Acta Cryst. (1995). D51, 177–189

Crystal Structure of Narbonin at 1.8 Å Resolution

BY MICHAEL HENNIG,* SABINE PFEFFER-HENNIG,† ZBIGNIEW DAUTER AND KEITH S. WILSON

European Molecular Biology Laboratory (EMBL) c/o DESY, Notkestrasse 85, 22603 *Hamburg, Germany*

AND BERNHARD SCHLESIER AND VAN HAI NONG

Institute of Plant Genetics and Crop Plant Research, Correnstrasse 3, 06466 *Gatersleben, Germany*

(*Received* 14 *February* 1994; *accepted* 23 *August* 1994)

Abstract

The three-dimensional structure of narbonin, a seed protein from *Vicia narbonensis L*, has been determined at 1.8 Å resolution. Phase information was obtained by multiple isomorphous replacement and optimized anomalous dispersion. The narbonin structure was initially traced with only 17% amino-acid sequence information and preliminarily refined to a crystallographic *R*-factor of 16.5%. It is now refined to 15.9% using full sequence information derived from cDNA and after the addition of more solvent molecules. The monomeric molecule of narbonin is an eight-stranded parallel β-barrel surrounded by α-helices in a β/α-topology similar to that first observed in triose phosphate isomerase. Differences exist in the N-terminal part of the polypeptide chain, where the first helix is replaced by a loop and the second β-strand is followed by an additional antiparallel α-sheet placed parallel on top of α-helices $\alpha 3$ and $\alpha 4$. Two short additional secondary structures are present. The first, an α-helix, is situated between the seventh β-strand and the following helix, and the second, which is a 3_{10} helix, between the eighth strand and the C-terminal helix. The most striking observation is the lack of a known enzymatic function for narbonin, because all TIM-like structures known so far are enzymes.

1. Abbreviations

TIM, triose phosphate isomerase; PRAI, *N*-(5′-phosphoribosyl)anthranilate isomerase; IGPS, indole-3-glycerol-phosphate synthase; TS, tryptophan synthase (α subunit); ARP, automated refinement procedure; MIR, multiple isomorphous replacement.

2. Introduction

Storage globulins are a family of proteins of different size and structural organization, found in all legume seeds. They share several properties, such as similar amino-acid composition, salting-in solubility, time of synthesis, and degradation and lack of enzymatic or biological activity. Subsequently, the classification of globulins from legume seeds was performed using their sedimentation coefficient (Derbyshire, Wright & Boulter, 1976). Narbonin is a member of the smaller 2*S* globulins, a monomer of 290 amino-acid residues and consists of a single polypeptide chain.

The three-dimensional structures of two 7*S* globulins, phaseolin and canavalin, have been determined to 3.0 and 2.6 Å resolution, respectively (Lawrence, Suzuki, Varghese, Davis, Van Donkelaar, Tulloch & Colman, 1990; Ko, Ng & McPherson, 1993). Comparison of the two structures shows high similarity. They are composed of two domains, which are almost exactly related by twofold symmetry, with each domain being subdivided into two structural elements: an eight-stranded β-barrel with Swiss roll topology and an extended loop with a helix-turn-helix motif. The homology of the proteins is confirmed by 60% sequence identity.

Until now, no three-dimensional structure for the 11*S* globulins has been published, although some 11*S* globulins have been crystallized (Hennig & Schlesier, 1994).

The crystal structure of narbonin from *Vicia narbonensis L* was determined at 1.8 Å resolution by multiple isomorphous replacement. Because of lack of sequence information, a tentative amino-acid sequence was derived by examining the shapes of the side chains in the electron-density map. The structure was preliminarily refined to the crystallographic *R*-factor 16.5% and the tentative amino-acid sequence published (Hennig, Schlesier, Dauter, Pfeffer, Betzel, Höhne & Wilson, 1992; Hennig & Schlesier, 1993). Narbonin turned out to be folded as a β/α-protein with β-strands and α-helices alternating to form an eight-stranded parallel β-barrel surrounded by a ring of seven α-helices. This topology is similar to that first observed in triose phosphate isomerase (Banner, Bloomer, Petsko, Phillips, Pogson, Wilson, Corran, Furth, Milman, Offord, Priddle & Waley, 1975) and later found in many different enzymes.

Here we present the refinement and analysis of the narbonin structure at 1.8 Å resolution, considering the

* Correspondence address: Michael Hennig, Biozentrum, Department of Structural Biology, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland.

† Present address: Sandoz AG, CH-4002 Basel, Switzerland.

ISSN 0907-4449

Table 1. *Statistics of the data for the parent protein*

X-ray source	Beamline X11/EMBL Hamburg storage ring DORIS	
Detector	Imaging plate	
Wavelength (Å)	0.96	
Crystal-to-plate distance (mm)	180	235
Maximum resolution (Å)	1.8	2.2
Total crystal rotation (°)	122	104
Oscillation angle/image (°)	2	2
Exposure time/image (s)	50	5
No. of measured reflections	124 224	
No. of unique reflections	27 227	
Completeness (1.8–10 Å) (%)	95.5	
$*R_{sym}$(1.8–10 Å) (%)	5.6	

$*R_{sym} = \sum(\sum_i |I_1 - \langle I\rangle|/\sum\langle I\rangle$, where $\langle I\rangle$ is the mean of the intensity measurements, I_i, and the summation extends over all reflections.

complete cDNA-derived amino-acid sequence. The consequences of these results for classifying narbonin as a true storage protein are discussed.

3. Experimental

3.1. *Crystallization, data collection and processing*

Narbonin was prepared from seeds of *Vicia narbonensis*, as described previously (Schlesier & Scholz, 1974; Schlesier, Manteuffel, Rudolph & Behlke, 1978).

Table 2. *Data collction and analysis for the derivatives used for phasing*

Compound	Concentration (mM)	Soak time (days)	Wavelength (Å)	Resolution (Å)	$*R_{iso}$ (%)	$\dagger K_{emp}$ (%)	Completeness (%)
$Hg(CH_3COO)_3$	10	8	0.95	2.1	26.4	6.2	07.4
$Hg(CH_3COO)_2$	10	4	1.009	3.8	22.7	9.9	92.1
$UO_2(NO_3)_2$	1	0.5	0.96	4.0	29.8	12.2	67.5
K_2PiCl_4	5	20	1.54	2.2	22.6	5.3	68.5
$SmCl_3$	7.5	2.5	1.54	2.2	11.9	4.3	93.1
$SmCl_3$	5	4	1.009	3.9	13.2	4.4	94.3

$*R_{iso} = |\sum(F_{PH} - F_p)|\sum F_p$, where F_{PH} is the derivative and F_p the parent structure-factor amplitude.

$\dagger K_{emp} = \sum |F_{PH} - F_p|/\sum F_{PH+} - F_{PH-}|$.

Large monoclinic crystals were obtained by dissolving microcrystals and recrystallization using a vapour diffusion-induced pH shift, simultaneous with the lowering of the ionic strength of the drop solution (Hennig, Schlesier, Pfeffer & Höhne, 1990). Purification was performed using citrate/phosphate buffer (McIlvaine, 1921) and the pH shift was achieved by Tris/HCl buffer. Crystals belong to space group $P2_1$ with the cell dimensions a = 46.9, b = 75.5, c = 50.9 Å and β = 120.5°. Rotation data were collected on an imaging plate scanner

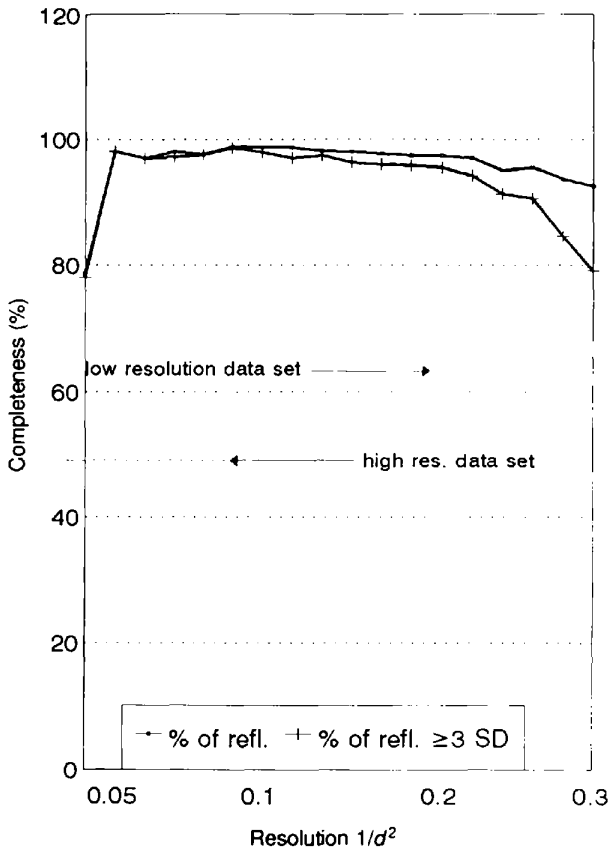

Fig. 1. Completeness of the native data as a function of resolution.

