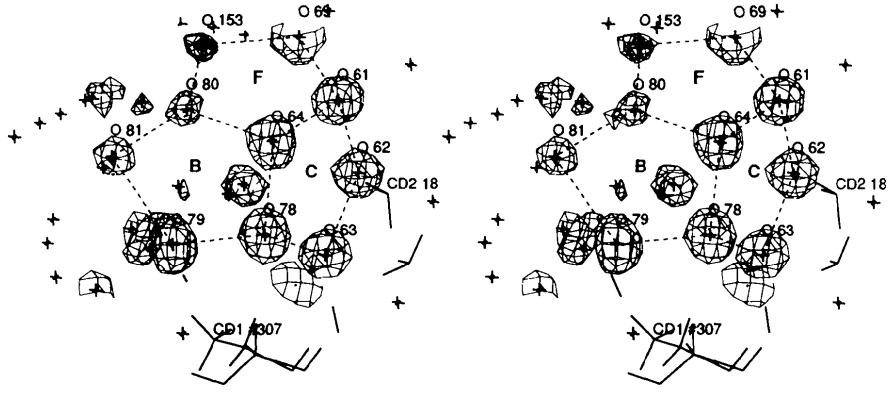

Fig. 10. Electron-density map (1.6σ level) showing several interconnected water rings B, C, F (major ring waters 54, 61 62, 63, 78 have occupancy 1). The newly found ring F is attached to the previously described (Teeter, 1984) rings B and C which are part of pentagonal clathrate structure. Even with diminishing occupancy (O153, $q = 0.35$; O80, $q = 0.50$; O81, $q = 0.65$; other disordered water molecules $q \simeq 0.35$) water molecules have a tendency to form well ordered structures of pentagonal rings.

the FMLS refined model than in the RLS (see Table 2). The comparison of B's and q's for disordered residues observed in those two models are presented in Table 2.

The biggest reduction in B's for disordered residues (1, 8, 12 and 39) is apparent from the B-factor plots (Fig. 13). This can be explained by the averaging property of a sparse-matrix least-squares refinement used in *PROLSQ*. In cases where close atoms in the same group should have very different temperature factors, *PROLSQ* causes the temperature factor to rise for the entire residue (Fig. 13 and Table 2) rather than for individual atoms.

To evaluate the accuracy of modeling of temperature-factor anisotropy by the three-parameter model (*PROLSQ*) *versus* the full six, we have calculated the differences in magnitudes and directions of eigenvectors for *PROLSQ* and FMLS–*RFINE* refined structures. The differences between directions of eigenvectors calculated from *PROLSQ* and the FMLS refined thermal ellipsoids were around 30°. The magnitudes of eigenvectors were very similar indicating similar modeling of 'thermal' motion by both methods.

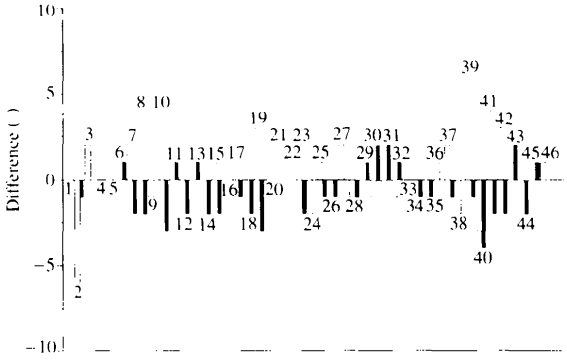

Fig. 11. Differences in backbone torsional angles between FMLS–*RFINE* and *PROLSQ* models. The φ difference is represented by open bars, ψ by full bars.

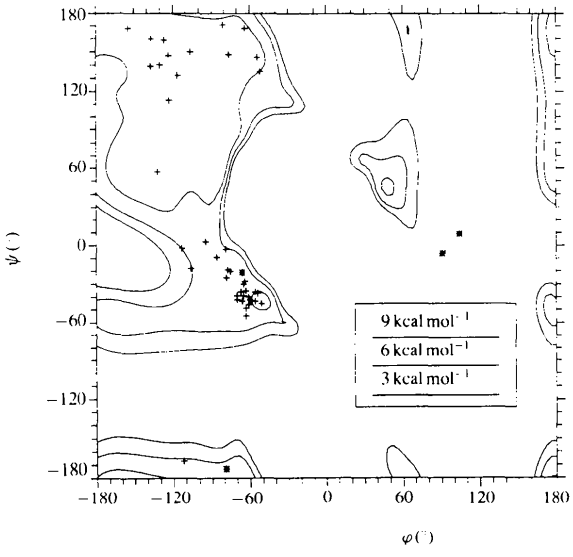

Fig. 12. Ramachandran plot for FMLS–*RFINE* refined model. Glycine residues are denoted by asterisks.

Analysis of individual temperature factors: shape and magnitude of motion

It is believed that the B values of temperature factors in most crystals of small molecules accurately reflect the thermal motion of the molecules. Usually proteins' B factors are more than six times larger than in small molecules. Therefore, the Gaussian distribution interpreted as temperature factors cannot possibly arise from oscillatory motion of individual atoms alone. They have to reflect the packing defects (weak intermolecular forces), high mosaicity and different conformational substates of individual molecules.

In this context, crambin's refinement provides once more a unique opportunity to check traditional interpretations, since B values are much smaller than usual for proteins. The average temperature factor for all protein atoms is 4.64 Å^2 with extensive regions of $B = 3.0$ Å^2. This is fully comparable to that of small molecules. It is interesting though that in certain regions B factors are significantly elevated ($\simeq 8$ Å^2) above this value (Fig. 13).

