# Structural Parameters for Proteins Derived from the Atomic Resolution (1.09 Å) Structure of a Designed Variant of the *ColE1* ROP Protein

Metaxia Vlassi,[a] Zbigniew Dauter,[b] Keith S. Wilson[b] and Michael Kokkinidis[c,d]*

[a]*National Centre for Scientific Research 'DEMOCRITOS', 15310 Ag. Paraskevi-Attikis, Athens, Greece,* [b]*EMBL, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany,* [c]*University of Crete, Department of Biology, PO Box 2208, GR-71409 Heraklion, Crete, Greece, and* [d]*Institute of Molecular Biology and Biotechnology, PO Box 1527, GR-71110 Heraklion, Crete, Greece. E-mail: kokkinid@imbb.forth.gr*

## Abstract

The crystal structure of a designed variant of the *ColE1* repressor of primer (ROP) protein has been refined with *SHELXL*93 to a resolution of 1.09 Å. The final model with 510 non-H protein atoms, 576 H atoms in calculated positions and 114 water molecules converged to a standard *R* factor of 10% using unrestrained blocked full-matrix refinement. For all non-H atoms six-parameter anisotropic thermal parameters have been refined. The majority of atomic vibrations have a preferred orientation which is approximately perpendicular to the bundle axis; analysis with the TLS method [Schomaker & Trueblood (1968). *Acta Cryst.* B**24**, 63–77] showed a relatively good agreement between the individual atomic displacements and a rigid-body motion of the protein. Disordered residues with multiple conformations form clusters on the surface of the protein; six C-terminal residues have been omitted from the refined model due to disorder. Part of the solvent structure forms pentagonal or hexagonal clusters which bridge neighbouring protein molecules. Some water molecules are also conserved in wild-type ROP. The unrestrained blocked full-matrix least-squares refinement yielded reliable estimates of the standard deviations of the refined parameters. Comparison of these parameters with the stereochemical restraints used in various protein refinement programs showed statistically significant differences. These restraints should be adapted to the refinement of macromolecules by taking into account parameters determined from atomic resolution protein structures.

## 1. Introduction

Compared with small-molecule crystals, which typically diffract to 1.0 Å or better, the resolution obtained from crystals of macromolecules is usually limited to less than atomic. So far, very few cases of protein crystals diffracting beyond 1.2 Å – which might be regarded as a threshold for atomic resolution – have been characterized. Whereas at 1.0 Å resolution the ratio of observa-

tions to parameters is approximately 5:1 (which is usually sufficient for the refinement of an anisotropic atomic model against X-ray data alone), the less than atomic resolution which is usual for proteins results in a much lower ratio. This dictates the need of additional stereochemical restraints in protein refinement which effectively increase the observations-to-parameters ratio to a value sufficient for the definition of a good least-squares minimum. This additional stereochemical information consists of geometric parameters (bond lengths, angles, torsion angles, planarity, chirality *etc.*) which are used as 'target' values. These values must be as accurate as possible, particularly when refinement against low-resolution data is performed. 'Target' values are typically obtained from small-molecule structures determined at atomic resolution and deposited with the Cambridge Structural Database (Allen *et al.*, 1983). The most recent compilation of peptide bond-length and angle parameters for macromolecular refinement is based on small-molecule structural data and has been reported by Engh & Huber (1991). These parameters, however, may not be completely valid for proteins: although it is possible that the mean values of bond lengths and angles are the same for peptides and proteins, it is not obvious that their natural variability (*i.e.* the distribution about their means) should be the same because of differences in their specific environments, as also noted by other investigators (Wilson *et al.*, 1998). On the other hand, stereochemical parameters obtained from structures deposited with the Brookhaven Protein Data Bank are biased by the restraints applied in the crystallographic refinement, which leave their mark to the final model (Laskowski *et al.*, 1993). Even from protein structures refined close to atomic resolution, there is no direct and reliable information on the stereochemical parameters and their standard errors. This is because of the lack of estimation of parameter variances which requires an unrestrained full-matrix or blocked full-matrix refinement.

Compared to small-molecule crystallography and with the exception of the structure crambin determined at 130 K (Stec *et al.*, 1995), all proteins refined to reso-

Table 1. *Summary of data collection and statistics*

| Data collection | High resolution | Medium resolution | Low resolution | |
|---|---|---|---|---|
| Synchrotron beamline | X11 | X11 | X11 | X31 |
| Wavelength (Å) | 0.92 | 0.92 | 0.92 | 0.92 |
| Crystal-to-plate distance (mm) | 75 | 125 | 230 | 500 |
| Minimum resolution (Å) | 2.5 | 3.5 | 30 | 30 |
| Maximum resolution (Å) | 1.09 | 2.5 | 2.5 | 2.5 |
| Oscillation per image (°) | 1.0 | 2.5 | 4.0 | 12.0 |
| Number of images | 139 | 56 | 45 | 15 |
| Statistics of merged data | | | | |
| Space group | $C2$ | | | |
| Unit-cell dimensions (Å, °) | $a = 47.09$, $b = 37.88$, $c = 31.65$, $\beta = 100.8$ | | | |
| No. of unique/total reflections | 22025/105106 | | | |
| $R_{merge}$† (%) | 4.5 | | | |
| Completeness (23.1–1.09 Å) (%) | 98.2 | | | |
| Overall $B$ factor (from Wilson plot) (Å$^2$) | 9.7 | | | |

| Resolution range (Å) | Completeness (%) | $R_{merge}$ (%)† |
|---|---|---|
| 30.0–2.99 | 98.7 | 3.8 |
| 2.99–2.00 | 96.2 | 4.2 |
| 2.00–1.55 | 99.8 | 7.1 |
| 1.55–1.30 | 98.2 | 9.1 |
| 1.30–1.19 | 98.6 | 13.4 |
| 1.19–1.09 | 97.5 | 19.7 |

† $R_{merge}(I) = \sum |I - \langle I \rangle| / \sum I$.

lutions close to 1 Å [*e.g.* the avian pancreatic polypeptide (Glover *et al.*, 1983), rubredoxin (Dauter *et al.*, 1992) and bovine pancreatic trypsin inhibitor determined at 125 K (Parkin *et al.*, 1996)] do not take full advantage of the high resolution, as no unrestrained full-matrix refinement against X-ray data alone has been performed. One reason for the choice of restrained *versus* unrestrained refinement in the above studies had been the poor statistics obtained from the stereochemistry of the unrestrained models when evaluated with the commonly used library values which, as already pointed out, may be inappropriate for proteins. A further paradox has been the deterioration of the refinement statistics at high resolution when compared with lower resolution cases, a behaviour which is most probably related to the use of 'improper' libraries in refinement programs (Laskowski *et al.*, 1993). Finally, anisotropic refinement which is essential for the description of the atomic motion, has been used rarely (Glover *et al.*, 1983; Stec *et al.*, 1995) or only partially as in an approach using a three-parameter anisotropic parameterization of temperature factors (Teeter *et al.*, 1993).

In this study we report the crystallographic refinement of a variant of the *ColE1* ROP protein at a nominal resolution of 1.09 Å. The dimeric ROP protein has become a rich source of information about the structural (Banner *et al.*, 1987) and thermodynamic properties of 4-$\alpha$-helical bundle proteins. On the basis of the wild-type protein structure, a number of ROP variants has been designed with the aim of addressing specific questions on protein folding (Paliakasis &

Kokkinidis, 1992). ROP variants with redesigned connections between the antiparallel $\alpha$-helices of the bundle are particularly important, as changes in these regions are expected to affect drastically the stability of the protein. In a recent paper (Vlassi *et al.*, 1994) we described in detail the 1.4 Å X-ray crystal structure and thermodynamic properties of a ROP mutant carrying two extra Ala residues in the bend region. This mutant (hereafter referred to as $\langle 2aa \rangle$) has been designed with the aim of characterizing specific aspects of the relationship between sequence periodicity and the folding of 4-$\alpha$-helical bundles (Banner *et al.*, 1987). The three-dimensional structure of this stable highly regular 4-$\alpha$-helical bundle (Fig. 1) is mainly determined by hydrophobic core packing interactions and was shown to tolerate an unfavorable local geometry at the region (Vlassi *et al.*, 1994). Despite the excellent quality of the crystals used in the original structure determination, further improvements of the crystallization strategy were sought; these led to an extension of the diffraction
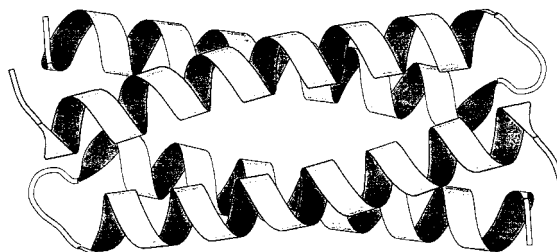


Fig. 1. Schematic representation of the structure of $\langle 2aa \rangle$.

limit to the atomic resolution range, *i.e.* to a nominal resolution of 1.09 Å.

The availability of high-resolution synchrotron data and the use of the program *SHELXL*93 (Sheldrick, 1993) with several useful options, including estimation of standard deviations on refined parameters, and six-parameter anisotropic temperature factor protein refinement, make the refinement of ⟨2aa⟩ a valuable source of information about macromolecular geometrical parameters and crystallographic refinement strategies at atomic resolution.

## 2. Experimental procedures

### 2.1. *Crystallization*

The crystals of ⟨2aa⟩ used in the present study, were obtained with seeding techniques. Crystals from an initial seed stock, grown as was previously described (Kokkinidis *et al.*, 1993), were first washed in water and subsequently in a sodium acetate buffer (150 m$M$, pH = 4.8) containing 0.4 $M$ ammonium sulfate. They were transferred to 10 μl hanging drops on siliconized cover slips, inverted over 1 ml reservoir of precipitating solution in 24-well cell-culture plates. Typically each drop consisted of 5 μl of protein solution (32 mg ml$^{-1}$) and 5 μl of sodium acetate buffer (150 m$M$, pH = 4.8) containing 1.6 $M$ ammonium sulfate mixed with 100 m$M$ NaCl. The precipitant solution consisted of 0.8 $M$ ammonium sulfate and 50 m$M$ NaCl in 75 m$M$ sodium acetate buffer, pH 4.8. These conditions provided optimum crystal growth and eliminated uncontrolled nucleation which generally characterizes the crystallization of ⟨2aa⟩. The crystals were increasing in size for approximately 12 h, and the whole procedure, starting from the washing, was repeated several times for each crystal, until a satisfactory final size was reached.