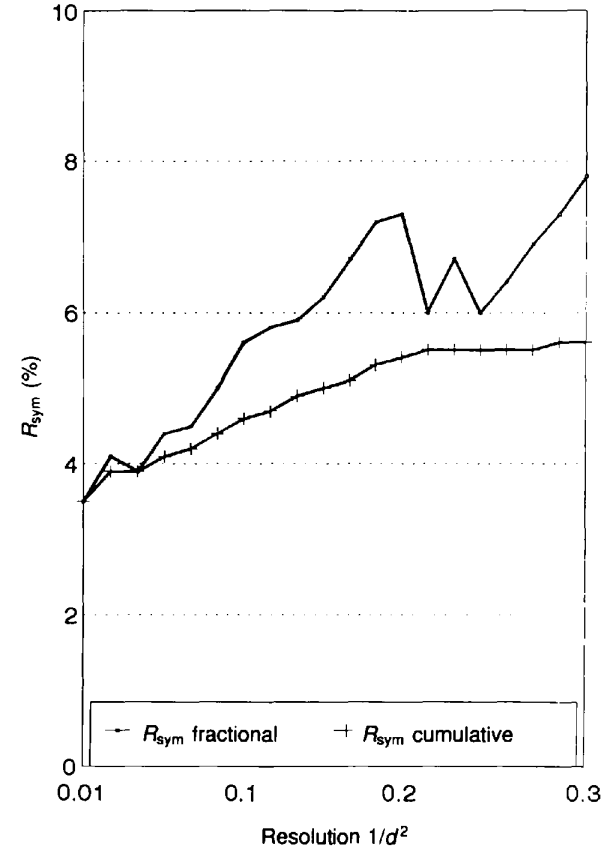

Fig. 2. R_{sym} ($R_{sym} = \sum_i |I_i - \langle I\rangle|/\sum\langle I\rangle$, where $\langle I\rangle$ is the mean of the intensity measurements I_i) as a function of resolution.

Table 3. *Results of the refinement of the heavy-atom parameters*

Compound	R_c (%)	x	y	z	B (Å^2)	Occupancy	Phasing power (Resolution, Å)
$Hg(CH_2COO)_2$	29	0.875	0.000	0.911	12.3	0.92	5.06
		9814	0.114	0.630	15.6	1	(2.2–10)
		0.574	0.011	0.332	23.9	0.78	
		0.551	0.070	0.369	40.5	0.35	
$Hg(CH_3COO)_2$	27	0.869	0.000	0.908	5.1	0.86	5.11
		0.815	0.114	0.632	18.9	0.98	(3.8–10)
		0.570	0.012	0.334	15.6	0.84	
		0.550	0.070	0.369	9.7	0.22	
$SmCl_3$	63	0.505	0.074	0.104	13.4	0.45	0.97
							(2.3–10)
K_2PtCl_4	51	0.879	0.000	0.918	4.7	0.69	2.55
		0.795	0.115	0.597	21.4	0.50	(2.2–3.5)
		0.588	0.011	0.360	25.2	0.51	
$UO_2(NO_3)_2$	38	0.507	0.079	0.111	−10.5	0.92	2.00
		0.467	0.071	0.516	−2.9	0.37	(4.0–10)
		0.154	0.277	0.019	40.6	0.94	
		0.684	0.020	0.083	58.1	0.50	
		0.904	0.318	0.425	64.0	0.55	

$R_c = \sum |F_{PH_{calc}} - F_{PH_{obs}}| / \sum F_{PH_{obs}}$, the Cullis R factor for centric reflections. The phasing power is defined as F_H/E, where F_H is the heavy-atom structure-factor amplitude and E the r.m.s. residual lack of closure. x, y, z = fractional coordinates, B is the isotropic temperature factor. Resolution range used for phase calculation, occupancy was scaled to 1.0 for the highest value.

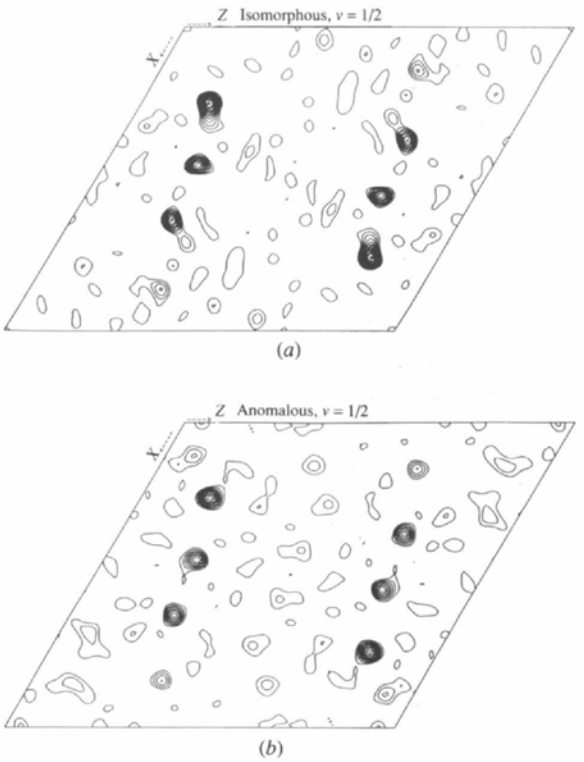

Fig. 3. Harker section (v = 1/2) of the Patterson syntheses calculated using (*a*) isomorphous differences between the mercury derivative and the native data and (*b*) the anomalous differences of the mercury derivative. Derivative data were collected on X31 with λ = 0.95 Å, resolution range 2.5–10 Å.

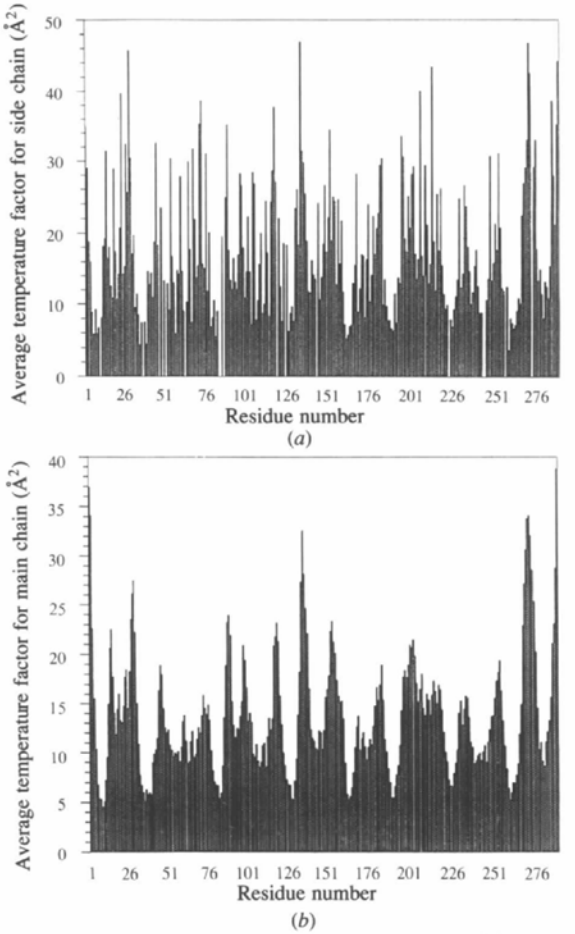

Fig. 4. Distribution of the average temperature factors (Å^2) along the polypeptide chain for (*a*) side-chain atoms and (*b*) main-chain atoms.

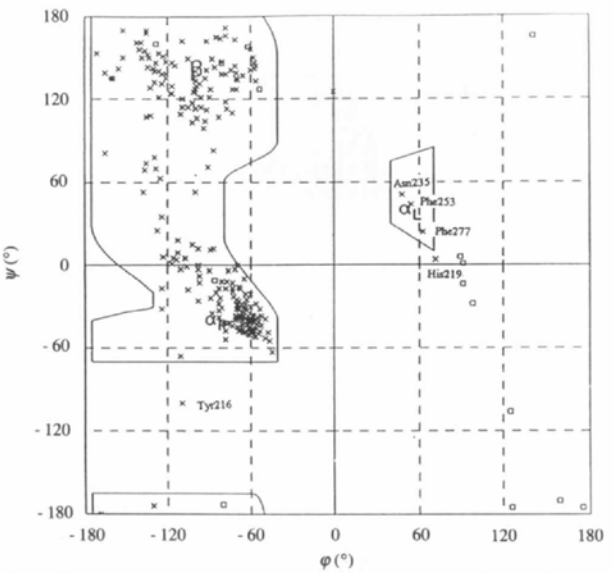

Fig. 5. Ramachandran φ, ψ plot for the refined structure of narbonin. The energetically favoured regions for all residues other than glycine are indicated as bounded parts of the plot. Only two non-glycine residues have unfavoured φ, ψ values (Tyr216 and His219, see text).