Inspection of the B tensors provides an interesting insight into the possible origin of the vibrational factors in proteins. The average values of the B tensor for all atoms are $B_{11} = 6.21$, $B_{22} = 7.30$, $B_{33} = 6.71$ Å^2 ($B_{11} = 4.24$, $B_{22} = 4.98$, $B_{33} = 4.73$ Å^2 for protein atoms). The largest B value is associated with the shortest crystal axis $b = 18.49$ Å which is presumably the fastest crystal growth axis. This would imply that packing and static disorder is the main contributor to the temperature factors rather than large average motion in the b direction

The temperature-factor plots provide further insight into potential vibrational units within the protein, later

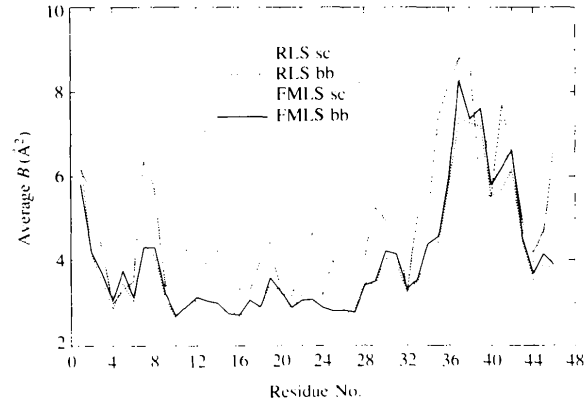

Fig. 13. Comparison of temperature-factor profiles between FMLS–*RFINE* and *PROLSQ* refinements. The backbone is presented by heavy lines and the side chains by thin lines. Broken lines denote *PROLSQ* refined model, continuous lines the FMLS–*RFINE* model.

confirmed by TLS refinement. The molecule can be divided into two distinct parts. The first constitutes a relatively immobile fragment with two α-helices (residues 9–32) with average $B = 3.52$ Å^2, and the second much more mobile fragment with a loop region and the β-sheet (residues 33–46 and 1–7) with average $B = 8.72$ Å^2 (see Figs. 13 and 14). The hinge is located at the junction of the helical and β-sheet regions of the protein. Significantly diminished mobility for the second part of the first strand in the β-sheet can be attributed to the two disulfide bonds (Cys3 and 4) as well as to the strong hydrogen bonds of Arg10. However, it should be noted that the mobility of residues 3–4 is less than expected for one rigid-body model (Fig. 15) and residues 33–44 move more than expected. The best description of motion required partitioning of the molecule into more than one rigid body (see below).

Influence of disorder on the modeling of thermal parameters.

The temperature-factor profile of a protein can provide important information about molecular motion but may also turn out to be deceiving if disorder is involved. For crambin, high temperature factors in the more immobile region of the protein defined above prior to modeling disorder (residues 1, 8, 12 and 6–32) were primarily associated with later modeled disorder. If disordered atoms were properly modeled, their temperature factors were significantly reduced by FMLS refinement (Fig. 13). Temperature factors of disordered residues with previously higher temperature factors (residues 1–4 and 34–46) after modeling of disorder match that of the neighboring residues.

Thr1 illustrates the best the superiority of the FMLS refinement in modeling disorder compared to *PROLSQ* (Fig. 7). This residue refined with high temperature factors in *PROLSQ* (before and after modeling of disorder). In FMLS it refined much better. After modeling the alternative positions ($q = 0.8$ and 0.2), the temperature factors on the backbone as well on the side chain dropped.

Temperature factors of well refined disordered atoms were lower than those in the original model. For example, the values of B factors for individual atoms were lowered in Thr1 from 11 to 6, in Ile7 from 18 to 7, in Arg10 from 10 to 2, in Asn12 from 12 to 4, in Phe13 from 8 to 3, *etc.*

In modeling disorder, as a rule, we tried to achieve a similar level of temperature factors for alternate positions by adjusting q manually. Two attempts of independent occupancy refinement failed presumably because of close spatial proximity of the disordered atoms. If, however, the differences in temperature factors persisted despite the efforts, we inspected the electron-density maps and in most such cases more dynamic disorder was detected (*e.g.* Val8). Here the electron density observed was not well defined and spread beyond reasonable stereochemical limits.

Again we note that without properly modeled disorder the real molecular vibrations cannot be understood. The temperature factors are higher than expected, reflecting

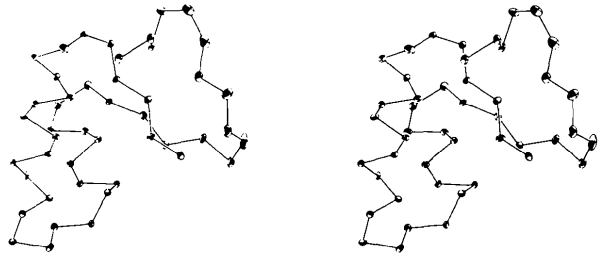

Fig. 14. C_α representation of the FMLS–*RFINE* model of crambin with vibrational ellipsoids at the 70% level.

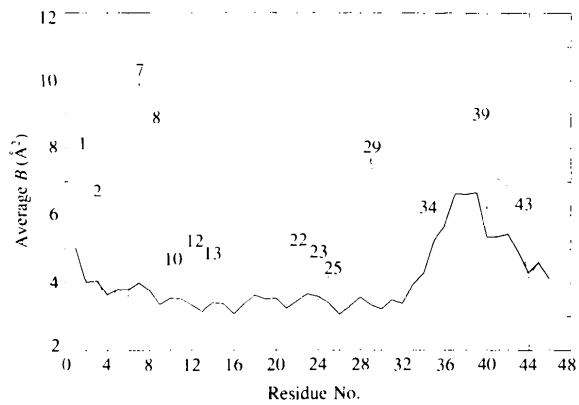

Fig. 15. Comparison of temperature-factor profiles obtained in FMLS–*RFINE* refinement and using the rigid-body TLS refinement: (*a*) one rigid body is depicted by a heavy continuous line and the FMLS–*RFINE* model without disorder modeled by a thin line. The numbers indicate the modeled disordered residues, (*b*) FMLS–*RFINE* refined model as a thin broken line, one rigid body is depicted as a heavy continuous line, two rigid bodies (first: residues 1–6 and 33–46, second: residues 7–32) as a dot–dashed line and three rigid bodies (residues 1–8, 9–31, 32–46, respectively) as a heavy broken line.

deficiencies of the disorder modeling rather than physical reality.