### 2.2. *Data collection and processing*

One single (parallelepiped shaped) crystal with the dimensions of 2.3 × 1.7 × 1.2 mm was mounted in a glass capillary with the [−1,−1,0] direction parallel to the spindle axis and was used for entire X-ray data collection. The crystallographic and data-collection parameters are given in Table 1. Data extending to 1.09 Å were measured using synchrotron radiation at the EMBL outstation at DESY/Hamburg using beamline X11 at the DORIS III storage ring. The wavelength was adjusted to 0.92 Å and reflections were recorded with a MAR Research imaging-plate scanner.

Owing to extensive overload effects, three sets of data from overlapping resolution ranges were collected at room temperature on beamline X11. Since most of the reflections lower than 2.5 Å saturated the detector, a fourth data set was measured for the low-resolution data with much shorter exposures at beamline X31. The crystal was exceptionally resistant to radiation damage

and showed no significant deterioration after all four data sets were collected.

Data processing and merging was carried out using the program *DENZO* (Otwinowski, 1991). All other computations were carried out with the *CCP*4 (Collaborative Computational Project, Number 4, 1994) suite of programs. Interpretation of electron-density maps and rebuilding of the atomic model were carried out using the programs *FRODO* (Jones, 1985) and *O* (Jones *et al.*, 1991) on Evans & Sutherland and Silicon Graphics workstations.

### 2.3. *Refinement*

2.3.1. *General strategy*. The starting model for the refinement comprised residues Met1 to Phe58 from the coordinate set 1RPO deposited with the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977) which corresponds to an earlier model of ⟨2aa⟩ refined at 1.4 Å resolution (Vlassi *et al.*, 1994). No water molecules from this structure were included. At the first stages of the refinement a restrained least-squares energy minimization using the conjugate-gradient algorithm with the program *X-PLOR* (Brünger, 1990) was performed, starting with 1.4 Å data and gradually extending the resolution until all data up to 1.09 Å were included. Building of the solvent structure was aided by an automated procedure implemented in the program *ARP* (Lamzin & Wilson, 1993). Because of extensive disorder effects of the C terminus of the protein (residues 60–65), an unambiguous interpretation of the electron-density map in this region was not possible and, therefore, the six C-terminal residues were not included in the refinement. Water molecules in the vicinity of the C-terminal region were thus conservatively assigned. The atomic model at this stage was refined against structure-factor amplitudes by least-squares minimization using the *CCP*4 version of the *PROLSQ* program (Konnert & Hendrickson, 1980). H atoms were included as geometrically fixed contributors. At the end of this stage it was decided to exclude low-resolution data (<8 Å) from further refinement due to the disorder effects and the incomplete solvent structure description at the C-terminal region. Subsequently, a restrained conjugate-gradient refinement was carried out with the program *SHELXL*93 (Sheldrick, 1993) using six-parameter anisotropic temperature factors for the non-H atoms and isotropic ones for the H atoms. Finally, unrestrained refinement was performed, using the blocked full-matrix option of *SHELXL*93; this allowed us to fully exploit the information of the high-resolution data and to obtain reliable estimates of the standard deviation (e.s.d.'s) for all refined parameters. The progress of the refinement was followed by applying the free $R$ value (hereafter referred to as $R_{\text{free}}$) test (Brünger, 1992) on a randomly chosen subset of data (10% of the total) which were not included in the

refinement and were thus unbiased from the protein model. The course of the *R* factors during refinement is shown in Fig. 2.

2.3.2. *Restrained refinement using X-PLOR.* The conjugate-gradient algorithm was used for restrained least-squares energy minimization with *X-PLOR*. Bond lengths, bond angles and the planarity of aromatic rings and peptide bonds were restrained to 'target' values from the *X-PLOR* dictionary. The *X-PLOR* refinement was performed in four steps.

Step 1. 33 cycles of positional refinement were followed by 20 cycles of restrained temperature-factor refinement. Data were restricted to 1.4 Å. The value of the *R* factor fell from 29.8 to 22.4%. Peaks on difference Fourier synthesis higher than $3\sigma$ from the mean electron density were interpreted as water molecules and were added to the model manually. For three disordered residues (Lys6, Asp31 and Thr17) extra electron density was modeled as an alternative side-chain conformation.

Step 2. 34 cycles of positional refinement were followed by 20 cycles of temperature-factor refinement and the *R* factor fell to 21.7%. 14 new water molecules were added manually to the model. On the basis of a difference Fourier synthesis, alternative side-chain conformations for three disordered residues (Lys3, Met11 and Glu28) were included in the model.

Step 3. Data extending to 1.3 Å were included, and 60 cycles of positional refinement followed by 20 cycles of temperature-factor refinement were performed. The resulting *R* value was 21.9%. The side chains of two further residues (Glu41, Glu49) revealed a second conformation. More water molecules were added manually.

Step 4. Data extending to 1.2 Å were introduced and after positional and temperature-factor refinement the resulting *R* value was 22%. Several cycles of model rebuilding were performed, whereby emphasis was placed on new water molecules and disordered side chains. This stage was by far the most time consuming, so it was decided to proceed using the automated fitting refinement program *ARP* (Lamzin & Wilson, 1993).

2.3.3. *ARP/PROLSQ refinement.* Alternating cycles of automated building of the water structure and restrained refinement of the model were performed using the programs *ARP* (Lamzin & Wilson, 1993) and *PROLSQ* (Konnert & Hendrickson, 1980), respectively. Water positions corresponding to an electron density of less than $1\sigma$ above the mean density were considered incorrect and were replaced automatically by new ones corresponding to peaks in difference electron-density maps higher than $2\sigma$ from the mean, if certain distance criteria were satisfied: The distances which were considered to be acceptable for water molecules were 2.3–3.5 Å from O atoms, 2.7–3.5 Å from N atoms and 3.7 Å from any other atom. These distance criteria were also used as restraints for contacts in the *PROLSQ* refinement. The *ARP/PROLSQ* refinement was performed in three steps using all data up to 1.09 Å.

Step 5. After three cycles of positional and tempera-ture factor *ARP/PROLSQ* restrained refinement, 17 new water molecules were automatically added and an *R* value of 22.2% was reached. At the end of this stage and in order to take into account specific features of the electron-density map which had become very clear at this stage, it was decided that H atoms had to be intro-duced as fixed contributors in geometrically calculated positions.

Step 6. Only H atoms associated with backbone atoms were included, and the *R* value after refinement fell to 21.2%. 14 new water molecules were found.

Step 7. H atoms attached to side-chain atoms were introduced. In case of disordered side chains, H-atom positions were calculated only for the major conforma-tion. The *R* value fell to 20.4% and the total number of water molecules included in the model increased to 109.

After this stage no further improvement could be achieved with conventional techniques, although several electron-density features on difference Fourier synth-eses could not be accounted for by the isotropic model, being clearly associated with the anisotropic motion of the atoms, for which no option is provided by the programs used up to this stage. The refinement was thus continued with the program *SHELXL*93 which, in addition to anisotropic temperature-factor refinement, offers the advantage of estimation of standard errors for refined parameters due to the use of full-matrix least-squares techniques. Extensive disorder effects at the C-terminal region made the electron-density maps non-interpretable for residues 60–65 and dictated a conser-vative allocation of density peaks to water positions and an omission of low-resolution data from further refine-ment.

2.3.4. *SHELXL*93 *refinement. Restrained refinement.* Restrained least-squares refinement was initially performed using the conjugate-gradient option of the program. As stereochemical restraints the 'target' values given by Engh & Huber (1991) were applied. For thermal parameters the following constraints were
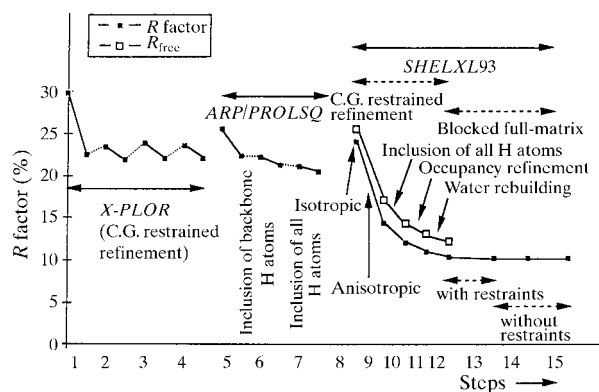


Fig. 2. Course of the *R* factors during the refinement of ⟨2aa⟩.

introduced: for a pair of bonded atoms the anisotropic displacement components along the bond were restrained to be equal within a standard deviation of 0.8 $Å^2$. In addition, for atoms closer than 1.7 Å, the directions of the thermal ellipsoids were restrained with the $U_{ij}$ ($i \neq j$) components being equal within an effective standard deviation of 0.8 Å. Water molecules were restrained within a standard deviation of 0.1 $Å^2$ to an isotropic behavior.

Step 8. Five cycles of isotropic refinement without H atoms were initially performed in order to be able to compare the *SHELXL*93 refinement with the previous stages. The resulting $R$ value was 23.9% for all reflections between 8 and 1.09 Å and the $R_{free}$ value was 25.4% for 2198 reflections. Maximum peaks of the residual density in difference Fourier syntheses after this step were reaching values of approximately $10\sigma$ above the mean, and reflected anisotropic motion effects and the presence of H atoms (Fig. 3).

Step 9. Six-parameter anisotropic temperature factors for all non-H atoms were included and after restrained refinement the $R$ value fell to 14.5% and the $R_{free}$ to 17.03%. The maximum residual density peaks in difference Fourier syntheses fell to approximately $5\sigma$ above the mean. The drop in the residual electron density in the difference Fourier map as well as the improvement of the $R$ factor and of the $R_{free}$ (Fig. 2) shows that inclusion of anisotropic motion was essential at this stage of the refinement. The remaining electron density in difference Fourier electron-density maps after this stage could be attributed to H atoms (Fig. 4).