Table 4. *Summary of the course of the least-squares restrained-parameter refinement*

Step	1		2	3	4	5	6	7	8	
No. of cycles	3	2	7	4	5	5	5	4	3	4
Resolution range (Å)	2.8–10	2.2–10	2.2–10	2.2–10	2.2–10	2.2–10	2.2–10	2.2–10	2.2–10	1.8–10
No. of reflections	7382	15 192	15 192	15 192	15 192	15 192	15 192	15 192	15 192	27 227
No. of atoms	1202	1202	1431	1508	1627	1687	1740	1803	1835	1835
% of the protein atoms	51	51	61	65	69	71	74	77	77	77
No. of amino acids	213	213	236	245	262	264	281	290	292	292
No. of solvent molecules	0	0	0	0	11	28	11	9	41	41
X-ray/geometry weighting	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.2	1.2	1.0
R (begins, %)*	51.8	49.6	50.8	41.2	42.8	38.2	25.3	35.8	33.7	37.6
R (end, %)*	45.7	47.3	40.2	37.3	38.1	33.5	31.8	31.9	30.3	34.6
Mean temperature (Å^2)	—	—	15.4	14.7	15.6	15.5	15.3	15.2	17.7	16.7
Residual deviations										
Bond lengths (1–2 distances, Å)	—	—	—	—	0.049	0.055	0.050	0.045	0.044	0.040
Planar groups (Å^2)	—	—	—	—	0.022	0.037	0.039	0.029	0.030	0.028
Chiral volumes (Å)	—	—	—	—	1.005	1.100	1.095	1.150	1.099	1.080
Torsion angles of peptide plane (°)	—	—	—	—	15.1	10.8	10.5	8.2	6.9	6.6

Step	9	10	11	12	13	14	15	16
No. of cycles	5	5	9	5	9	6	6	9
Resolution range (Å)	1.8–10	1.8–10	1.8–10	1.8–10	1.8–10	1.8–10	1.8–10	1.8–10
No. of reflections	27 227	27 227	27 227	27 227	27 227	27 227	27 227	27 227
No. of atoms	1846	1924	2175	2428	2527	2775	2947	2847
% of the protein atoms	77	79	89	95	97	100	100	100
No. of amino acids	290	289	290	288	288	289	289	289
No. of solvent molecules	51	67	87	198	261	441	596	513
X-ray/geometry weighting	1.0	1.0	1.0	0.8	0.8	1.0	0.8	0.5
R (begins, %)*	33.1	36.2	32.1	24.4	20.3	20.8	18.8	16.1
R (end, %)*	31.5	29.6	21.9	19.3	17.2	16.5	15.8	15.9
Mean temperature (Å^2)	17.1	16.5	17.6	17.9	18.2	20.0	20.6	21.0
Residual deviations								
Bond lengths (1–2 distances, Å)	0.038	0.030	0.027	0.018	0.012	0.025	0.019	0.015
Planar groups (Å^2)	0.026	0.022	0.020	0.019	0.014	0.019	0.018	0.014
Chiral volumes (Å)	1.071	0.87	0.48	0.52	0.27	0.30	0.20	0.17
Torsion angles of peptide plane (°)	6.0	5.5	5.3	4.1	3.5	3.5	3.3	3.1

* *R* factor calculated without σ-cut.
† Mean isotropic *B* factor using all atoms, including solvent.

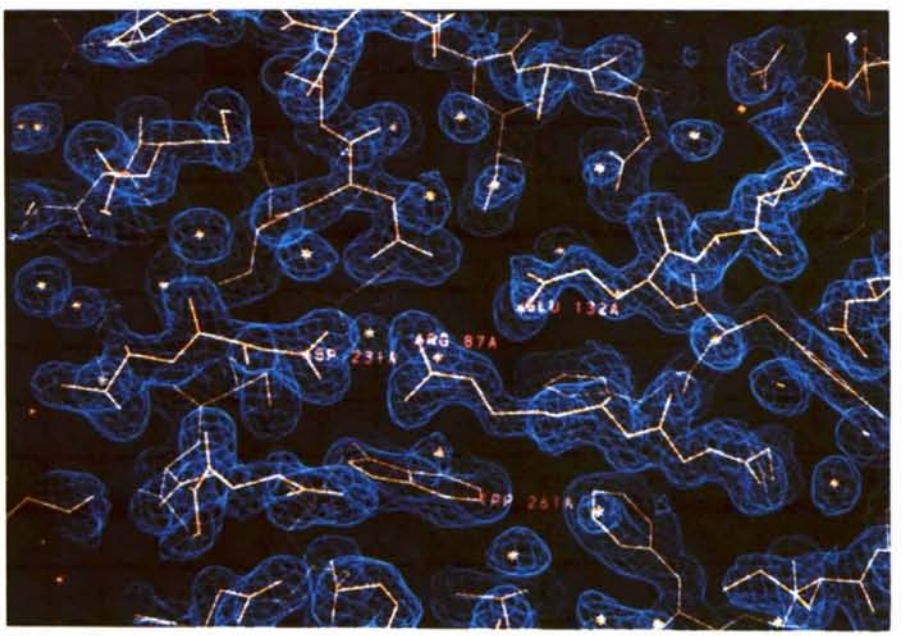

Fig. 6. View of a representative region of the final $(3F_o - 2F_c)$ map contoured at 1.2σ.

using synchrotron radiation on the EMBL X11 beamline (DESY, Hamburg). The needle-shaped crystals of dimensions *ca* 0.3 × 0.3 × 1.5 mm were mounted in glass capillaries. To avoid the necessity of rotation through 180° for complete data, the crystal was oriented with the c* axis approximately parallel to the rotation axis. Because of the high stability of the crystal, the possibility of its translation in the X-ray beam, the use of short wavelengths and the high intensity of the X-ray source, only one crystal was necessary for the parent protein and for each of the derivatives. A rotation angle of 2° was used for each exposure. Two data sets were collected with different exposures. Details on data collection parameters for the parent crystal are summarized in Table 1. The data were processed with the program package *MOSFLM* (Leslie, Brick & Wonacott, 1986). All the programs used were from the *CCP*4 package (Collaborative Computational Project, Number 4, 1994), unless otherwise stated. The initial orientation matrices were determined from still photographs and refined using the program *POSTCHK*. The reflections were integrated with the program *MOSFLM*, using profile fitting. The integrated intensities were scaled with the program *ROTAVATA* and merged using verted to structure-factor amplitudes using *TRUNCATE*. The native data are of high quality in all resolution shells. The overall percentage completeness and reflections with intensities greater than three standard deviations are shown as a function of resolution in Fig. 1. Even in the highest resolution range, 80% of the data are greater than 3σ. The overall R factor, based on intensities for the symmetry-equivalent reflections for the combined low- and high-resolution data, was 5.6%. In the highest resolution shell of 1.8–1.85 Å, R_{sym} is 7.8% (Fig. 2).

3.2. *Derivatives*

Heavy-atom derivatives were prepared by soaking the crystals in the mother liquor containing 0.1 *M* Tris/HCl buffer, pH 7.2, and the appropriate heavy-atom compound. Possible derivatives were screened by selecting those crystals which showed no serious damage after soaking. Subsequently, a preliminary data set at 4 Å was collected using synchrotron radiation, which required not more than 3 h total beam time. The derivatives were analysed and data to higher resolution collected, if the derivative proved to be useful. In addition, two derivatives were investigated using a sealed-tube generator (Cu $K\alpha$ radiation). The statistics of the heavy-atom derivative data are summarized in Table 2. The mercury acetate derivative gave very clear anomalous and isomorphous difference Patterson maps, as the wavelength was optimized to just below the absorption edge of mercury (λ = 0.95 Å, Fig. 3). Two binding sites were identified from the Harker section. The third peak of the Harker section corresponds to a cross vector. The remaining sites of this derivative and all sites of the other derivatives were found by cross-phased difference Fourier syntheses. The refinement of the heavy-atom sites was performed using *REFINE* and phase calculations with *PHARE*. The heavy-atom positions and refined parameters are summarized in Table 3. The phasing power of the mercury derivative is exceptionally high, which indicates the tight binding and high isomorphism of this derivative, even at high resolution. In contrast, the samarium and platinum derivatives were relatively poor. A comparison of the phase angles, calculated from the refined structure, shows that the use of only the mercury and uranyl derivatives produced the best phases (mean $\Delta\varphi = 40°$, $\Delta\varphi = \varphi_{MIR} - \varphi_{final\ model}$, 10–2.2 Å resolution range). In contrast, phasing without these resulted in a very noisy map and a mean $\Delta\varphi$ of 69°. The initial figure of merit, calculated using all the derivatives mentioned above, was 0.71 for all reflections between 2.2 and 10 Å resolution (mean $\Delta\varphi = 43°$). The application of solvent flattening (Wang, 1985) did not improve the quality of the electron-density map significantly and, therefore, was not used in model building.