Differences in various stages of refinement: modeling of disorder with two isotropic atoms versus one anisotropic atom (2×5 *versus* 10 *atomic parameters*)

Gradual introduction of disordered atoms improved the electron-density maps and decreased the R factor. B factors of newly created alternates were changed from anisotropic to isotropic. Thus, the number of parameters was not increased, since two isotropic atoms have the same number of free parameters (x, y, z, B, q) as one anisotropic atom, *i.e.* ten parameters. However, on a few occasions, when the atoms were widely separated (Thr2, Ile34, Ile7), we attempted to refine B factors anisotropically (Fig. 3). In all these cases, the atoms chosen for anisotropic refinement had to be converted to isotropic after several refinement cycles when previously defined conversion rules were applied. Ultimately, the B factors dropped for these alternates and electron-density maps improved indicating better modeling.

High correlations are usually found for atoms positioned closer than the resolution limit. For both RLS as well as FMLS refinements, such close atoms have a tendency to move away to less correlated positions at the cost of rising temperature factors and an increased R factor, although FMLS is more sensitive to these problems. The positions of atoms (especially those with low occupancy) reflect a compromise between maintaining program stability and refinement-based positioning. Therefore, for residues which show larger than average differences in temperature factors between disordered mates (Phe13, Tyr29), the difference should be interpreted as an artifact of the method and not as a real vibrational difference.

For many disordered residues refined with FMLS, the stereochemistry has been somewhat distorted from ideality. It is most visible for disorder that involves C_β. Here one C_α—C_β distance is usually shorter and one is longer than an ideal C—C bond.

Modeling temperature factors by TLS

We have attempted to understand crambin's molecular motion by reproducing temperature factors with the rigid-body–TLS method (Schomaker & Trueblood, 1968). The TLS refinement in *RFINE* is run by optimization of structure factors expressed in terms of TLS tensors (real refinement) and not by fitting the TLS tensors to the refined temperature factors. This allows checking of the accuracy of the modeling by inspecting R factors and individual thermal ellipsoids.

The reproduction of the individual temperature factors by the TLS method is not in itself proof that the molecule is a rigid body. Other models of motion can fit the data equally well. However, it also means that the possibility that the molecule is rigid cannot be ruled out. The notion of rigidity has to be supplemented by the additional chemical knowledge about the molecule. As for benzene, it can be argued that delocalized electrons make the peptide bond rigid and secondary-structure hydrogen bonds can add rigidity to helix, sheet and turn.

Another argument for rigidity of proteins comes from the observed local correlations in temperature factors. The apparent correlation between the direction and magnitude of the motion of the adjacent atoms reinforces the notion of the rigidity. The changes in individual torsional angles would create much larger deformations than that observed from temperature factors of distant atoms. The results of diffuse-scattering experiments (Caspar, Clarage, Salunke & Clarage, 1988; Clarage, Clarage, Philips, Sweet & Caspar, 1992) which put the upper limit on the correlation length around 8 Å can be understood by looking at the behavior of the side chains. Side chains are flexible and constitute a buffering zone between a relatively rigid core and mobile solvent. Our studies confirm that correlated disorder between the side chains (Yamano & Teeter, 1994) would explain the results from diffuse-scattering experiments.

In order to assess how rigid crambin is, we have partitioned the structure and conducted TLS refinement to determine tensors for the following numbers of rigid bodies: one (whole protein), two (first: residues 1–6 and 33–46; second: 7–32) and three (residues 1–8, 9–32 and 33–46, respectively). Temperature-factor plots were calculated (Fig. 15) to compare the different TLS partitioning schemes to the temperature factors from unrestrained FMLS refinement. The directions of motion are represented by the components of eigenvectors, and magnitudes of motion are described by eigenvalues expressed in rad^2 for L tensors and in $Å^2$ for T tensors (Table 6, note all tensors and L eigenvalues are multiplied by a factor of 10^3).

The average temperature factor for individual B refinement was identical to that obtained with one rigid body for the TLS method, indicating that the TLS method gave a good approximation of average temperature factors. Modeling the whole molecule as a rigid body gave rotations approximately around the major axes of inertia which pivoted on the center of weight of the molecule. It also showed (Fig. 15) that a few fragments move more than would be explained by simple single rigid-body motion. Some of these were already modeled as disordered (residues 7, 8 and 29) and for others this suggests unmodeled disorder (residues 30, 37, 41 and 46).

Furthermore, Fig. 15(*a*) shows that if the B factors of the structure with unmodeled disorder are plotted then the B's for those residues are much higher than that produced by TLS. Kuriyan & Weis (1991) noticed that if the temperature factors could not be reproduced by the TLS method, the structure probably contained errors of unknown origin. We extend this by identifying that disorder rather than errors in protein conformation is

Table 6. *Summary of the TLS refinement*

TLS tensor elements as well as eigenvectors and eigenvalues for the T and L tensors are presented for one ($T1$, $L1$, $S1$), two ($T21$, $L21$, $S21$, *etc.*) and three rigid-body refinements. Values presented in the table need to be multiplied by 10^{-3} to obtain the T, L, S tensor elements and the eigenvalues for the L tensor. All the tensors are presented in unreduced form as calculated with the origin at $x = 0$, $y = 0$, $z = 0$. Reduced forms can be obtained by applying the Schomaker & Trueblood (1968) formula (20*a*,*b*).