Step 10. H atoms were introduced in geometrically calculated positions and were refined using the 'riding model' which ensures a sensible geometry during refinement (Sheldrick, 1993): the coordinates of a H atom (H) were re-idealized before each refinement cycle and 'ride' on an atom ($X$) to which the H atom is attached. The length of the $X-$H vector and its direction relative to the C$-X$ bond are kept constant but the position of $X$ may move during refinement. After five cycles of anisotropic refinement for all non-H atoms and isotropic for H atoms, the $R$ value fell to 12.1% and the $R_{free}$ to 14.3%. The highest peak in difference Fourier maps corresponded to a third conformation for the disordered Thr17 which was also incorporated in the model.

Step 11. After 15 cycles of anisotropic refinement for non-H atoms and occupancy refinement for atoms belonging to disordered side chains, the $R$ value dropped to 11.1% and the $R_{free}$ to 13%. The sum of the occupancies of the multiple conformations for each disordered side chain was constrained to 1.0.

Step 12. 100 cycles of water-structure modeling were performed. Water molecules with $U_{ij}$ values greater than 1.0 $Å^2$ (corresponding to a $B$ value of 79 $Å^2$) were rejected and peaks from a difference Fourier map calculated by *SHELXL*93 were interpreted as partially occupied water molecules and refined without van der Waals repulsion. Some partially occupied waters which were located close to disordered side chains were assumed to be associated only with a subset of the alternative side-chain conformations (Fig. 5). After all atoms were included, the relative weight between observed and calculated data was updated. The $R$ factor fell to 10.3% for data between 8 and 1.09 Å and $R_{free}$ to 12.3%.
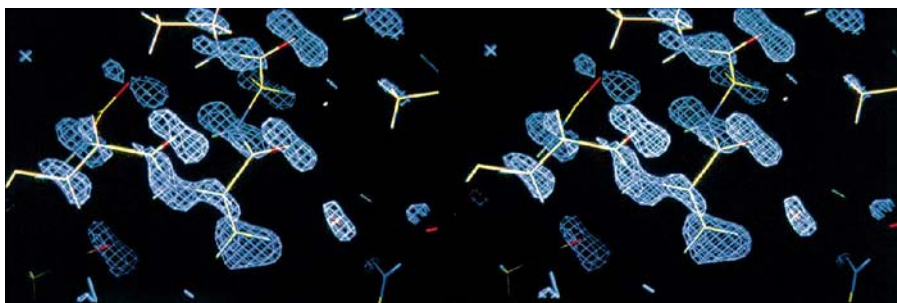


Fig. 3. Difference Fourier electron-density map (contoured at $3\sigma$ above the mean density) after completion of the isotropic refinement. The region of residue Ala47 is shown. The remaining density corresponds to anisotropic motion and H atoms.
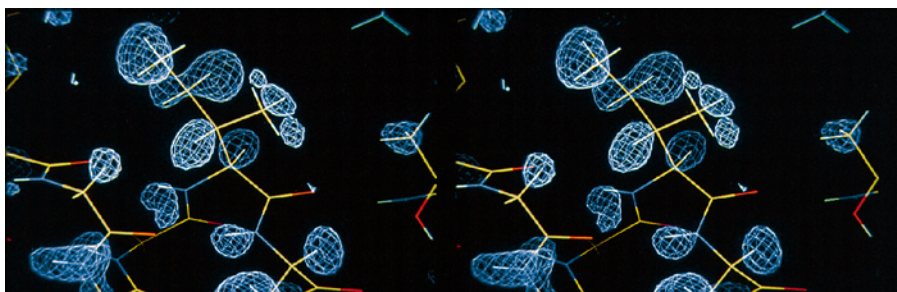


Fig. 4. Difference Fourier electron-density map (contoured at $2\sigma$ above the mean density) after completion of the six-parameter anisotropic temperature factor. The region of residue Ile15 is shown. The remaining density corresponds to H atoms (shown in white).

Table 2. *Final refinement parameters for three models*

Model *A*: restrained refinement, Model *B*: unrestrained geometry and restrained temperature-factor refinement, Model *C*: stereochemically restrained refinement, unrestrained temperature-factor refinement.

| | Model *A* | Model *B* | Model *C* |
|---|---|---|---|
| *R* factor (%) | 10.1 | 10 | 10 |
| GooF† | 1.068 | 1.047 | 1.169 |
| Restrained GooF‡ | 0.969 | 0.962 | 1.017 |
| Mean shift/e.s.d. | 0.088 | 0.089 | 0.132 |
| Max. density in the $\Delta F$ map (e Å$^{-3}$) | 0.25 | 0.25 | 0.33 |
| Mean e.s.d. of atomic positions (Å) | 0.021 | 0.022 | 0.021 |
| Mean e.s.d. of $B_{eq}$§ (Å$^2$) | 2.85 | 3.86 | 4.04 |

† Goodness of fit: GooF = $S = [\sum^{\text{reflections}} w(F_o^2 - F_c^2)^2/(n - p)]^{1/2}$, where $n$ = number of reflections, $p$ = number of parameters and $w$ is a weight depending on the standard deviation of the intensities. ‡ Restrained goodness of fit: $\sum^{\text{restraints}} w(yt - y)^2$, where $y$ is the quantity being restrained and $yt$ is its 'target' value. § $B_{eq} = 8\pi^2/3 \sum_i \sum_j U_{ij}a_i a_j a_i^* a_j^*$, where $a_i$, $a_i^*$ the unit-cell axes vectors in the direct and reciprocal space, respectively.

The conjugate-gradient algorithm used in the previous steps cannot provide estimation of standard deviations on the refined parameters. The full-matrix or blocked full-matrix least-squares refinement options of *SHELXL*93 offer the advantage of providing e.s.d.'s for all refined parameters and in additionally, a global minimum can be reached in cases where conjugate-gradient techniques can get trapped in a local one. In the case of ⟨2aa⟩ however, the full-matrix refinement option was not applicable, due to the unacceptably low observations-to-parameters ratio when anisotropic temperatures are used. Thus, a blocked full-matrix least-squares refinement was used instead. Six blocks of approximately 1500 parameters each were used, with sufficient overlap between the blocks in order to ensure that every e.s.d. is calculated with all contributing atoms in at least one of the refinement cycles.

Step 13. After 30 cycles of blocked full-matrix least-squares refinement (equivalent to five full-matrix cycles) convergence was achieved with an *R* value of 10.1%. The protein model obtained at the end of this stage will be termed as model *A*. Statistics on the refinement parameters of this model are summarized in Table 2.

*2.3.5. Unrestrained refinement.* The least-squares refinement described so far, is influenced by the restraints used as additional 'observations' so as to reach a well defined minimum. Consequently, the e.s.d.'s of the refined parameters are to some extent biased by the restraints. It should be noted however, that in *SHELXL*93 stereochemical restraints are directly applied only to bond lengths and angles and not to other parameters such as torsion angles and hydrogen-bonding geometries. In order to obtain unbiased e.s.d.'s, the stereochemical restraints and constraints on $U_{ij}$ values applied by the program were removed. 'Unrestrained' refinement was performed in two parallel steps.

Step 14. All stereochemical restraints were removed by imposing very small weights to the corresponding geometrical terms. However, the constraints on $U_{ij}$'s were kept, since otherwise an unstable refinement is expected due to a rather unfavourable ratio (4:1) of observations to parameters which approaches the limit for fully unrestrained refinement. This refinement was stable and led to a protein model (model *B*) with an *R* value of 10%. Statistics on model *B* are also given in Table 2.

Step 15. It was attempted to model anisotropic motion by $U_{ij}$ values which are unbiased from any temperature-factor constraints imposed during refinement. Constraints on thermal motion were thus removed, but stereochemical restraints were again introduced. The refinement at this stage was less stable than the previous ones. The *R* factor for the model obtained (model *C*) is 10% and the refinement statistics are given in Table 2.
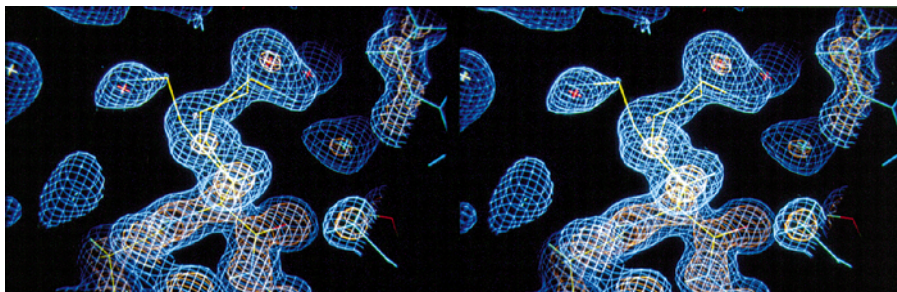


Fig. 5. $2F_o - F_c$ electron-density map showing the disordered residue Lys3 (blue: contoured at $1\sigma$ above the mean density, red: contoured at $2.5\sigma$ above the mean density). The peak on the left-hand side of the continuous electron-density distribution which has been allocated to the side chain of Lys3 is interpretable both as a water molecule and as an alternative conformation of the side chain.