3.3. *Model building and refinement*

MIR-phases based on all derivatives were used to calculate a map weighted according to the figure of merit. Data were limited to the resolution range 2.2–8 Å. The electron density was plotted in steps of 0.8σ, starting at 1.2σ on Plexiglas sheets. Several strands could be easily recognized, but large parts of the molecule remained unclear. Some Cα positions were labelled and used as a marker on a graphics station. All further work was performed using the *FRODO* software package (Jones, 1985) on Evans & Sutherland PS300 and later ESV30 graphics display systems. In the first step of electron-density interpretation, only alanine and glycine residues were built in regions where a continuous chain was visible. The chain directions defined by the carbonyl O atoms were unclear in many cases, but in retrospect, only three of the ten segments were traced in the wrong direction. After refinement of the partial model using the restrained least-squares procedure (Hendrickson & Konnert, 1980) implemented in *PROLSQ*, a phase combination of the model and MIR phases was performed using the program *COMBINE* (Bricogne, 1976). The incompleteness of the model (*ca* 50% of all atoms) was taken into account by lowering the weight of the model contribution. The resulting slightly improved electron density was used to correct misfit parts of the model and to extend it. After several cycles of model building, refinement of the model, phase combination and recalculation of the electron density, the molecule was extended considerably and several side chains identified. After step 8 of the refinement, Fourier synthesis using data to 1.8 Å resolution based on the coefficients $(2F_o - F_c)\times \exp(i\alpha c)$ was used simultanously with the combined map for manual rebuilding. The N-terminus and two

further internal peptides (overall, 17% of the residues), which were chemically sequenced, could be located in the molecule. An automated refinement procedure (ARP) was applied at a crystallographic R factor of 30.1% for all data in the resolution range 1.8–10 Å. The method and its application to narbonin is published elsewhere (Lamzin & Wilson, 1993). ARP led to a substantial improvement of the electron density and the model was rebuilt in several parts of the molecule, according to the features of the ARP map. Two further cycles of application of ARP and manual rebuilding using interactive graphics led to the preliminary refined narbonin structure with most of the side chains identified (Hennig, Schlesier, Dauter, Pfeffer, Betzel, Höhne & Wilson 1992). After the cDNA-derived amino-acid sequence became available, the model was corrected and some more water molecules added. The interpretation was carried out using $(3F_o - 2F_c)$ and $(F_o - F_c)$ electron-density maps. In order to find the positions of water molecules, the highest positive peaks in the $(F_o - F_c)$ difference Fourier map were located using the program *PEAKMAX* and water molecules were placed at these maxima, if they were in a geometrically reasonable position. Subsequently, two further steps of refinement were performed. In summary, the refinement was carried out in 16 steps comprising 110 cycles of least-squares minimization, Table 4.

4. Structure of narbonin

4.1. *Quality of the structure*

The refined model of narbonin contains 289 of the 290 residues and 513 water molecules. The backbone atoms of all residues, except the C-terminal Arg290 which is disordered and, therefore, not included in the model, have well defined electron density. Nevertheless, some of the side chains could not be fully modelled because of disorder and weak or missing density. Thus, Lys289 was traced as an alanine. Further poorly defined residues can be recognized by their high B values. These parts of the molecule are the N- and three C-terminal residues and the side chains of Glu29, Arg73, Arg135, Lys214, Arg271 and Asp272. Regions of weak electron density correspond to positions with high thermal-motion parameters (Fig. 4). In the main-chain B-factor plot (Fig. 4), the N- as well as the C-terminal and β/α-loop regions with B factors > 30 Å^2 can be identified. In contrast, all eight parallel β-strands have low temperature factors.

The R factor, based on 27 227 observed structure factors between 10 and 1.8 Å resolution, is 15.9%. The r.m.s. deviation from ideal bond length is 0.015 Å and from ideal bond angles is 3.1°. Further stereochemical and crystallographic data of the refined narbonin structure are given in Table 4. The φ, ψ plot of the final model (Ramachandran & Sasisekharan, 1968) is presented in Fig. 5 and shows only two residues outside the allowed regions. Tyr216 lies directly at the end of α-helix 6, but we have no explanation for its unusual geometry. His219 lies close to the α_L section and, like Phe253, at the start of an α/β-loop (Table 6). Fig. 6 shows a representative part of the structure with the electron density superimposed, displaying the high quality of the map. The plot of the R factor as a function of resolution is shown in Fig. 7, together with the lines for a series of theoretical coordinate errors (Luzzati, 1952). It allows a rough estimation of the mean coordinate error, which can be much lower or higher in different parts of the molecule. The deviation at low resolution is caused by the relatively poor model for the bulk solvent structure.

4.2. *Comparison of the X-ray sequence and the cDNA sequence*

Comparison of the sequence derived by interpretation of the electron density at 1.8 Å resolution and cDNA-derived amino-acid sequence is shown in Fig. 8. The alignment of the sequences shows general agreement. Only 49 of the 290 amino acids (17%) were determined before publication of the X-ray sequence by peptide sequencing. In the remaining polypeptide chain, 203 of the 241 residues were correctly identified by X-ray sequencing (84.2%), a high quota compared with earlier trials of X-ray sequencing (Colman, Jansonius & Matthews, 1972; Herriott, Watenpaugh, Sieker & Jensen 1973). In addition, no renumbering of the residues was necessary, because the cDNA-derived sequence has no deletions or insertions compared with the X-ray sequence.

Nevertheless, at this resolution it is difficult to distinguish between O, N and C atoms in X-ray crystallography, because of their similar atomic form factors and small difference in bond length. This results in difficulties in distinguishing between Asn, Asp and Leu, as well as between Gln and Glu and between Thr and Val. On the other hand, there are differences in the environment of the amino-acid side chains. For instance, valine and leucine are hydrophobic and do not form hydrogen bonds. In addition, the conformation

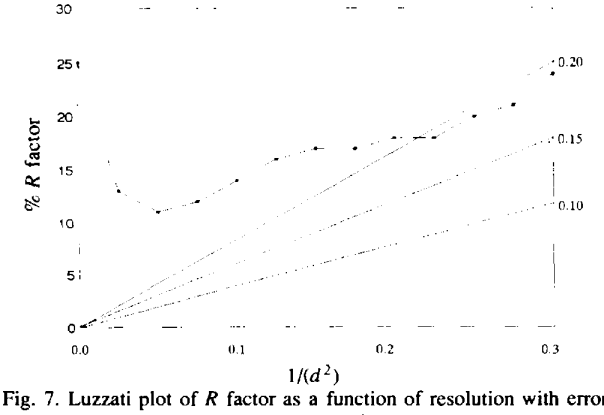

Fig. 7. Luzzati plot of R factor as a function of resolution with error levels shown as lines.

of the leucine side chain differs significantly from Asn and Asp. The X-ray sequencing includes Asn and Asp together (Asx or B in one letter code) and Leu separately. Only the residue Leu287 was inaccurately identified as Asx.

For glutamine and glutamic acid, Glx or Z were used in the X-ray sequence. Only for clearly identified salt bridges were Asp or Glu included in the sequence, which turned out to be correct in all cases.

The differentiation between valine and threonine, based on differences in the environment of the residues, was less successful. Eight of the 35 Val/Thr residues were mismatched. This supports the unusual property of narbonin having hydrophobic residues accessible to solvent, because seven of the eight residues identified as Thr in the X-ray sequence and turned out to be Val, have their position on the protein surface exposed to solvent.

A comparison of the correct identification of the residues by amino-acid type is shown in Fig. 9. All the aromatic and hydrophobic side chains of Trp, Tyr, Phe, Pro, Ala, Gly and Cys were identified correctly by interpretation of the highly accurate electron-density distribution of these side chains. On the other hand, the hydrophilic and charged amino acids are more difficult to identify. Especially the flexible Lys and Arg side chains were often wrongly sequenced, if they are not stabilized by ionic interactions. Even with the knowledge of the sequence, it is not possible to trace the side chains of some residues because of disorder. The single methionine could not be found, because of its two conformations.