	Tensor elements $\times 10^{-3}$			Eigenvectors			Eigenvalues
*T*1	11	12	13	$\mathbf{V}1_{1,2,3}$	$\mathbf{V}2_{1,2,3}$	$\mathbf{V}3_{1,2,3}$	$\lambda_{1,2,3}$
11	55.86816	−9.63520	−4.13496	0.2419	0.7444	−0.6223	0.0369
21	−9.63520	46.82086	8.16880	0.6265	0.3699	0.6860	0.0457
31	−4.13496	−8.16880	45.02290	−0.7409	0.5559	0.3769	0.0652
*L*1							$\lambda \times 10^{-3}$
11	0.06114	−0.03178	0.07830	−0.1382	0.9673	0.2126	0.1631
21	−0.03178	0.16088	−0.01051	0.9315	0.1999	−0.3039	0.0288
31	−0.07830	−0.01051	−0.26182	0.3365	−0.1560	0.9287	0.2920
*S*1							
11	0.84486	−0.53800	−0.49355				
21	−1.01888	−0.08355	0.93353				
31	2.43200	−0.89093	−0.76131				
*T*21	11	12	13	$\mathbf{V}1_{1,2,3}$	$\mathbf{V}2_{1,2,3}$	$\mathbf{V}3_{1,2,3}$	$\lambda_{1,2,3}$
11	95.20644	9.40192	−34.2446	−0.0752	−0.9159	−0.3943	0.0301
21	9.40192	42.09945	−29.5996	−0.8011	0.2909	−0.5231	0.0694
31	−34.2446	−29.5996	105.4156	−0.5938	−0.2765	0.7556	0.1432
*L*21							$\lambda \times 10^{-3}$
11	−0.34504	−0.10600	−0.11188	−0.6239	−0.7672	0.1489	0.1879
21	−0.10600	0.25415	−0.10300	−0.2221	0.3567	0.9074	0.0581
31	0.11188	−0.10300	0.12600	0.7493	−0.5331	0.3929	0.6791
*S*21							
11	1.90154	2.22625	−4.51915				
21	−3.59601	−0.05247	2.62805				
31	2.28203	1.43730	−1.84907				
*T*22	11	12	13	$\mathbf{V}1_{1,2,3}$	$\mathbf{V}2_{1,2,3}$	$\mathbf{V}3_{1,2,3}$	$\lambda_{1,2,3}$
11	43.12774	−8.38651	−3.72965	0.3178	1.000	−0.2796	0.0200
21	−8.38651	49.07728	8.58987	−0.2902	1.000	0.3256	0.0762
31	−3.72965	8.58987	46.50221	0.6780	−0.4394	1.000	0.0425
*L*22							$\lambda \times 10^{-3}$
11	0.03772	−0.00207	0.07732	0.2482	−0.9444	0.2155	0.1127
21	−0.00207	0.12010	0.03470	−0.8953	−0.1387	0.4233	0.0008
31	0.07732	0.03470	0.17575	0.3699	0.2980	0.880	0.2200
*S*22							
11	0.55561	−0.59422	−0.34093				
21	−0.22981	−0.11774	0.89064				
31	1.45499	−0.70324	−0.43787				
*T*31	11	12	13	$\mathbf{V}1_{1,2,3}$	$\mathbf{V}2_{1,2,3}$	$\mathbf{V}3_{1,2,3}$	$\lambda_{1,2,3}$
11	383.1125	−310.561	−78.6429	0.5047	0.6289	−0.05915	0.0883
21	−310.561	377.5527	42.57727	0.4946	0.3509	0.7951	0.0363
31	−78.6429	42.57727	66.45299	−0.7076	0.6938	0.1339	0.7025
*L*31							$\lambda \times 10^{-3}$
11	0.87762	0.65299	−0.49619	−0.7815	−0.9980	1.000	2.3134
21	0.65299	0.98244	−0.76511	−0.6067	1.000	0.4537	0.2265
31	−0.49619	−0.76511	1.19519	0.8043	0.3395	1.000	0.5472
*S*31							
11	−11.4516	14.90596	0.26886				
21	−17.7404	15.36622	43.1908				
31	18.47816	−17.0417	−3.91463				
*T*32	11	12	13	$\mathbf{V}1_{1,2,3}$	$\mathbf{V}2_{1,2,3}$	$\mathbf{V}3_{1,2,3}$	$\lambda_{1,2,3}$
11	38.88821	−5.40044	−4.29081	0.9074	0.4203	0.0008	0.0364
21	−5.40044	48.02494	9.24738	0.2550	−0.5520	0.7939	0.0372
31	−4.29081	9.24738	45.02925	−0.3341	0.7202	0.6081	0.0583
*L*32							$\lambda \times 10^{-3}$
11	0.02240	0.01270	0.07520	−0.3085	−0.7664	−0.5634	0.1913
21	0.01270	0.17150	0.01998	0.3549	−0.6423	0.6793	0.1433
31	0.07520	0.01998	0.12295	0.8825	−0.0096	−0.4702	−0.0178
*S*32							
11	0.56203	−0.58927	−0.12254				
21	−0.45041	0.10095	0.98489				
31	0.97920	−0.39228	−0.66298				

Table 6 (*cont.*).

	Tensor elements $\times 10^{-3}$			Eigenvectors			Eigenvalues
*T*33	11	12	13	$V1_{1,2,3}$	$V2_{1,2,3}$	$V3_{1,2,3}$	$\lambda_{1,2,3}$
11	92.61140	9.86204	−36.5171	−0.0211	−0.9566	−0.2907	0.0367
21	9.86204	43.70368	−23.9276	−0.8347	0.1269	−0.5215	0.0677
31	−36.5171	−23.9276	118.0423	−0.5503	−0.2316	0.8022	0.1500
*L*33							$\lambda \times 10^{-3}$
11	0.21742	−0.16090	0.10766	0.3437	0.7727	0.5337	0.0228
21	−0.16090	0.17893	−0.12238	0.6788	0.1883	−0.7098	0.0602
31	0.10766	−0.12238	0.13070	0.6489	−0.6062	0.4598	0.4440
*S*33							
11	2.79438	0.81988	−4.28831				
21	−2.96944	−0.07802	2.49858				
31	2.58127	0.93512	−2.71636				

a primary cause of this discrepancy in high-resolution structures.

We would propose a method for identifying the disordered regions in proteins refined at high resolution. If the value of an average temperature factor for a particular residue (before modeling disorder) exceeds significantly the value obtained by the TLS method with one rigid body, then it is highly probable that the residue would exhibit disorder.