## 3. Results and discussion

### 3.1. *The crystal structure*

In the following discussion, the unrestrained model *B* will be referred to as the refined structure of ⟨2aa⟩. The models resulting from the refinement at 1.09 Å resolution and from the previously reported refinement at 1.4 Å (Vlassi *et al.*, 1994) are essentially identical. The r.m.s. difference is 0.09 Å for 59 $C_\alpha$ atoms (for residues Met1–Gly59), and 0.1 Å for all backbone atoms; the biggest differences are associated with solvent-exposed residues, while the agreement in the hydrophobic core is excellent. No significant changes in the conformational angles $\varphi$, $\psi$ occur after refinement at 1.09 Å. The mean coordinate error in model *B* based on the e.s.d.'s of the atomic coordinates (obtained by inversion of the blocked full matrix) is 0.021 for the 625 non-H atoms (including water O atoms). However, all conventional constraints used in protein refinement programs overestimate the mean coordinate error: for example, the r.m.s. deviation of 0.06 Å for bond distances obtained from the restraints of *SHELXL*93 corresponds to an r.m.s. coordinate error of 0.04 Å. Furthermore, the plot of the final *R* factor against resolution (Luzzati, 1952) estimates the r.m.s. coordinate error to 0.05–0.1 Å. On the other hand, the *SIGMAA* plot defined by Read (1986) gives an estimate of the coordinate error of 0.026 Å, which is remarkably close to the value obtained from the blocked full-matrix approach.

### 3.2. *Disordered residues*

Nine residues with disordered side chains had been found in the 1.4 Å model (Vlassi *et al.*, 1994). From those residues only five (Lys3, Lys6, Met11, Ser17 and Glu49) show unambiguous density for alternative side-chain conformations in the 1.09 Å resolution structure. Interestingly, for the other four residues (Ile15, Thr21, Lys25 and Ser42) only one conformation was identified at high resolution. This can be explained for Thr21, Lys25 and Ser42 by the non-inclusion of the anisotropic motion which was detectable even with the 1.4 Å data, leading to a residual electron density; this density was erroneously interpreted as an extra conformation. Residue Ile15 is tightly packed in the hydrophobic part of the protein. This constrains severely its freedom to adopt alternative side-chain conformations. At 1.09 Å resolution, Ile15 exhibits a nearly isotropic motion and unambiguously a unique side-chain conformation. The electron density for this residue after correction for anisotropic motion effects and inclusion of H atoms, is shown in Fig. 6. The residual electron density which was interpreted at 1.4 Å as a second conformation, corresponds to the H atoms attached to the $C_{\gamma1}$ atom of the side chain (Fig. 7) and were not included in the 1.4 Å structure. This is an indication that H atoms may contribute to X-ray intensities even at 1.4 Å.

The high-resolution structure reveals additional disorder effects for five side chains: Asn27, Glu28, Asp31, Glu41 and His46. Indication of a possible disorder was already present at 1.4 Å due to broad and low electron density for all atoms. The improved quality of the atomic resolution electron-density map, revealed alternative conformations of the above side chains. However, because the density associated with the alternative conformations was weak, only partial fitting was possible.

The disorder residues are 17% of the total number in the model and are exposed to the solvent. The disordered residues form clusters on the surface of the bundle. If the specific pattern of hydrophilic–hydrophobic residues is described in terms of heptads (Vlassi *et al.*, 1994) the disorder affects mainly positions *b*, *c* and *f* which are the most exposed ones in the topology of the bundle. Interestingly, there is an asymmetry in disorder effects: several of the semi-buried positions of type *g* are disordered (*e.g.* Met11, Fig. 8), none of the topologically equivalent *e* positions, however. Finally, the six C-terminal residues are extremely disordered and were not visible even in the atomic resolution map.

### 3.3. *Solvent structure*

Using the $V_m$ value (Matthews, 1966) of 1.9 $\text{Å}^3\,\text{Da}^{-1}$ (Kokkinidis *et al.*, 1993) for the ⟨2aa⟩ crystals, a solvent content of 35% corresponding to 161 water molecules is calculated. The final model contains only 65 fully occupied water molecules and 49 partially occupied. Modelling of water molecules in the disordered regions was performed conservatively, and this explains the
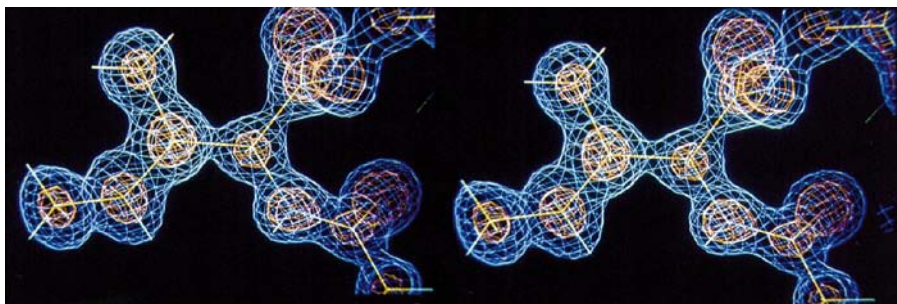


Fig. 6. $2F_o - F_c$ electron-density map (blue: contoured at $2\sigma$ above the mean density, red: contoured at $4.5\sigma$ above the mean density) for Ile15 showing that the residue is well ordered.

Table 3. *Weighted mean values and (in parentheses) weighted standard deviations of the mean (see Appendix) for various classes of torsion angles in the unrestrained model (model B)*

| | | | | | |
|---|---|---|---|---|---|
| $\omega_1$ ($C_\alpha{-}C{-}N{-}C_\alpha'$) (°) | +178.32 (0.34) | | | | |
| $\omega_3$ ($O{-}C{-}N{-}C_\alpha$) (°) | | | | | |
|   For all peptide bonds | −1.84 (0.39) | | | | |
|   Helical part only | −1.92 (0.41) | | | | |
| $\chi_1$ ($N{-}C_\alpha{-}C_\beta{-}C_\gamma$) (°) | | | | | |
|   g − | +65.18 (3.08) | | | | |
|   g + | −67.22 (1.19) | | | | |
|   t | +179.84 (1.22) | | | | |
| Chirality ($C_\alpha{-}N{-}C_\beta{-}C$) (°) | −2.41 (0.06) | | | | |
| $\alpha$-Helices | $\varphi_i$ (°) | $\psi_i$ (°) | | | |
| | −62.57 (0.48), | −42.37 (0.63) | | | |
| | $\varphi_{i+1}$ (°) | $\psi_i$ (°) | | | |
|   'Hydrophobic' environment† | −62.32 (0.68) | −42.60 (1.01) | | | |
|   'Hydrophilic' environment† | −63.02 (0.70) | −41.98 (0.98) | | | |
| $\alpha$-Helical hydrogen-bond geometry | O$\cdots$N (Å) | CO$\cdots$N (°) | CO$\cdots$H (°) | O$\cdots$H (Å) | NH$\cdots$O (°) |
|   'Hydrophobic' environment | 2.92 (01) | 156.2 (8) | 161.4 (8) | 2.09 (02) | 161.4 (8) |
|   'Hydrophilic' environment | 3.04 (03) | 153.2 (6) | 159.8 (6) | 2.20 (03) | 160.1 (1.2) |

† As defined by Blundell *et al.* (1983).

discrepancy between the expected and the actually identified water positions. The majority of the partially occupied waters that are located close to disordered side chains occur only in connection with one of the alternative side-chain conformations. Part of the solvent structure is organized in pentagonal or hexagonal rings which bridge neighboring protein molecules. Details of the solvent structure will be published elsewhere.

Comparison of the solvent structure of the high-resolution model and the one obtained at 1.4 Å (Vlassi *et al.*, 1994) shows that 26 out of the 65 fully occupied water molecules (40% of the well ordered water structure) are located at very similar positions (deviations less than 1.5 Å) and exhibit identical hydrogen-bonding partners. Almost half of those waters (13 out of 26) are common with wild-type ROP (Banner *et al.*, 1987) and represent 20% of the ordered solvent structure. These well conserved water molecules are located in two regions: one region includes a network of waters located in the highly hydrated region around residue Thr17. The second region contains water molecules involved in intermolecular contacts. One buried water molecule is conserved in both wild-type ROP and the ⟨2aa⟩ structure.
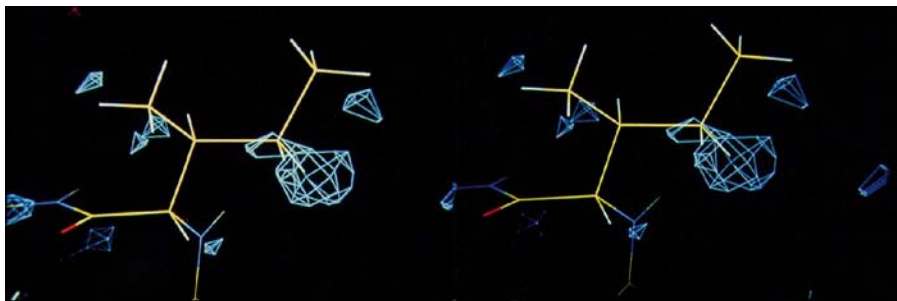


Fig. 7. Difference Fourier electron-density map (contoured at $2\sigma$ above the mean) for Ile15 calculated with the old data set at 1.4 Å (Vlassi *et al.*, 1994). The remaining density which was initially interpreted as a second side-chain conformation (Vlassi *et al.*, 1994), corresponds to the H atoms attached to $C_{\gamma1}$.
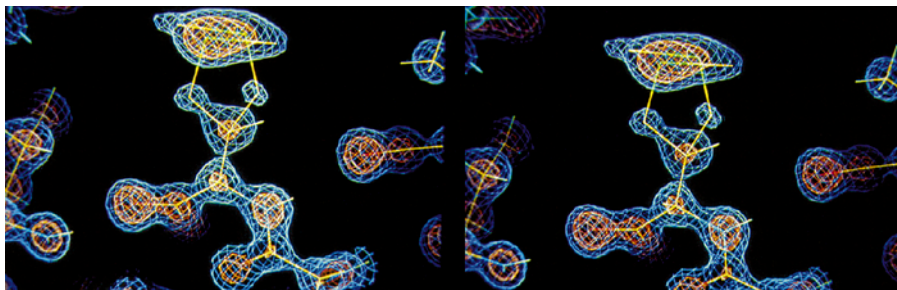


Fig. 8. $2F_o - F_c$ electron-density map (blue: contoured at $2\sigma$ above the mean density, red: contoured at $4.5\sigma$ above the mean density) in the region of Met11 showing two alternative side-chain conformations for this residue.