For the estimation of the reliability and quality of X-ray structures and their correct tracing in the electron density, several factors, including temperature factor, real-space correlation coefficient, solvent accessibility *etc.*, for every atom or residue, have been used in protein crystallography. We introduced a probability factor to

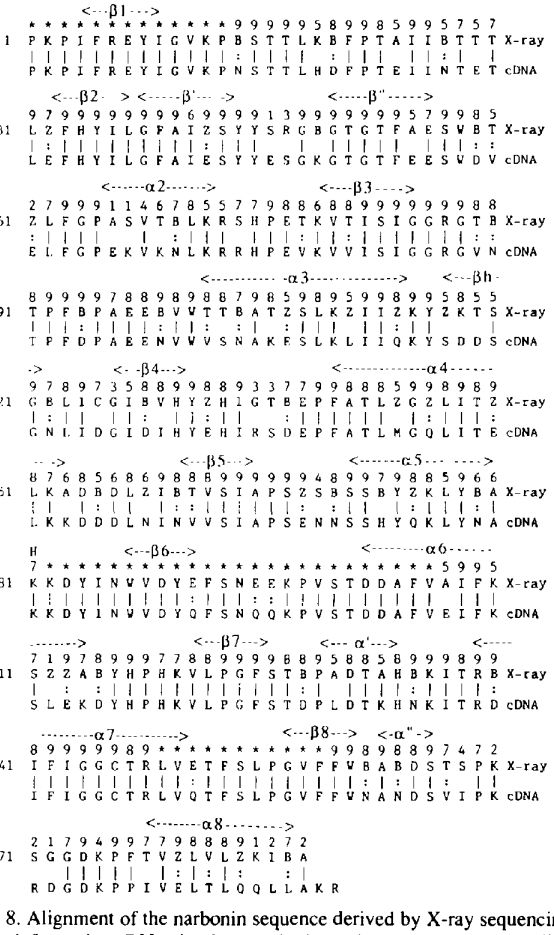

Fig. 8. Alignment of the narbonin sequence derived by X-ray sequencing and from the cDNA in the standard one-letter code. The top line shows the secondary structure. The second line shows the estimation of the accuracy of X-ray sequencing, where * indicates parts known by peptide sequencing, 1 substantial disorder in the electron density and 9 certainty in interpretation (see Hennig, Schlesier, Dauter, Pfeffer, Betzel, Höhne & Wilson, 1992). The third line shows the identity of the X-ray- and cDNA-derived sequence using | and : for identity with X-ray sequencing, because the residues Val/Thr, Glu/Gln and Asp/Asn cannot be distinguished by X-ray sequencing. The bottom line shows the cDNA-derived sequence.

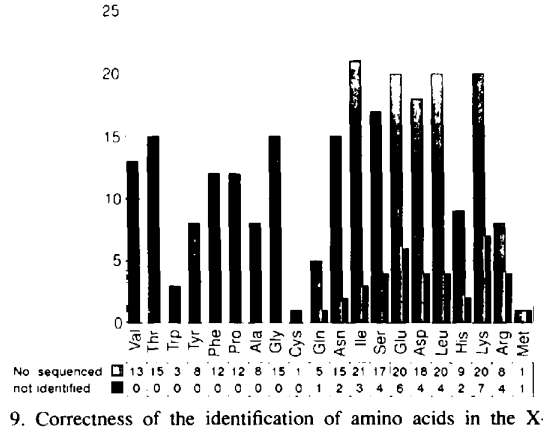

Fig. 9. Correctness of the identification of amino acids in the X-ray sequencing of narbonin according to residue type. Val/Thr were stated to be correctly identified if they were determined as one of the residue types in the X-ray sequence.

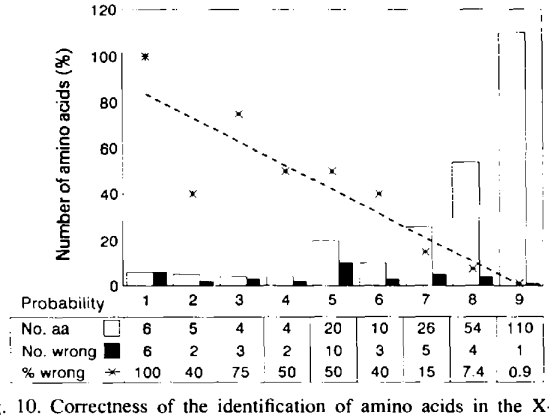

Fig. 10. Correctness of the identification of amino acids in the X-ray sequencing of narbonin according to the probability factor.

estimate the certainty of the correct identification of the side chain. A low factor was used to indicate substantial disorder in the electron density, allowing no identification at all and the highest number 9 indicates well ordered side-chain density with high confidence in the interpretation. The result in Fig. 10 shows the high conformity of this estimation with reality by counting the number of wrongly identified amino acids for each probability factor. This confirms that it is possible to indicate the correctness of structure determination in protein crystallography of every atom or residue by careful consideration of the quality factors, especially the electron-density distribution.

4.3. *Overall structure*

The narbonin molecule consists of a parallel β/α-barrel domain having the dimensions *ca* 45 × 45 × 30 Å. A schematic cartoon of the structure is shown in Fig. 11. The positions, length and mean φ, ψ angles of helices and β-strands are summarized in Table 5. The comparison of the fold with other β/α-barrel proteins for which three-dimensional crystal structures have been determined so far shows some differences. The N-terminal region of the chain shows most differences. The first and second β-strands are connected by an extended loop comprising two β-turns. However, the *B* factors of the main-chain atoms are not significantly higher in this region without regular secondary structure compared with the neighbouring helices. The second strand ($\beta 2$) extends almost without interruption of the β-sheet hydrogen-bonding network and the β-strand conformation angles (φ, ψ) to an antiparallel β-sheet motif being on the top of α-helices $\alpha 3$ and $\alpha 4$. This structural element is formed by the β'- and β''-strands and does not cover the C-terminal entrance of the β-barrel. It follows the $\alpha/\beta/\alpha$ folding pattern found in all TIM-barrel proteins. Short secondary-structural elements are additionally found in the loop regions between strands and helices at the top of the molecule and between helices and strands at the bottom.

In several TIM-barrel proteins there is an additional helix found in the loop before helix 8, promoting the binding of phosphate because of its macrodipole moment. In narbonin, a turn of type III (3_{10}-helix) is situated here and there is a short α-helix in the neighbouring loop before helix 7. The N-termini of both helices are adjacent. Between $\alpha 3$ and $\beta 4$ there is a hairpin structure (Fig. 15).

4.4. *Parallel β-barrel and β/α loop regions*

The β-barrel in narbonin is constructed from 46 residues and is elliptical with maximum cross sections 13 × 18 Å. Most of the β-strands consist of six residues, only $\beta 3$ is longer with eight residues and $\beta 5$, $\beta 6$ and $\beta 8$ are shorter (Table 5). The β-strands of the barrel form 44 main-chain hydrogen bonds. The core of the barrel is formed not only by hydrophobic side chains. The charges of the two polar residues Arg6 and Asp189 are compensated by forming a salt bridge. In addition, Arg6 forms hydrogen bonds to OG of Ser163, Wat8 and His34. There are two internal waters (Wat8 and Wat49), forming hydrogen bonds to surrounding main-chain and side-chain atoms (Fig. 12). They have *B* factors of 8.0 and 8.4 Å^2, respectively, which are in the same range as surrounding protein atoms. Asp128 forms hydrogen bonds to OH of Tyr8, to OG of Ser83 and Ser163 and the water molecule Wat49. The side chains of the parallel β-strands which point into the centre of the barrel form four layers (for details, see Hennig, Schlesier, Dauter,