A single rigid body was sufficient to describe the individual temperature factors relatively well (Fig. 15*b*). The differences between the temperature factors obtained from individually refined thermal factors (FMLS) and those reproduced by the TLS method with one rigid body were less than 2 Å^2. The magnitude of rotational motion as estimated from the *L* tensor (see Table 6) varies between 1.6°^{2} and 7.6°^{2} with directional cosines of the rotational axis with largest the motion being 0.2126, –0.3039, 0.9287. Translational motion is small with a magnitude around 0.5×10^{-1} Å^2 as estimated from the reduced *T* tensors. The numbers above are fully comparable with those derived for lysozyme (6.9°^{2}, 1.0×10^{-1} Å^2, Sternberg, Grace & Philips, 1979) but indicate decreased mobility for crambin as would be expected from the lower average temperature factor in crambin crystals.

When a single rigid body does not fit the *B*-value profile well, then the fragments exhibiting higher mobility need to be treated separately. After identifying high temperature factors originating from disorder, we tried to model the molecular motion by dividing the molecule into more rigid-body segments and refining the separate TLS tensors for each fragment (Fig. 15).

Modeling of two and three rigid bodies not only improved the description of the temperature-factor profile but also revealed the different mobility of the various molecular fragments (see Fig. 15 and Table 6). The direction and magnitude of the motion is presented in Table 6 and changes in mobility between different fragments can be seen by comparing values of *T* and *L* tensors as well as their eigenvalues and eigenvectors.

The *L* tensor is independent of the origin of coordinate system but *T* and *S* tensors depend on origin of the coordinate system in which the tensors were calculated. The *S* tensor describes the correlation between the translation (*T*) and libration (*L*) of the molecule. If the coordinate system is rotated to the main pivots of the *L* tensor and shifted to make the *S* tensor symmetric, then the reduced *T* tensor (^rT) becomes minimal and independent on the choice of origin (Schomaker & Trueblood, 1968).

We hypothesize that if we can isolate a common part of all the tensors obtained for different rigid bodies, we can hope to isolate the internal modes of vibration from external ones. If the common part is independent of the choice and number of the rigid bodies used in the refinement, it will reflect the maximal allowable common motion of all the fragments. This motion would represent the motion of the lattice rather than individual separate fragments or molecules and by definition this motion would be called external. This definition is close to Diamond's (Diamond, 1990) interpretation of the results of normal mode refinement of bovine pancreatic trypsin inhibitor (BPTI). In order to partition the motion, we have taken six sets of TLS tensors obtained from one, two, and three rigid-body partitions and extracted the largest common parts of all the tensors. In the calculation we have used the *S*, *L* and reduced ^rT tensors according to the formulae,

$$^rT_i = {}^rT_c + {}^rT_{si},\ L_i = L_c + L_{si},\ S_i = S_c + S_{si}.$$

The symbols ^rT_c and L_c denote the biggest common part and the symbols $^rT_{si}$ and L_{si} the 'characteristic parts' of the every single rigid-body *T*, *L* and *S* tensors. The biggest common part was derived by the requirement that the 'characteristic part' of all six tensors would be minimal. So we searched for a maximal common value (A_{ij}) of each tensorial element in all six matrices representing the appropriate tensors.

As described above, the common parts have been interpreted as a lattice contribution to molecular motion. Fig. 16 shows the Cα representation of crambin with ellipsoids obtained from one rigid-body TLS model, from external (common) modes of vibration and from internal ones. The internal representation was taken by subtracting the common part (external) from the one rigid-body TLS representation. The comparison shows that an average contribution from external modes is more

than 60% to the total mobility of the fragments. This finding is in rough agreement with Diamond (1990) who stated that external vibrations dominate. It is also in qualitative agreement with the results for small molecules (Willis & Pryor, 1975).

It is important to note that this result calls into question numerous attempts to reproduce the distribution of temperature factors by molecular dynamics (*e.g.* van Gunsteren, Berendsen, Hermans, Hol & Postma, 1983; Yu, Karplus & Hendrickson, 1985). By definition, molecular dynamics should reproduce only the internal motion of the individual molecules (less than 40% of the X-ray determined B factors) but should not describe the lattice contribution (more than 60%) because of periodic removal of the motion of the center of mass. While sometimes the average mobility is smaller in molecular dynamics (Teeter & Case, 1990) than that observed in X-ray refinement (Teeter & Hendrickson, 1979), in the vast majority of other examples the molecular dynamics produces higher temperature factors than X-ray crystallography (for example, Yu *et al.*, 1985).

The mobility would be significantly reduced in simulations carried out in crystalline and liquid environment but in most cases it is still larger than detected in X-ray refinement. This can be attributed partially to temperature-factor restraints used in protein refinement. However, the omission of the collective lattice vibrations in molecular-dynamics simulations would suggest a deficiency in the description of motion by molecular dynamics rather than the insufficient modeling of temperature factors by X-ray crystallography that Kuriyan suggested (Kuriyan, Petsko, Levy & Karplus, 1986).