Table 4. *Weighted mean values and (in parentheses) weighted standard deviations of the mean (see Appendix) for various classes of bond lengths and angles in the unrestrained model (model B) distributions*

Standard deviations of the distributions are $(N-1)^{1/2}$ times the standard deviations of the mean given in parentheses. The number of contributors ($N$) is given in square brackets. Superscripts refer to the $\chi^2$ test for the distributions obtained: $A$ corresponds to an accepted $\chi^2$ test showing that the distribution is normal at a level better than 5%. $A1$: accepted $\chi^2$ test at a level of 1%. $R$: distribution is unlikely to be normal. The function $t$ (see *Appendix*) is used to evaluate the significance of the difference between the stereochemistry of the model and stereochemical parameters widely used in protein refinement. For parameters which have passed the $\chi^2$ test, significant differences (at a confidence level of 99%) on the basis of a $t$-test between the mean values from model $B$ and the values reported by Engh & Huber (1991), Laskowski *et al.* (1993), or the restraints used in the *X-PLOR* (Brünger, 1990) and the *TNT* (Tronrud *et al.*, 1987) are given in bold.

| | | $\langle 2aa\rangle$ | EH‡ | $t$§ | X-PLOR library | $t$§ | TNT† library | $t$§ | 186 'best'¶ | $t$§ |
|---|---|---|---|---|---|---|---|---|---|---|
| Bond lengths | | | | | | | | | | |
| C—N | [58] | 1.327 (002)$^A$ | 1.329 | 1 | 1.33 | 1.5 | 1.33 | 1.5 | 1.323 | 2 |
| C—O | [58] | 1.237 (002)$^A$ | 1.231 | **3** | 1.23 | **3.5** | 1.23 | **3.5** | 1.240 | 1.5 |
| $C_\alpha$—C | [58] | 1.508 (003)$^A$ | 1.525 | **6** | 1.52 | **4** | 1.52 | **4** | 1.525 | **6** |
| $C_\alpha$—$C_\beta$ (Ala) | | 1.512 (009) | 1.521 | 1 | 1.52 | 0.9 | 1.52 | 0.9 | 1.530 | 2 |
| (Thr, Ile) | | 1.513 (008) | 1.540 | 3.5 | 1.53 | 2 | 1.52 | 0.9 | 1.548 | 4.5 |
| (the rest) | [52] | 1.516 (004)$^R$ | 1.530 | 3.5 | 1.52 | 1 | 1.54 | 6 | 1.533 | 4.2 |
| N—C$\alpha$ | [58] | 1.458 (002)$^{A1}$ | 1.458 | — | 1.45 | **4** | 1.46 | 1 | 1.466 | **4** |
| Bond angles | | | | | | | | | | |
| C—N—$C_\alpha$ | [58] | 121.37 (16)$^A$ | 121.7 | 2 | 120.0 | **8.5** | 121.9 | **3.3** | 121.69 | 2 |
| $C_\alpha$—C—N | [58] | 118.16 (15)$^R$ | 116.2 | 13 | 117.5 | 4.4 | 115.6 | 17 | 116.31 | 12 |
| $C_\alpha$—C—O | [58] | 120.47 (17)$^A$ | 120.8 | 2 | 121.5 | **6** | 121.1 | **3.7** | 120.06 | 2.4 |
| $C_\beta$—$C_\alpha$—C(Ala) | | 111.36 (56) | 110.5 | 1.5 | 106.5 | 8.7 | 111 | 0.6 | 110.48 | 1.6 |
| (Thr, Ile) | | 111.95 (57) | 109.1 | 5 | 110.0 | 3.4 | 111 | 1.7 | 111.33 | 1 |
| (the rest) | [52] | 111.66 (21)$^R$ | 110.1 | 7 | 109.5 | 10 | 111 | 3 | 110.33 | 6.3 |
| N—$C_\alpha$—C | [58] | 111.17 (17)$^A$ | 111.2 | 0.16 | 111.6 | 2.5 | 112 | **5** | 110.77 | 2.4 |
| N—$C_\alpha$—$C_\beta$ (Ala) | | 110.20 (31) | 110.4 | 0.64 | 108.5 | 5.5 | 110.9 | 2.3 | 110.52 | 0.96 |
| (Thr, Ile) | | 110.91 (53) | 111.5 | 1.1 | 110.0 | 1.7 | 112 | 2.1 | 111.05 | 1.5 |
| (the rest) | [52] | 109.93 (26)$^R$ | 110.5 | 2 | 110.0 | 0.3 | 112 | 8 | 110.61 | 2.6 |
| O—C—N | [58] | 121.35 (15)$^A$ | 123.0 | **11** | 121.0 | 2.3 | 123.2 | **12** | 123.40 | **13** |

‡ Values for the parameters given by Engh & Huber (1991). § Value of function $t$ (see *Appendix*) for a comparison between the $\langle 2aa\rangle$ parameters and the library listed in the preceding column. † Values used in the refinement program *TNT* (Tronrud *et al.*, 1987). ¶ 186 best refined structures of the Brookhaven Protein Data Bank (Laskowski *et al.*, 1993).

### 3.4. *Stereochemical analysis of the model*

The geometrical parameters of $\langle 2aa\rangle$ are summarized in Tables 3 and 4. The mean values and standard deviations in these Tables are the weighted mean and weighted standard deviations, respectively (see *Appendix*).

3.4.1. *Dihedral angles. Peptide planarity ($\omega_1$, $\omega_3$).* The average $\omega_1$ ($C_\alpha$—C—N—C'$\alpha$) is $+178.32°$ $(0.34°)$ and deviates significantly from the value of $180°$ used so far as 'target' value in restrained refinements of proteins. This was also observed by other investigators (Stec *et al.*, 1995; Sevcik *et al.*, 1996) and may reflect a general property of proteins. The sign of $\omega_1$ for the $\alpha$-helical residues is positive, in agreement with observed preferences in $\alpha$-helical (Remington *et al.*, 1982) and $\beta$-sheet-containing structures (Marquart *et al.*, 1983).

The average $\omega_3$ (O—C—N—$C_\alpha$) deviates significantly from the ideal planarity ($\omega_3 = 0°$). For the 50 $\alpha$-helical residues of the model, the average $\omega_3$ is 1.92 $(0.41°)$, a value which is closer to the average value reported for beta structures (Marquart *et al.*, 1983) and differs from the value given for the all-$\alpha$ protein, citrate synthase (Remington *et al.*, 1982). This observation contradicts

the suggestion by Marquart *et al.* (1983) that the average $\omega_3$ depends on the helical content of a protein.

*The $\chi_1$ angle.* The $\chi_1$ angles in our model cluster around the average values of $+65.18°$ $(3.08°)$, $-67.22°$ $(1.19°)$ and $+179.84$ $(1.22°)$ for the $g^-$, $g^+$ and t conformers (Janin *et al.*, 1978), respectively. Alternative conformations in disordered side chains were also included in the above statistics. The $\chi_1$ angles for the $\alpha$-helical residues show almost the same preference for both t (53.9%) and $g^+$ (43%) conformers and only 3.8% for the $g^-$. The frequency ratio g/t for the helical residues does not agree with the value of 2.04, found for helices in other high-resolution structures (Teeter *et al.*, 1993) or in the more general analysis by Janin *et al.* (1978). This observation contradicts the hypothesis by Teeter *et al.* (1993) about the establishment of specific patterns in side-chain conformations for successive residues in $\alpha$-helical structures. In the case of $\langle 2aa\rangle$ the less common $g^-$ conformation is found only for Thr19 and Asp34. Notably, both residues are located in local distortions of the helical structure: The side chain of Thr19 is hydrogen bonded in a bifurcated fashion to C=O groups of the preceding turn of the helix (Ile15 and Arg16), which may have a destabilizing effect on the

helix (Richardson & Richardson, 1989). Asp34 starts the second helix of the monomer, with its side chain imitating an $\alpha$-helical hydrogen bond.

$\varphi, \psi$ *angles in $\alpha$-helices.* The mean $\varphi, \psi$ angles for the helical part (50 residues) of the molecule are similar to those reported for other high-resolution protein structures (Glover *et al.*, 1983; Barlow & Thornton, 1988; Teeter *et al.*, 1993). In this geometry of the helices, the C=O vector turns away from the direction of the helix axis, and the average hydrogen-bond angle NH$\cdots$O is 160.3° (0.7°). This slightly distorted hydrogen-bonding geometry is more pronounced in the case of residues exposing their carbonyl O atoms to solvent [or 'hydrophilic' environment as defined in Blundell *et al.* (1983)]. These carbonyl O atoms usually form additional hydrogen bonds to water molecules or to side chains. In the $\langle 2aa \rangle$ structure, the N$\cdots$O distances in the 'hydrophilic' environment are longer (Table 3) compared with the less distorted $\alpha$-helical hydrogen bonds in the 'hydrophobic' environment of the protein core. In addition the NH$\cdots$O angles show a stronger deviation from 180° (Table 3). The difference of the peptide angles $\varphi_{i+1}, \psi_i$ for 'hydrophobic' and 'hydrophilic' residues is not statistically significant on the basis of a *t*-test. There is thus no grouping of the values of the $\varphi_{i+1}, \psi_i$ pairs, which are associated with the direction of the peptide bond, in contrast to the grouping observed in several cases by Blundell *et al.* (1983).