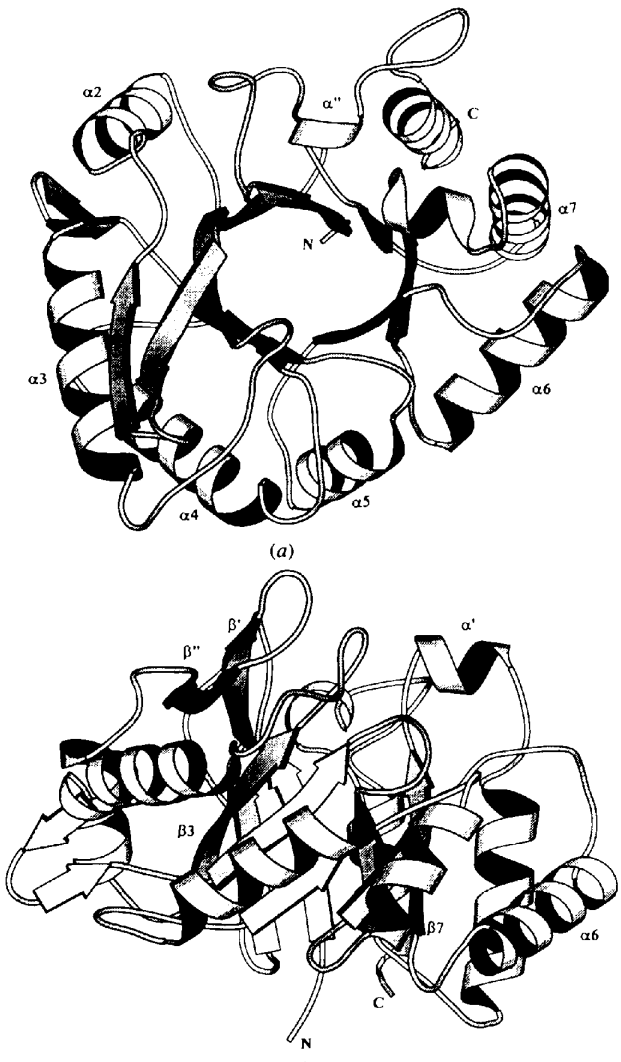

Fig. 11. Schematic ribbon plot of the fold of narbonin from *Vicia narbonensis L.* The view is down the axis of the eight-stranded β-barrel (*a*) and rotated 90° to give a side view (*b*). The missing $\alpha 1$-helix and the additional antiparallel β-sheet are at the beginning of the chain. The loop regions between the β-strands and α-helices at the top of the barrel are more extended than those at the bottom (*MOLSCRIPT*, Kraulis, 1991).

Table 5. *Positions, length and mean φ, ψ angles of helices and β-strands (the very short antiparallel β-sheet in the $\alpha 3\beta 4$ loop is not included)*

Name	Position	Length	Mean φ, ψ angles
$\beta 1$	4–9	6	−115, 134
$\beta 2$	32–37	6	−101, 120
β'	38–44	7	−115, 143
β''	50–56	7	−95, 137*
$\alpha 2$	65–74	10	−58, −40
$\beta 3$	79–86	8	−108, 137†
$\alpha 3$	101–115	15	−58, −42
βh‡	117–122	4	—
$\beta 4$	126–131	6	−111, 105
$\alpha 4$	140–153	14	−63, −42
$\beta 5$	161–164	4	−125, 144
$\alpha 5$	171–180	10	−63, −40
$\beta 6$	187–191	5	−84, 115
$\alpha 6$	202–215	14	−64, −39
$\beta 7$	221–226	6	−108, 138
α'	229–234	6	−62, −36
$\alpha 7$	239–251	13	−62, 40
$\beta 8$	257–261	3	−136, 135
α''	263–266	4	−71, −28
$\alpha 8$	278–288	11	−63, −46

* Omitting residues 51/52 (bulge).

† Omitting residues 85/86.

‡ βh is a β-turn type I extended to an antiparallel β-sheet by two residues on each side.

Pfeffer, Betzel, Höhne & Wilson, 1992). The C-terminal entrance of the barrel is of special interest, because all α/β-barrel structures known to date have their active site situated in this region of the structure. The entrance shows a pocket formed by the loop regions between the β-strands of the barrel and the following α-helices. This top region of the barrel is subdivided by the side chain of Trp261 into two parts, a more hydrophilic cavity filled by eight solvent molecules and a hydrophobic pocket pointing towards the antiparallel β-sheet. A schematic drawing of the conformation of the loops and the most striking side chains making up the C-terminal entrance of the barrel is shown in Fig. 13. The cavity is closed by a salt bridge formed by Arg87, Asp231 and Glu132. These side chains are very well defined in electron density, Fig. 6. Three β/α-loop regions are held in their observed conformations, $\beta 3\alpha 3$ by Arg87, $\beta 4\alpha 4$ by Glu134 and α', which is in the loop $\beta 7\alpha 7$, by Asp231. The salt bridge is accessible to solvent and makes up the bottom of a flat groove on the surface of the molecule. The surface is closed completely and there is no channel between the solution and the internal cavity. On the other hand, in the derivative structures there is an Hg or Pt atom (at almost the same position) located in the cavity. The mercury derivative structure was refined to an R factor of 17.4% for all reflections between 2.1 and 10 Å resolution, to allow a detailed analysis of its binding sites. An Hg atom binds to His130 (the other three mercury-binding sites are Cys246, His133 and His173) without major distortion of the surrounding structure. Only the side chain of Arg87 moves significantly, but without breaking the salt bridge (Fig. 14). The hydrophobic pocket is formed mainly by side chains of the C-terminal end of $\beta 1$ (Tyr8 and Val11) and the beginning of β' (Phe39 and Ile41). On the top of the loops there are two hydrophobic residues, Val89 from the $\beta 3\alpha 3$ loop and Leu230 from the α'-helix in the $\beta 7\alpha 7$ loop [Fig. 13(*b*)]. The side chains are completely exposed to solvent and reject all water molecules in a sphere of 3.5 Å.

4.5. *α/β-loop regions*

Many α/β-loops of different TIM-barrel proteins can be classified as type $\alpha\beta 1$ or $\alpha\beta 3$ (Edwards, Sternberg & Thornton, 1987; Scheerlinck, Lasters, Claessens, Maeyer, Pio, Delhaise & Wodak, 1992). The loops can be identified by a positive φ angle at position L1. The energetically unfavourable positive φ angle is frequently formed by a glycine residue. In narbonin there is one loop with a glycine residue (Gly121). The loop connects

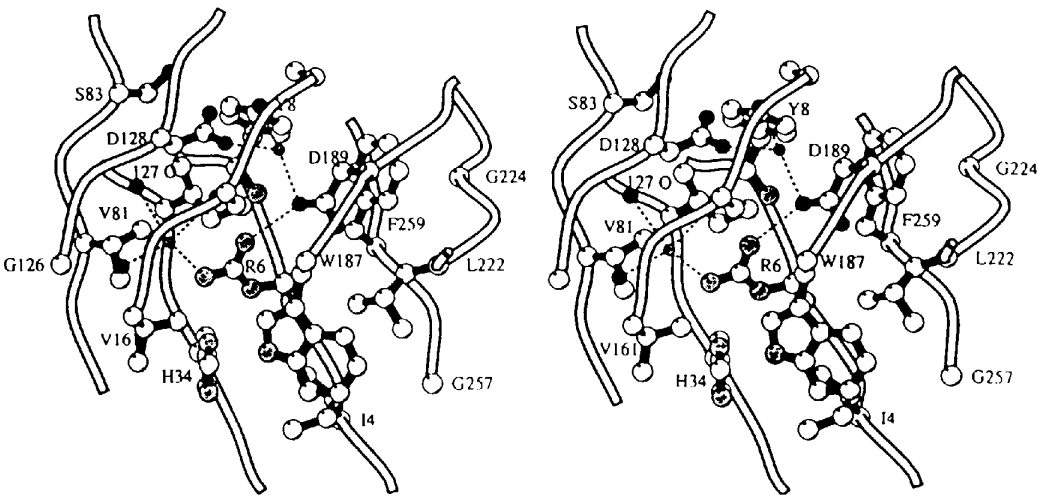

Fig. 12. Schematic drawing of the conformation of the β-barrel. The main-chain trace of the eight parallel β-strands and the packing of the side chains in the interior of the barrel is shown. The coordination of the internal water molecules Wat8 and Wat49 (the upper one) and the salt bridge of Arg6 and Asp189 are marked by dashed lines (*MOLSCRIPT*, Kraulis, 1991).

$\alpha 3$ and $\beta 4$ and is of the type $\alpha\beta 3$. However, Gly121 is involved in a hairpin structure (Fig. 15). In loop $\alpha 6\beta 7$ and $\alpha 7\beta 8$ the L1 position is taken by His219 and Phe253, respectively. These residues can be identified in the Ramachandran plot (Fig. 5) as α_L residues. The φ, ψ angles of the loop residues are listed in Table 6.

4.6. *Salt bridges*

The side-chain–side-chain interactions include 19 intramolecular salt bridges. All intramolecular and the five intermolecular salt bridges are listed in Table 7. Many of the salt bridges are found in α-helices and loops between the secondary-structural elements. Most of the β/α-loops and the additional secondary structures are stabilized by at least one salt bridge. Only Arg6 and Asp189 form a salt bridge between β-strands and compensate the charges in the interior of the parallel β-barrel (Fig. 12).