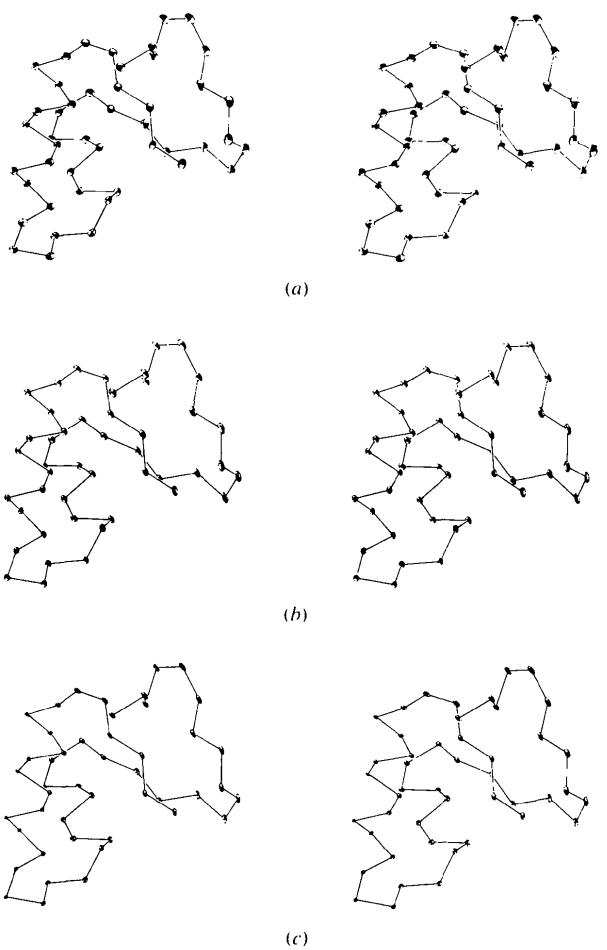

Fig. 16. C_α representation of crambin with vibrational ellipsoids at the 70% level derived from: (*a*) one rigid-body TLS refinement, (*b*) common part of TLS tensors representing external modes of vibration, (*c*) leftover part of one rigid-body TLS tensors representing internal motion.

Differences between F^2 and F refinement

Except for minor changes there are no geometrical differences between the models obtained with either of the refinement schemes for FMLS–*RFINE* refinement. There are, however, major differences in stability, speed of reaching the minimum and accuracy of modeling the temperature factors of individual atoms.

The refinement on F seems to be more stable and a particular local minimum is reached in fewer cycles. The stability is measured by the extent of change in individual parameters (shifts). In a stable refinement, the shifts diminish from cycle to cycle, while in an unstable one, they grow and refinement diverges. In the case of unstable refinement, the scale factor started to oscillate with increasing magnitude until the program reached unreasonably high R values, despite strong damping imposed in such cases on positional and on B-factor shifts (60%).

Additional advantages in using the F mode of the refinement were observed in modeling disorder. When F^2 refinement was chosen, a few atoms with partial occupancy always drifted away from one cycle to the next (*e.g.* Val7 CG1 or Ile34 CD1), never reaching a stable position or temperature-factor level. When refinement on F was chosen, they stabilized and settled into reasonable positions. We noticed that the lower the occupancy assigned to the atom, the more difficult it was to reach a stable minimum. It was necessary at times to fix the position or temperature factor for several cycles and slowly release them.

Finally, the level of the R factor was always lower in F refinement indicating better modeling of structure factors and intensities. When F^2 refinement was chosen, certain local minima were lost. This could be observed in an experiment when we changed the refinement mode from F to F^2. The R factor, in this case, increased until it stabilized at a higher level. The initial R-factor value was almost identical to that obtained in F refinement. The final R value rose by 0.5%. Incidentally, identical observations were made in small-molecule refinement where structures 'differed slightly, but significantly' and

the R factor for refinement against structure factors converged to the lower value by 0.5% (Freer & Kraut, 1965). A compilation of different refinement statistics for a variety of models is presented in Table 3.

The differences described might be at least partially attributed to the weighting scheme used by the program. Employment of different weighting schemes for F and F^2 refinement than those used in *RFINE* (G. Sheldrick, personal communication) should certainly change the situation bringing R factors in both cases closer together.

Discussion and conclusions

We have shown that it is feasible to refine a small protein (large small-molecule, 917 atoms) by a standard full-matrix least-squares method. The final model has good geometry and well defined positional and thermal factors comparable to that of small molecules. The averaged estimated standard deviations (e.s.d.) of positional parameters or, as it is commonly interpreted, positional errors obtained from the FMLS-*RFINE* correlation matrix are 0.0096 Å for backbone atoms, 0.0168 Å for side-chain atoms and 0.0409 Å for solvent atoms giving for all atoms an average positional e.s.d.'s = 0.022 Å. The thermal e.s.d.'s (0.71 $Å^2$ for the protein atoms and 0.94 $Å^2$ for all atoms) are also comparable to that from small molecules.

Modeling disorder played a significant role in achieving convergence and a good agreement between calculated and observed structure factors. The total extent of disorder encountered during the refinement (30%) was more than expected (15%, Smith *et al.*, 1988) but in line with 15% disorder for two heterogeneous sequences. Modeling of this disorder proved crucial to lowering the R factor and reaching a stable minimum.

With increasing resolution and decreasing R value, more and more disorder becomes apparent. There are some indications (Fig. 15*a*) that despite such a detailed modeling as presented in this paper, not all the disorder has been modeled. Even more accurate data would be necessary to resolve such disorder.

Modeling of disorder proved to diminish the overall temperature-factor variation. Because the FMLS method does not have restraints, it was a sensitive tool for detecting disorder. However, it also proved to be difficult to handle and slower to converge (more time taken by one cycle) than other methods that use stereochemical restraints. The very limited bias introduced by the RLS method leads us to recommend the use of *PROLSQ* (or any other restrained-refinement method) as the faster, more stable and easier to handle program. Despite this recommendation, in very high resolution structures, the FMLS method would be desirable as an ultimate tool for providing a less biased picture of protein structure.

Successful rigid-body refinement calls into question the accuracy of using Luzzati plots for estimating protein-refinement errors. If the molecule is moving as a rigid body, as it is suggested by this and other studies (Sternberg *et al.*, 1979; Diamond, 1990), then the crucial assumption of the random positional errors of the Luzzati theory does not hold. Errors in atomic positions are not distributed randomly because B factors increase with distance from the center of a rigid body (Fig. 17), and the different shells of atoms do not contribute equally to the R factor as a function of resolution. On the other hand, Luzzati theory has provided a reasonable estimate of positional errors so far, so these two conflicting ideas have to be somehow reconciled. A reconciliation comes from the fact that the largest contribution to the individual temperature factors may come from lattice vibrations (~60%) which are contributing equally to all the atoms.