3.4.2. *Bond lengths, angles.* Statistics on bond lengths and angles for the polypeptide backbone and the $C_\beta$ atoms are given in Table 4. Due to the high $\alpha$-helical content of $\langle 2aa \rangle$, it is possible that the values in Table 4 are typical for predominantly $\alpha$-helical proteins. The very few non-helical residues of the protein are insufficient for a sensible analysis. We have examined the
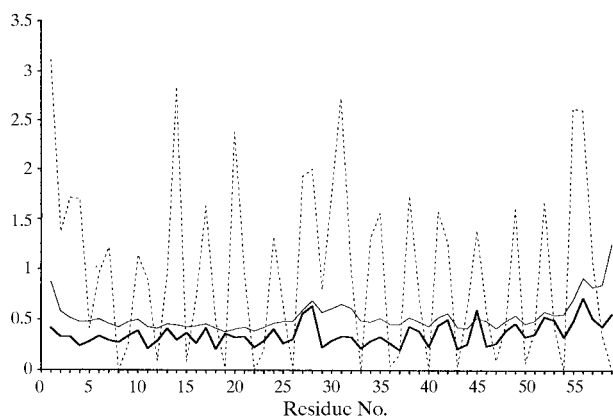
Table 5. *Anisotropy measures for various types of atoms in $\langle 2aa \rangle$ (X=C, O or N from a side chain)*

|  | A1† | A2† | $\langle \Delta r^2 \rangle$‡ (Å²) |
|---|---|---|---|
| All atoms | 0.39 | 0.11 | 0.78 |
| Backbone | 0.37 | 0.12 | 0.52 |
| Side chains | 0.42 | 0.11 | 1.02 |
| N (backbone) | 0.36 | 0.12 | 0.49 |
| C (backbone) | 0.37 | 0.12 | 0.50 |
| =O | 0.40 | 0.13 | 0.56 |
| $C_\alpha$ | 0.36 | 0.11 | 0.50 |
| $C_\beta$ | 0.37 | 0.10 | 0.56 |
| $X_\gamma$ | 0.42 | 0.10 | 0.64 |
| $X_\delta$ | 0.40 | 0.11 | 0.78 |
| $X_\varepsilon$ | 0.43 | 0.11 | 0.94 |
| $X_\zeta$ | 0.45 | 0.11 | 1.01 |
| $X_\eta$ | 0.45 | 0.14 | 1.23 |

† Anisotropy measure (see text).　‡ $\langle \Delta r^2 \rangle = \langle U_X^2 \rangle + \langle U_Y^2 \rangle + \langle U_Z^2 \rangle$ where $\langle U_X^2 \rangle, \langle U_Y^2 \rangle, \langle U_Z^2 \rangle$ are the mean-square displacements along the vibrational axes $X, Y, Z$ as calculated by *SHELXL*93.

distributions of bond lengths and angles using the $\chi^2$ test. The majority of various types of bond lengths and angles follow the normal distribution. A statistical validation using the *t*-test of the differences between the mean bond lengths and angles obtained for the $\langle 2aa \rangle$ structure and values reported by Engh & Huber (1991), Laskowski *et al.* (1993), or the values used as stereochemistry constraints in the *X-PLOR* (Brünger, 1990) and the *TNT* (Tronrud *et al.*, 1987) dictionaries is also given in Table 4. For several types of bond lengths and angles statistically significant differences have been found. A similar result has also been reported by other investigators (Laskowksi *et al.*, 1993) on the basis of an analysis of a large database of well refined protein structures. Thus, it may be concluded that the libraries of stereochemical constraints used in protein refinement are not completely valid and should be updated with new parameters derived from protein structures refined at atomic resolution. Such updated libraries would be even more crucial for molecular dynamics simulations. Alternatively, the expected variations of the geometric parameters could be introduced as a force constant in programs that use energy terms, or as an additional weighting scheme in protein refinement programs instead of the uniform scheme used so far. A similar suggestion was made by Engh & Huber (1991).

3.4.3. *Anisotropic motion.* Model *C*, refined without thermal motion constraints, was initially used to analyse the anisotropic atomic motion in $\langle 2aa \rangle$. Despite poor refinement statistics (Table 2), the unrestrained anisotropic temperature factors of model *C* follows the same general trends as in models *A* and *B*. The restrained model *A* was thus used in the analysis of anisotropic motion.

Anisotropic temperature-factor refinement reduced the *R* value by 9.4% (Fig. 2). Thus, it would appear that anisotropic treatment of atoms in the refinement of high-resolution structures of macromolecules is crucial



Fig. 9. Variation of (*a*) the relative accessibility (broken line) (*b*) the mean-square displacement (Å²) (thin line-see text), (*c*) the anisotropy measure *A1* (bold line, see text) along the sequence of $\langle 2aa \rangle$. The relative accessibility of a residue is defined as the fraction of its solvent accessibility as calculated with *DSSP* (Kabsch & Sander, 1983) to its average accessibility in proteins (Lesser & Rose, 1990).

for obtaining a good agreement between the observed and the calculated structure factors. Backbone atoms have on the average smaller fluctuations (defined by the quantity $\langle \Delta r^2 \rangle$, see Table 5) than side-chain atoms. The fluctuations of the side-chain atoms increase with the distance from the main chain. Fig. 9 shows the magnitude of the average of fluctuations for backbone atoms along the sequence. The $\alpha$-helices have, on average, smaller fluctuations than the non-helical parts of the molecule: the mean value of the quantity $\langle \Delta r^2 \rangle$ for the backbone atoms of helical residues is 0.487 and 0.568 Å$^2$ for the bend-region residues (Ala30–Asp34). Minimum values of $\langle \Delta r^2 \rangle$ occur in the tightly packed hydrophobic core of the protein (Fig. 9). The mean fluctuation for backbone atoms of helix I (Lys3–Glu28) is 0.452 Å$^2$ which is significantly lower than the value of 0.528 Å$^2$ obtained for helix II (Gln35–Phe58). These differences

in mobility could be explained by differences in crystal contacts made by the two helices. The mean-square displacement of all atoms along the major component $X$ of their thermal fluctuation, is 2.5 times that along the direction of the smallest displacement; atomic vibrations in $\langle 2aa \rangle$ are thus clearly anisotropic. A better measure of anisotropy is given by the quantity

$$A1 = \left( \frac{2\langle U_X^2 \rangle}{\langle U_Y^2 \rangle + \langle U_Z^2 \rangle} \right)^{1/2} - 1,$$

(Brooks *et al.*, 1988) which determines the extent by which the fluctuation along the principal direction $X$ deviates from isotropy (for which $A1 = 0$). A second measure of anisotropy is given by the quantity
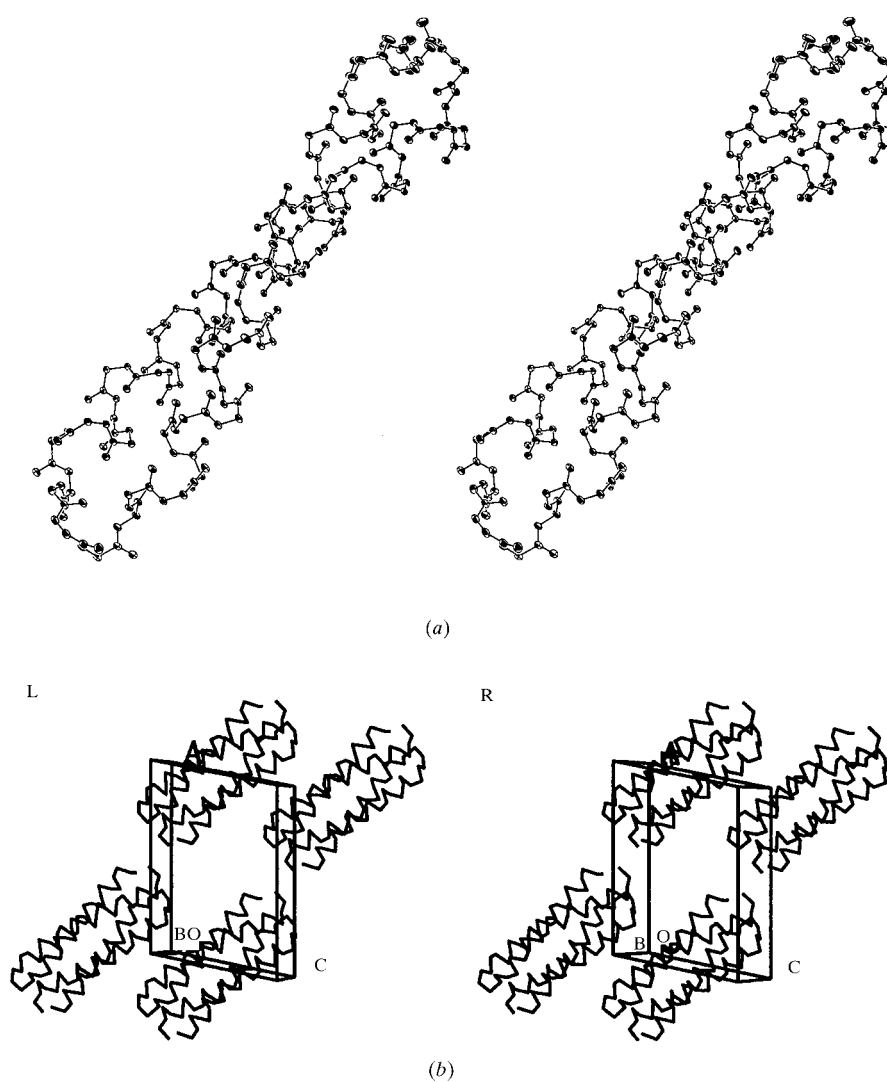


(a)



(b)

Fig. 10. (a) Thermal ellipsoids (calculated at a 99% probability level) for the backbone atoms. (b) Crystal packing diagram for $\langle 2aa \rangle$. Same orientation as in (a). The more pronounced character of atomic motion along one direction (a), is due to crystal packing effects.