4.7. *Solvent structure*

The model includes 513 solvent molecules. The occupancy of the molecules was fixed at 1.0 and molecules

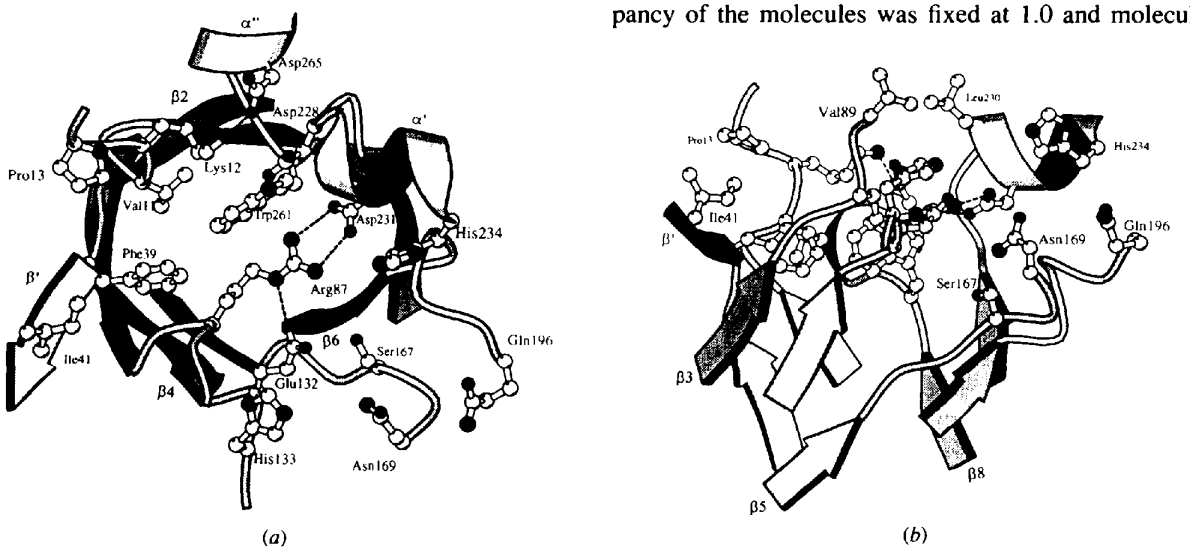

Fig. 13. Drawings of the C-terminal entrance of the barrel; (*a*) shows the view along the barrel axis and (*b*) the side view rotated by 90°. For clarity, not all the side chains are repeated here (*MOLSCRIPT*; Kraulis, 1991).

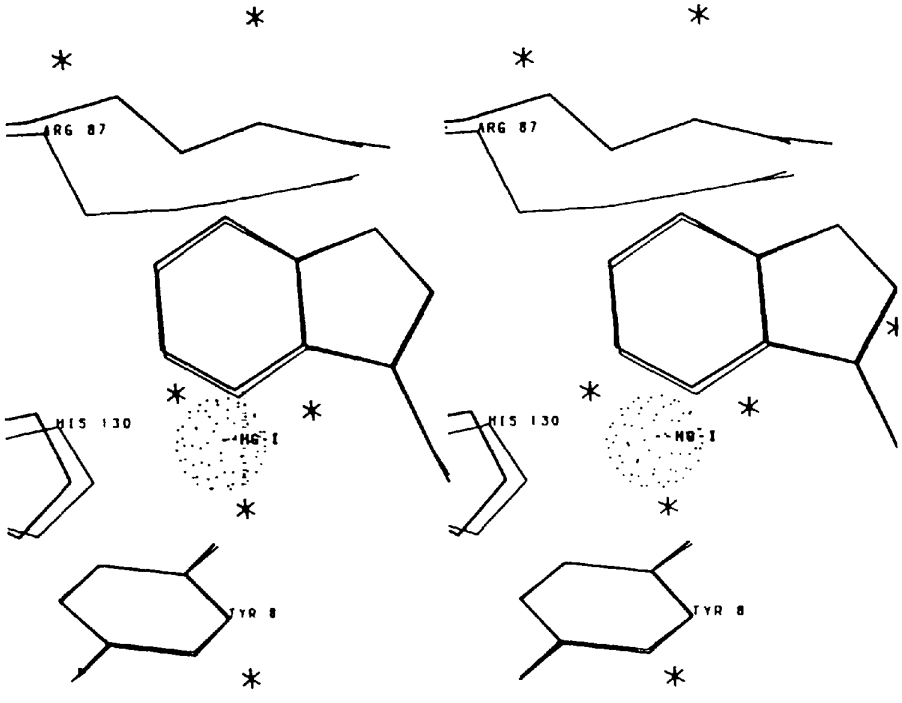

Fig. 14. Superposition of the C-terminal entrance of the barrel without metal binding in heavy lines and with a mercury ion bound on His130 in light lines. The dotted sphere indicates the position of the mercury ion and * indicates conserved water molecules.

Table 6. *Position and backbone φ, ψ dihedral angles (°) of residues forming $\alpha\beta1$ and $\alpha\beta3$ loops*

Position in the structure	Type	Residue L_1 φ, ψ	Residue L_{1+1} φ, ψ	Residue L_{1+2} φ, ψ	Residue L_{1+3} φ, ψ
$\alpha3\beta4$	$\alpha\beta3$	Gly121	Asn122	Leu123	Ile124
		121, −175	−75, 142	−99, −4	−95, 129
$\alpha6\beta7$	$\alpha\beta1$	His219	Lys220	Val221	Leu222
		68, 8	−111, −6	−96, 128	−111, 121
$\alpha7\beta8$	$\alpha\beta3$	Phe253	Ser254	Leu255	Pro256
		53, 45	−115, 6	−106, 111	−75, −17

Table 7. *Intramolecular and intermolecular salt bridges in narbonin*

All salt bridges with distances less than 3.5 Å between the N and the nearest O atom are included. Histidine is assumed to carry a positive charge if it is hydrogen bonded to a negatively charged residue.

Positively charged	Position in the structure	Negatively charged	Position in the structure	Length (Å)
Arg6 NH1	$\beta1$	Asp189 OD1	$\beta6$	2.92
Lys12 NZ	$1\beta2$*	Asp228 OD1	$\beta7\alpha'$	3.37
Lys12 NZ	$\beta1\beta2$*	Asp265 OD1	α''	2.61
His34 NE2	$\beta2$	Glu32 OE1	$\beta2$	2.60
Lys69 NZ	$\alpha2$	Asp118 OD2	$\alpha3\beta4$	3.30
Arg73 NH2	$\alpha2$	Asp118 OD1	$\alpha3\beta4$	3.47
Arg87 NH1	$\beta3\alpha3$	Asp231 OD1	α'	3.02
Arg87 NH2	$\beta3\alpha3$	Asp231 OD2	α'	3.12
Arg87 NE	$\beta3\alpha3$	Glu132 OE2	$\beta4\alpha4$	2.84
Lys110 NZ	$\alpha3$	Asp154 OD1	$\alpha4$	3.14
	$\alpha3$	Asp154 OD2	$\alpha4$	3.03
Lys115 NZ	$\alpha3$	Glu 56 OE2	β''	2.94
Arg135 NH1	$\beta5\alpha5$	Asp94 OD1	$\alpha3\alpha3$	3.37
Lys176 NZ	$\alpha5$	Glu138 OE2	$\beta4\alpha4$	2.90
Arg239 NH1	$\alpha7$	Asp274 OD1	$\alpha''\alpha8$	2.90
Arg239 NH2	$\alpha7$	Asp274 OD2	$\alpha''\alpha8$	3.25
Arg248 NH1	$\alpha7$	Asp203 OD2	$\alpha6$	2.89
Arg248 NH2	$\alpha7$	Asp203 OD2	$\alpha6$	2.69
	$\alpha7$	Asp156 OD1	$\alpha4\beta5$	3.09
Intermolecular salt bridges				
Lys49 NZ	$\beta'\beta''$	Glu77 OE2	$\alpha2\beta3$	2.74
Arg239 NH2	$\alpha7$	Glu97 OE2	$\beta3\alpha3$	2.76
Arg239 NE	$\alpha7$	Glu97 OE1	$\beta3\alpha3$	2.66
Arg271 NH1	$\alpha''\alpha8$	Glu97 OE2	$\beta3\alpha3$	3.02
Lys275 NZ	$\alpha''\alpha8$	Glu107 OE2	$\alpha3$	3.48

* This is in the long loop which replaces the first helix.