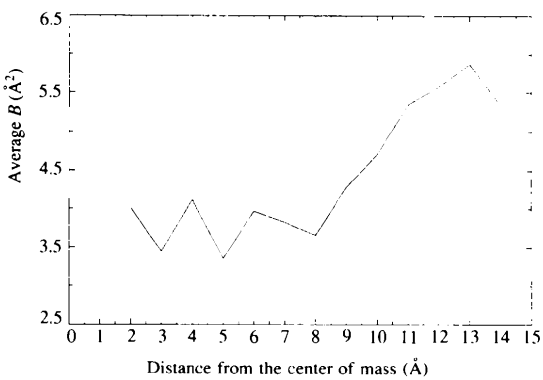

Fig. 17. Plot of average B factors for atoms in spherical shells at an increasing distance from the center of mass of the molecule. Because the center of mass lies outside of the molecule in a cleft between helices and β-sheet, the initial shells do not represent well the behavior of internal protein atoms.

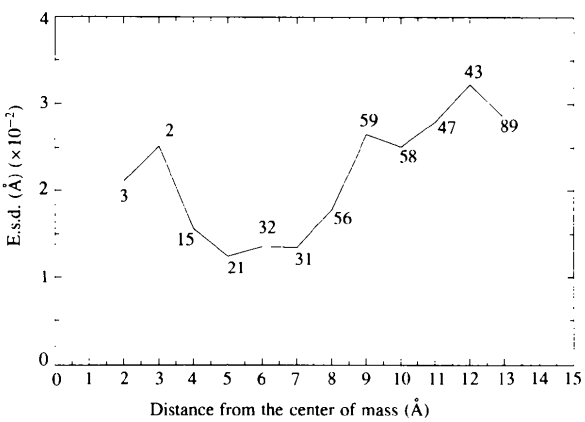

Fig. 18. Plot of e.s.d.'s for all atoms estimated from the FMLS method in the same spherical shells as described in Fig 17. The number of atoms in a particular shell is indicated along the curve. Those numbers differ slightly between figures because some atoms were restrained during the course of the refinement.

As expected for rigid-body motion, the e.s.d.'s for different atoms change with the distance from the rigid-body center (Fig. 18), but they are in fact quite close to the estimates obtained from the Luzzati plot. The magnitude of the overall e.s.d.'s (0.022 Å) is only 2–3 times smaller than a positional estimate of 0.055 Å obtained from a Luzzati plot (Fig. 19). Combining this information with the opinion of Stout & Jensen (1989) that errors obtained from e.s.d.'s are at least two times underestimated leads to the conclusion that Luzzati estimates, despite their weak theoretical basis for protein rigid-body motion, are reliable for crambin.

The restraints used in *PROLSQ*, one of the restrained protein refinement programs, seem to be different not only from the newly compiled Engh & Huber parameters but also from the average values obtained in this refinement. Therefore, we would recommend new values for the atomic parameters be adopted from this study (Table 4). Other important recommendations we can derive from our refinements are: the weights for planar group restraints (peptide bonds, ω-angles) as well as the restraints for on temperature factors for neighboring atoms are generally too strong. Secondly, the side-chain conformations, rotameric positions (staggered, χ angles), should be restrained more strongly. Third is that the least-squares refinement should be run on amplitudes (F) instead of on corrected intensities (F^2).

The increase of temperature factors as calculated with the distance from the backbone (C_α, C_β, C_γ, *etc.*) is relatively small (data not shown). The slope is much smaller than that obtained from molecular-dynamics simulations (Yu *et al.*, 1985), indicating much stronger correlation between bonded atoms. However, we recommend *PROLSQ* refinement be run with weak temperature-factor restraints for neighboring atoms not because it would improve modeling of the real motion but because it would allow one to detect modeling errors much faster than in strongly restrained refinement.

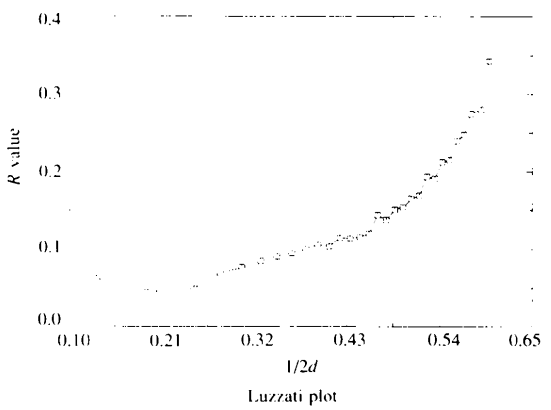

Fig. 19. Luzzati plot indicating estimated positional errors around 0.055 Å. The dashed lines, in descending order, indicate lines for 0.07, 0.06, 0.05 Å positional errors.

The analysis of FMLS temperature factors and its comparison to the TLS refined temperature factors provided the first step in defining the criterion for selecting the disordered regions. The analysis of motion as modeled by rigid-body motion provided information about molecular mobility in a more quantitative way. This information can be used to interpret the molecular-dynamics simulations (Teeter & Case, 1990).

The TLS refinement provided dynamical information necessary to understand the molecular coupling between individual molecules and the origin of temperature factors. The rigid-body motion was partitioned into internal end external modes of vibration. The external modes (lattice vibration) proved to contribute the most (> 60%) to the individual temperature factors even at low temperature. Internal modes of vibrations might dominate at temperatures lower than 130 K. However, even at this temperature which is lower than the postulated temperature of 200 K (Parak & Knapp, 1984) above which proteins can act as catalysts (Rasmussen, Stock, Ringe & Petsko, 1992), the external modes dominate.

The accuracy of atomic details resolved by the FMLS method will provide a valuable source of information for understanding the protein's static as well as dynamic behavior. Different aspects of theoretical methods such as molecular dynamics can now be tested against the structure presented. Thus, it should be possible to address problems associated with multiple minima (reproduction of the observed disorder), the harmonic approximation (directions of thermal ellipsoids) and the reproduction of realistic temperature factors (magnitudes of motion) in a more quantitative fashion.