Table 6. *Tensor elements, eigenvalues and eigenvectors of the **T** and **L** tensors for the rigid-body model*

The elements of tensor **L** are given in rad$^2$ and those of **T** in Å$^2$.

| | | | | Eigenvectors | | | Eigenvalues |
|---|---|---|---|---|---|---|---|
| | | | | $V_1$ | $V_2$ | $V_3$ | $\lambda_1, \lambda_2, \lambda_3$ |
| | | 0.0001 | 0.0001 | 0.0002 | 0.9405 | 0.2087 | 0.2682 | 0.0000 |
| **L** | | 0.0001 | 0.0003 | 0.0001 | −0.2682 | 0.9405 | 0.2087 | 0.0003 |
| | | 0.0002 | 0.0001 | 0.0008 | −0.2087 | −0.2682 | 0.9405 | 0.0009 |
| | | 0.1600 | −0.0010 | −0.0061 | 0.1938 | −0.6499 | 0.7349 | 0.1345 |
| **T** | | −0.001 | 0.1472 | −0.0123 | 0.6906 | 0.6224 | 0.3682 | 0.1569 |
| | | −0.0061 | −0.0123 | 0.1484 | 0.6967 | 0.4362 | −0.5695 | 0.1642 |
| | | −0.0008 | −0.0030 | −0.0006 | | | | |
| **S** | | −0.0008 | 0.0002 | 0.0000 | | | | |
| | | −0.0020 | −0.0031 | 0.0006 | | | | |

$$A2 = \left( \frac{2\langle U_Y^2 \rangle}{\langle U_Y^2 \rangle + \langle U_Z^2 \rangle} \right)^{1/2} - 1,$$

(Brooks *et al.*, 1988), which determines the anisotropy of motion in the principal plane $Y$–$Z$. The anisotropy measure $A1$ along the sequence is shown in Fig. 9 and indicates that the motion of hydrophobic core residues which are inaccessible to solvent is generally more isotropic compared to exposed residues. Maximum anisotropy occurs at the C-termini of the helices. Interestingly, the non-helical residues at the N-terminus and the bend region are almost isotropic. This probably reflects the involvement of these regions in crystal packing interactions which severely constrain the directional degrees of freedom of atomic fluctuations.

The anisotropy measures $A1$ and $A2$ for various atom types in the crystal structure of $\langle 2aa \rangle$ (Table 5) are, in general, lower than those derived from the molecular dynamics treatment of myoglobin (Kuriyan *et al.*, 1986) and lysozyme (Ichiye & Karplus, 1987). There are, however, common trends in these three cases. For side
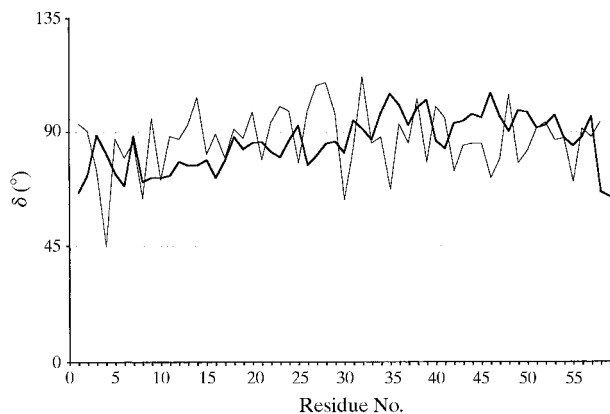


Fig. 11. Variation along the sequence of the parameter $\delta$ defined as the angle formed between the direction of the axis of the 4-$\alpha$-helical bundle and the direction of major vibrational axis of a thermal ellipsoid. Bold line: average value of $\delta$ for the backbone atoms of each residue. Thin line: average value of $\delta$ for the side-chain atoms of each residue.

chains the anisotropy is lower at the $C_\beta$ atoms and increases along the side chains. The carbonyl O atoms exhibit more pronounced anisotropic effects compared with the rest of the backbone atoms, with the direction of the largest displacement being perpendicular to the C=O bond. The value of $A2$ remains uniformly low for all classes of atoms. This indicates that anisotropic character of thermal motion is generally more pronounced along one direction (that of the largest displacement) and that a more isotropic behavior is established for the other two directions. Fig. 10(*a*) shows the directional preference of the thermal ellipsoids. For most atoms of the protein, and especially for the backbone atoms, this preferred direction is dictated by crystal packing (Fig. 10*b*) effects. The atomic thermal motion is more pronounced along a direction which is nearly perpendicular to the axis of the 4-$\alpha$-helix bundle (Fig. 11), although the motion of solvent-exposed side-chain atoms is less constrained and correlated to the atomic motions of neighboring residues. Such an example is the side chain of Arg57, which follows the motion of the bend residues to which it is hydrogen bonded (Fig. 12). In addition, atomic motions of phenylalanine and tryosine rings, as well as of imidazole rings show directional preferences depending on the environment of the ring. Finally there is a correlation, both in amplitude and direction, of the motions of atoms which are involved in crystal contacts; this represents an additional example of long-range effects in coordinated atomic motions which were also observed in crambin (Teeter *et al.*, 1993).

3.4.4. *TLS analysis.* The correlation between the direction and magnitudes of the atomic motion found in the structure of $\langle 2aa \rangle$ was validated by an interpretation of the anisotropic refinement using the full TLS rigid-body analysis (Schomaker & Trueblood, 1968) based on anisotropic temperature factors. The full TLS method analyses the rigid-body motion in terms of three tensors based on the anisotropic vibrational parameters $U_{ij}$ of all atoms: a symmetric tensor **T**, which describes the translational motion, a symmetric tensor **L**, which describes the librational motion and an asymmetric tensor **S**, which describes the correlation of librational and translational axes.

In the TLS analysis the whole 4-$\alpha$-helical bundle was treated as a rigid body. The observational equations [equation (7) of Schomaker & Trueblood (1968)] were formed for the cartesian components of the anisotropic temperature factor $U$ for all 510 non-H atoms of the protein. The values of the elements of the **T**, **L** and **S** tensors (21 independent parameters) were obtained from a linear least-squares fit to the observed individual anisotropic $U$'s (Table 6). The goodness of fit $\{[\sum(\Delta U_{ij})^2/n - s]^{1/2}$, where $n$ is the number of observations and $s$ the number of parameters} between the observed $U_{ij}$'s and those calculated from the TLS rigid-body approximation is 0.0065 Å$^2$ which indicates a relatively good agreement (Fig. 13). The mean difference on $B$ values is 3.9 Å$^2$ corresponding to 18% of the mean $B$ value of the molecule. The differences are more pronounced for disordered residues exposed to the solvent which indicates that these residues may perform an additional motion to that modelled by the rigid-body approximation, as it was also pointed out by Stec *et al.* (1995).

The relatively good agreement between the individual anisotropic model and the TLS model justifies an analysis in terms of rigid-body motion. The magnitudes and directions of the principal rotational and translational axes can be determined from the Eigenvalues and Eigenvectors of **L** and **T** tensors, respectively (Table 6). The magnitude of the libration varies between 0 and 9 × $10^{-4}$ rad$^2$ with the direction of the major librational axis given by $V_3$ of **L** (Table 6) and the mean-square translation of the molecule being equal to 0.15 Å$^2$. The major translational axis, with direction given by $V_3$ of **T** (Table 6), is almost perpendicular to the axis of the bundle. This result agrees with the directional preference of the major individual vibrational axes of all atoms described in Fig. 11. The libration and translation amplitudes are fully comparable with those derived for lysozyme (Sternberg *et al.*, 1979) and crambin (Stec *et al.*, 1995).

## 4. Conclusions

The crystallographic refinement of the ROP mutant $\langle 2aa \rangle$ represents one of the rare studies of proteins at a resolution which is sufficient to provide information on the individual atomic features of the crystal structure. These studies clearly bridge the division between the approaches of small-molecule and macromolecular crystallography and provide a more reliable and comprehensive knowledge of the stereochemistry of proteins. The specific refinement strategy implemented for $\langle 2aa \rangle$ at a resolution of 1.09 Å, comprised two major steps. The structure was initially refined with stereo-chemical restraints which were subsequently removed in order to obtain an unbiased model. The essential details introduced to the protein model include H atoms as fixed contributors, multiple conformations of side chains and anisotropic treatment of temperature factors. The largest part of the solvent structure was modelled using an automated procedure Blocked full-matrix refinement was then used and allowed a reliable estimation of the standard errors of the individual atomic parameters to be obtained.

The main lessons that can be learnt from this refinement are as follows. (i) Modelling disorder becomes
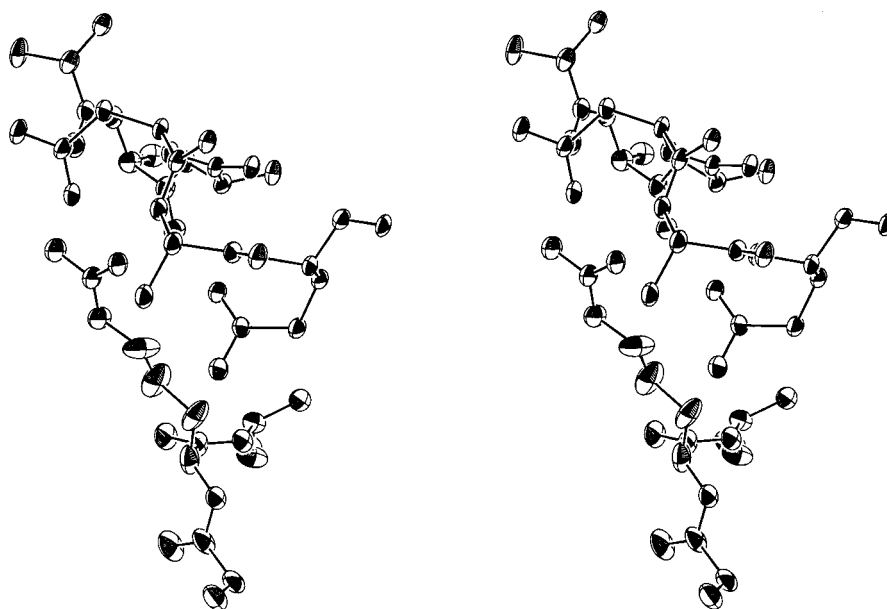


Fig. 12. Thermal ellipsoids calculated at 20% probability level for residue Arg57 and its neighbors. The side-chain atoms of Arg57 follow after C$_\gamma$ the motion of the bend region residues to which they are hydrogen bonded.

increasingly important for structures refined at an atomic resolution. This observation was also made for crambin (Stec *et al.*, 1995), rubredoxin (Dauter *et al.*, 1992) and ribonuclease (Sevcik *et al.*, 1996). There are ten residues with disordered side chains in ⟨2aa⟩, and these form clusters on the surface of the protein. In five cases there was clear evidence for two alternative conformations and these were introduced to the model. On the other hand, the inclusion of H atoms and anisotropic motion gives an unambiguous interpretation to specific features in electron-density maps, which were explained as disorder effects when the crystal structure was solved at a resolution of 1.4 Å (Vlassi *et al.*, 1994). The most serious disorder effects in ⟨2aa⟩ occur at the C terminus. Even at an atomic resolution a modelling of this disorder is not possible, and this dictated an omission of the seven C-terminal residues from the refinement. As various electron-density features in the C-terminal region could not be clearly interpreted, fitting of the local solvent structure was carried out conservatively and low resolution X-ray data had to be excluded from the refinement.