Some of the problems associated with the disordered residues described in this paper were alleviated with the use of the recently released program *SHELXL*93 by Sheldrick. For comparison, we have refined the final model with *SHELXL*93 using 1σ and 2σ cutoffs for the data. The results are summarized in Table 3. Despite different mode of refinement (F^2 *versus* F) we obtained almost identical R factors and e.s.d.'s but with better goodness-of-fit (GOF) which was close to unity. Nevertheless, the stereochemistry of the final model was almost identical to that obtained with *RFINE*–FMLS but with considerably lower σ's on the distributions of bonds and angles (Table 4). In addition, some small distortions at the disordered sites were detected.

In the future, we plan to repeat the full-matrix refinement with *SHELXL*93 on the newly collected data for one of the pure forms of crambin. The reduction in the number of disordered residues should help to resolve better the water structure and allow a clearer interpretation of the temperature factors. Problems with closely positioned atoms might be alleviated by application of selective restraints.

We would like to thank Drs G. Sheldrick, H. Bürgi and J. Dunitz for in-depth discussions and their comments.

References

ASHIDA, T., TSUNOGAE, Y., TANAKA, I. & YAMANE, T. (1987). *Acta Cryst.* B**43**, 212–218.

BHAT, T. N., SASISEKHARAN, V. & VIJAYAN, M. (1979). *Int. J. Peptide Protein Res.* **13**, 170–184.

BLUNDELL, T., BARLOW, D., BORKAKOTI, N. & THORNTON, J. (1983). *Nature (London)*, **306**, 281–283.

CASPAR, D. L. D., CLARAGE, J. B., SALUNKE, D. M. & CLARAGE, M. S. (1988). *Nature (London)*, **332**, 659–662.

CLARAGE, J. B., CLARAGE, M. S., PHILIPS, W. C., SWEET, R. M. & CASPAR, D. L. D. (1992). *Proteins*, **12**, 145–157.

DIAMOND, R. (1990). *Acta Cryst.* A**46**, 425–435.

DUNITZ, J. D., SCHOMAKER, V. & TRUEBLOOD, K. N. (1988). *J. Phys. Chem.* **92**, 856–867.

ENGH, R. A. & HUBER, R. (1991). *Acta Cryst.* A**47**, 392–400.

FINGER, L. W. & PRINCE, E. (1975). *Natl Bur. Stand. (US) Tech. Note* 854. *A System of Fortran* IV *Computer Programs for Crystal Structure Computations.*

FRAUENFELDER, H., PARAK, F. & YOUNG, R. D. (1988). *Annu. Rev. Biophys. Biophys. Chem.* **17**, 451–479.

FREER, S. T. & KRAUT, J. (1965). *Acta Cryst.* **19**, 992–1002.

VAN GUNSTEREN, W. F., BERENDSEN, H. J. C., HERMANS, J., HOL, W. G. & POSTMA, J. P. M. (1983). *Proc. Natl Acad. Sci. USA*, **80**, 4315–4319.

HARRIS, G. W. & MOSS, D. S. (1992). *Acta Cryst.* A**48**, 42–45.

HENDRICKSON, W. A. & KONNERT, J. H. (1980). *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN, pp. 13.01–13.23. Bangalore: Indian Academy of Sciences.

HOPE, H. (1988). *Acta Cryst.* B**44**, 22–26.

KURIYAN, J., PETSKO, G. A., LEVY, R. M. & KARPLUS, M. (1986). *J. Mol. Biol.* **190**, 227–254.

KURIYAN, J. & WEIS, W. I. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 2773–2777.

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

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

PARAK, F. & KNAPP, E. W. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 7088–7092.

PILLING, D. E., CRUICKSHANK, D. W. J., BUJOSA, A., LOVELL, F. M. & TRUTER, M. R. (1961). *Computing Methods and Phase Problem in X-ray Crystal Analysis*, p. 32. Oxford: Pergamon Press.

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

RASMUSSEN, B. F., STOCK, A. M., RINGE, D. & PETSKO, G. A. (1992). *Nature (London)*, **357**, 423–424.

SCHOMAKER, V. & TRUEBLOOD, K. N. (1968). *Acta Cryst.* B**24**, 63–76.

SHELDRICK, G. M. (1993). *SHELXL*93, *Program for Crystal Structure Refinement*, Univ. of Göttingen, Germany.

SMITH, J. L., CORFIELD, P. W. R & HENDRICKSON, W. A. (1988). *Acta Cryst.* A**44**, 357–368.

STERNBERG, M. J. E., GRACE, D. E. P. & PHILIPS, D. C. (1979). *J. Mol. Biol.* **130**, 231–253.

STOUT, G. H. & JENSEN, L. H. (1989). *X-ray Structure Determination*. New York: John Wiley.

TEETER, M. M. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 6014–6018.

TEETER, M. M. (1985). *Proceedings of Workshop on Molecular Dynamics and Protein Structure*, Univ. of North Carolina, Chapel Hill, North Carolina, USA.

TEETER, M. M. & CASE, D. (1990). *J. Phys. Chem.* **94**, 8091–8097.

TEETER, M. M. & HENDRICKSON, W. A. (1979). *J. Mol. Biol.* **127**, 219–223.

TEETER, M. M. & HOPE, H. (1986). *Trans. NY Acad. Sci.* **482**, 163–165.

TEETER, M. M., ROE, S. M. & HEO, N. H. (1993). *J. Mol. Biol.* **230**, 292–311.

VAN ETTEN, C. H., NIELSEN, H. C. & PETERS, J. E. (1965). *Photochemistry*, **4**, 467–473.

WILLIS, B. T. M. & PRYOR, A. W. (1975). *Thermal Vibrations in Crystallography*. Cambridge Univ. Press.

YAMANO, A. & TEETER, M. M. (1994). *J. Biol. Chem.* **269**, 13956–65

YU, H., KARPLUS, M. & HENDRICKSON, W. A. (1985). *Acta. Cryst.* B**41**, 191–201.