(ii) Refinement of anisotropic motion played a very significant role in achieving convergence and lowering the *R* value. Atomic vibrations are clearly anisotropic, with the major component of the mean-square displacement being on the average 2.5 times larger than the smallest one. The thermal ellipsoids reveal a clear directional preference of atomic fluctuations which is more pronounced for the backbone atoms, with the major vibrational axes being almost perpendicular to the axis of the 4-$\alpha$-helical bundle. This directionality of the anisotropy is dictated from the crystal packing. The anisotropy of temperature factors is, in general, lower than that obtained from molecular dynamics simulations (Kuriyan *et al.*, 1986; Ichiye & Karplus, 1987) and increases with the distance from the backbone. Maximum anisotropy of backbone atoms occurs at the

carbonyl O atoms with the major displacement being perpendicular to the C=O bond. Generally, hydrophobic core residues move isotropically and with correlated amplitudes.

(iii) Analysis of the temperature factors with the TLS method (Schomaker & Trueblood, 1968) showed a relatively good agreement between the individual atomic motion and a rigid-body motion.

(iv) A realistic estimate of positional errors has been obtained from the inversion of blocked full matrix. The errors of the refined atomic positions are generally overestimated if the structure is evaluated using conventional restraint libraries. In addition the unbiased stereochemical parameters differ significantly from the values used so far as restraints in protein refinement. Therefore, new values for stereochemical parameters derived from atomic resolution protein structures should be adopted as restraints, a suggestion also made in earlier studies (Engh & Huber, 1991; Laskowski *et al.*, 1993). Alternatively, weighting schemes based on the expected standard deviations of the stereochemical parameters could replace the uniform weighting used so far in refinement.

Luzzati plots also overestimate positional errors in ⟨2aa⟩; the accuracy of this plot however is questionable for two reasons. If the molecule moves as a rigid body, as suggested by the TLS analysis, then as already pointed out in a previous work (Stec *et al.*, 1995), the crucial assumption of the random errors of the Luzzati theory is no longer valid because *B* factors increase with distance from the center of the rigid body and different shells of atoms do not contribute equally to the *R* factor as a function of resolution. In fact, we have observed that errors in atomic positions in ⟨2aa⟩ are correlated with the distance from the center of the rigid body (correlation coefficient = 49%). Furthermore, the *R* values do not realistically reflect the accuracy of the structure, since the atomic model is incomplete due to missing C-terminal residues and the conservatively interpreted solvent structure. The *SIGMAA* plot (Read, 1986) which allows for incomplete atomic models, provides a much better agreement with the estimation of errors from blocked full-matrix refinement.

(v) As also observed by other authors (Stec *et al.*, 1995; Sevcik *et al.*, 1996) the $\omega$ angle has an average value of 178° rather than 180°, which may reflect a general property of protein structure.

The considerations concerning the validity of stereochemical parameters for proteins are particularly important both for the refinement of low to medium resolution protein structures and for the evaluation of very high resolution protein models, as was also pointed out by Wilson *et al.* (1998). The advantage of analyzing and refining proteins at atomic resolution is, therefore, self evident. These studies have been made possible by recent technological advances in the fields of synchrotron radiation, area detectors and cryogenic techniques.
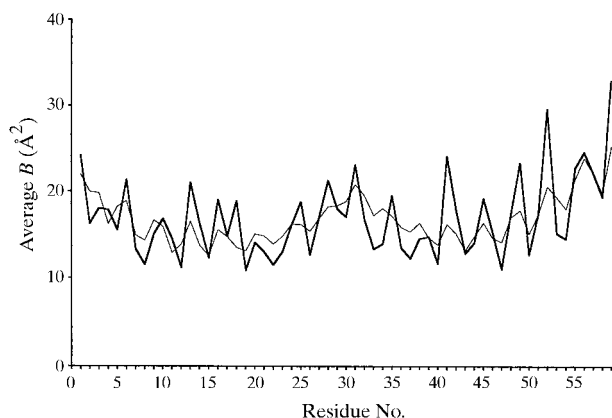


Fig. 13. Variation of the *B* factors (averaged over all residue atoms) along the sequence for the refined model (bold line) and the TLS model (thin line).

Structural data have been deposited with the Protein Data Bank.†

## APPENDIX A

If $\sigma_i$ is the estimated standard deviation for the refined parameter $x_i$ (*i.e.* $x,y,z,U_{ij}$), the commonly used mean value can be replaced by the more reliable weighted arithmetic mean,

$$\bar{x}_w = \frac{\sum_i w_i x_i}{\sum_i w_i},$$

with $w_i = 1/\sigma_i^2$. Each parameter $x_i$ is thus weighted by a factor which depends on the standard deviation associated with this parameter.

The standard deviation of the mean is given by

$$\sigma_w^2(\bar{x}) = \frac{\sum_i w_i(x_i - \bar{x})^2}{(n-1)\sum_i w_i}.$$

The function $t$ defined by

$$t = \frac{\bar{x}_w - m}{\sigma_w(\bar{x})},$$

follows the Student's distribution (Fisher & Yates, 1953) and is used to test whether the distribution of parameters $x_i$ deviates significantly from a distribution with the mean value $m$.

## References

Allen, F. H., Kennard, O. & Taylor, R. (1983). *Acc. Chem. Res.* **16**, 146–153.

Banner, D. W., Kokkinidis, M. & Tsernoglou, D. (1987). *J. Mol. Biol.* **196**, 657–675.

Barlow, D. J. & Thornton, J. M. (1988). *J. Mol. Biol.* **201**, 601–619.

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

Blundell, T., Barlow, D., Borkakoti, N. & Thornton, J. (1983). *Nature (London)*, **306**, 281–283.

Brooks, C. L., Karplus, M. & Pettitt, B. M. (1988). In *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*, pp. 75–83. New York: John Wiley.

Brünger, A. T. (1990). *X-PLOR Manual*, Version 2.1, The Howard Hughes Medical Institute, New Haven, Connecticut, USA.

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**54**, 760–763.

Dauter, Z., Sieker, L. C. & Wilson, K. S. (1992). *Acta Cryst.* B**48**, 42–59.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Fisher, R. A. & Yates, F. (1953). *Statistical Tables for Biological, Agricultural and Medical Research*, 5th ed., Table III. New York: Oliver and Bond Ltd.

Glover, I., Haneef, I., Pitts, J., Wood, S., Moss, D., Tickle, I. & Blundell, T. (1983). *Biopolymers*, **22**, 293–304.

Ichiye, T. & Karplus, M. (1987). *Proteins*, **2**, 236–259.

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

Jones, T. A. (1985). *Methods Enzymol.* **115**, 157–171.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.

Kokkinidis, M., Vlassi, M., Papanikolau, Y., Kotsifaki, D., Kingswell, A., Tsernoglou, D. & Hinz, H. J. (1993). *Proteins*, **16**, 214–216.

Konnert, J. H. & Hendrickson, W. A. (1980). *Acta Cryst.* A**36**, 344–350.

Kuriyan, J., Petsko, G. A., Levy, R. M. & Karplus, M. (1986). *J. Mol. Biol.* **190**, 227–254.

Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.

Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). *J. Mol. Biol.* **231**, 1049–1067.

Lesser, G. J. & Rose, G. D. (1990). *Proteins*, **8**, 6–13.

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

Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983). *Acta Cryst.* B**39**, 480–490.

Matthews, B. W. (1966). *Acta Cryst.* **20**, 82–86.

Otwinowski, Z. (1991). *DENZO. A Film Processing Program for Macromolecular Crystallography*. Yale University, New Haven, Connecticut, USA.

Paliakasis, C. D. & Kokkinidis, M. (1992). *Protein Eng.* **5**, 739–748.

Parkin, S., Rupp, B. & Hope, H. (1996). *Acta Cryst.* D**52**, 18–29.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Remington, S., Wiegand, G. & Huber, R. (1982). *J. Mol. Biol.* **581**, 111–152.

Richardson, J. S. & Richardson, D. C. (1989). *Principles and Patterns of Protein Conformation*. In *Prediction of Protein Structure and the Principles of Protein Conformation*, edited by G. D. Fasman, pp. 1–98. New York: Plenum Press.

Schomaker, V. & Trueblood, K. N. (1968). *Acta Cryst.* B**24**, 63–77.

Sevcik, J., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1996). *Acta Cryst.* D**52**, 327–344.

---

† Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1NKD, R1NKDSF).

Sheldrick, G. M. (1993). *SHELXL*93, *Program for crystal structure refinement*, University of Göttingen, Germany.

Stec, B., Zhou, R. & Teeter, M. M. (1995). *Acta Cryst.* D**51**, 663–681.

Sternberg, M. J., Grace, D. E. & Phillips, D. C. (1979). *J. Mol. Biol.* **130**, 231–252.

Teeter, M. M., Roc, S. M. & Heo Na, H. (1993). *J. Mol. Biol.* **230**, 292–311.

Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *Acta Cryst.* A**43**, 489–501.

Vlassi, M., Steif, C., Weber, P., Tsernoglou, D., Wilson, K., Hinz, H.-J. & Kokkinidis, M. (1994). *Nature Struct. Biol.* **1**, 706–716.

Wilson, K. S., Butterworth, S., Dauter, Z., Lamzin, V. S., Wals, M., Wodak, S., Pontius, J., Richelle, J., Vaguine, A., Sander, C., Hooft, R. W. W. W., Vriend, G., Thornton, J. M., Laskowski, R. A., MacArthur, M. W., Dodson, E. J., Murhudov, G., Oldfield, T. J., Kaptein, R. & Rullmann, J. A. C. (1998). *J. Mol. Biol.* In the press.