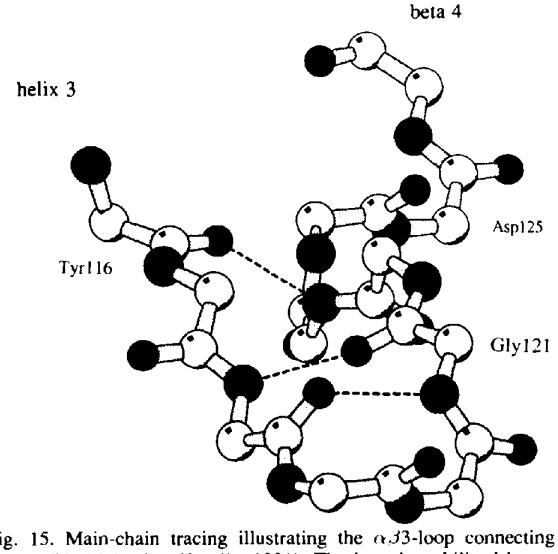

Fig. 15. Main-chain tracing illustrating the $\alpha\beta3$-loop connecting $\alpha3$ and $\beta4$ (Molscript; Kraulis, 1991). The loop is stabilized by crystal contacts and has well defined electron density. The β-turn extends to a hairpin (βh). The L_1 position of the $\alpha\beta3$-loop is Gly121; φ, ψ angles, see Table 6. The labels of the residues are close to their Cα positions.

with B higher than 80 Å^2 were rejected from the model. 360 solvent molecules have at least one hydrogen bond to the protein molecule and another 112 have hydrogen bonds to other water molecules. Only 41 molecules lie outside the distance cut-off 3.5 Å used for hydrogen bonds. The temperature factors vary between 8.0 Å^2 for the internal water molecule Wat8, situated in the β-barrel, and 77.1 Å^2 for Wat470. The average B factor of all waters is 47 Å^2 compared with 15 Å^2 for all protein atoms and 21 Å^2 for the complete structure. The number of solvent molecules in the B-factor ranges is shown in Table 8.

No features were observed in the electron-density maps which could be interpreted as buffer, metal or phosphate ions. In many β/α-barrel structures, a phosphate-binding site is formed by residues close to the N-terminus of the additional α-helix in the $\beta8\alpha8$ loop (Wilmanns, Hyde, Davies, Kirschner & Jansonius, 1991). In narbonin there are two additional helices, the α'-helix with six residues and α'', a 3_{10}-helix, which are situated between loops $\beta7\alpha7$ and $\beta8\alpha8$, respectively. The N-termini of both helices are close together. In contrast to β/α-barrel proteins having a

Table 8. *Number of solvent molecules in each B-factor range*

B factor (Å^2)	Number of solvent molecules
8–10	6
10–20	27
20–30	55
30–40	77
40–50	95
50–60	115
60–70	206
70–80	33

phosphate-binding site, there are significant differences. In narbonin there are several negatively charged residues in this region and there are no glycines comparable to those found in several other β/α-barrel proteins after the β-strands. Asp228 of α' *via* a water molecule and Asp265 of α'' directly form salt bridges with Lys12 of the first β/α-loop (note that the first helix is missing in the β/α-barrel topology of narbonin, Fig. 13).

4.8. *Crystal packing*

Narbonin has an accessible surface area of 13 200 Å^2 (*DSSP*; Kabsch & Sander, 1983). Most of the surface is covered by hydrophilic residues and many hydrogen bonds are formed by these residues with the surrounding solvent. The packing in the crystal is very compact and there are many interactions between neighbouring molecules. The packing involves layers of molecules with very close contacts and fewer contacts between the layers. The interaction of the molecules is mainly formed by loop regions. Subsequently, the loops $\beta'\beta''(\beta2\alpha2)$, $\beta4\alpha4$, $\beta6\alpha6$, $\beta7\alpha7$ and $\alpha'\alpha8$ of one molecule are in contact with the loops before $\beta2$, $\alpha2\beta3$, $\alpha3\beta4$, $\alpha4\beta5$ and $\alpha7\beta7$ of the adjacent molecule. Therefore, the C-terminus of the β-barrel lies on a solvent channel and is accessible to solvent. At least 17 intermolecular hydrogen bonds can be observed, including five salt bridges (Table 7). They stabilize extended loop regions of the molecule. The loop containing Arg271 and Lys275 fixes α-helix 3 of the neighbouring molecule at two positions like a claw. In addition to the hydrophilic residues, there are several hydrophobic residues accessible to the solvent. Some of these are involved in intermolecular interactions. Phe253 forms a hydrophobic cluster with Ile268 and Val267 of another molecule. Ile243 is in contact with the main chain of the residues Asn99 and Val100, and Pro76 is close to the 'back' of Ser47 of the next molecule. Nevertheless, there are several positions in which hydrophobic side chains are exposed to the solvent.

5. Discussion

The refined structure of narbonin from *Vicia narbonensis L* represents the first β/α-barrel structure for which there is no known enzymatic function. Only some eye lens proteins, crystallins, have additional functions in eye lenses, independent from their enzymatic activity (Wistow & Piatigorsky, 1987). Narbonin has been structurally classified into family C in the Farber and Petsko nomenclature (Farber & Petsko, 1990; Farber, 1993). Other members of this family with known three-dimensional structure are TIM, IGPS, PRAI and TS. All these enzymes and narbonin consist of a single domain, have the major axis of the barrel ellipse close to strand $\beta3$ and have a small helix between $\beta8$ and $\alpha8$. In all four enzymes, the small helix is responsible for the phosphate binding (Wilmanns, Hyde, Davies, Kirschner & Jansonius, 1991), but in narbonin there is no evidence for such a binding site. There is an additional α-helix between $\beta7$ and $\alpha7$, which has not yet been observed for other β/α-barrel proteins.

The lack of the first helix and the additional antiparallel β-strand are deviations from other β/α-barrel proteins. In mandelate racemase (Neidhart, Howell, Petsko, Powers, Li, Kenyon & Gerlt, 1991) and muconate cycloisomerase (Goldman, Ollis & Steitz, 1987) the $\alpha8$-helix is missing. A further deviation from all other α/β-barrel proteins is the extended antiparallel β-sheet which follows $\beta2$ almost without interruption and is situated parallel to $\alpha3$ and $\alpha4$. The β-turn which connects the antiparallel strands points into the solution and is stabilized by crystal contacts. Phe39 and Ile41 of β' contribute to the formation of the hydrophobic pocket on the C-terminus of the barrel and, therefore, may contribute to the formation of the prospective active site of narbonin.

The most striking question is whether narbonin has an enzymatic function and what it could be. All attempts to identify an activity have failed to date. From the structural point of view, the salt bridge formed by the three residues Arg87, Asp231 and Glu132 may protect the prospective active site from being active, but salt bridges are also found in structurally and functionally interesting regions in other α/β-barrel proteins.

For instance, in TIM (Lolis, Alber, Davenport, Rose, Hartmann & Petsko, 1990; Noble, Zeelen, Mainfroid, Goraj, Gohimont, Martial & Wierenga, 1993) residues Lys12 (in TIM from yeast, Lys13 in TIM from *E. coli*) and Glu97 (Glu99 in TIM from *E. coli*) form a salt bridge in a similar way to that observed in narbonin by the residues Lys12 and Asp265 directly, and Asp228 *via* a water molecule. In both structures, the negatively charged residue is provided from an additional α-helix. In narbonin Asp265 is from α'' between $\beta8$ and $\alpha8$, Asp228 from α' between $\beta7$ and $\alpha7$. In TIM the Glu is from the helix between $\alpha4$ and $\beta4$. However, despite the fact that the negatively charged residues are from different loops, both salt bridges connect loops on the top of the barrel in a very similar way. The alignment of the sequences in this region at least does not show further identities. In summary, no structural features can be described which prevent narbonin from exercising a binding or catalytic function, but we cannot

identify striking structural similarities with other TIM-like structures in their active site.

A comparison of narbonin with closely related proteins is impossible because narbonin is the first representative of the 2*S* globulins for which a crystal structure and amino-acid sequence is known. Biochemical characterization, sequencing and crystal structural analysis based on X-ray data to 1.9 Å resolution of a narbonin-like protein from *Vicia pannonica* is in progress (Schlesier, Hennig, Kraft & Horstmann, 1993). The narbonin from *Vicia pannonica* has a similar molecular weight, N-terminal sequence and crystallizes in the same space group, but deviations in peptide mapping suggest differences in the amino-acid sequence.

Concanavalin B from *Canavalia ensiformis* is classified as a 2*S* globulin like narbonin (Ko, Ng & McPherson, 1993). Preliminary structural results suggest concanavalin B has an α/β-barrel topology similar to narbonin, despite which there is no significant sequence similarity in the known parts of the sequence (concanavalin B has not been fully sequenced; Ko, Ng & McPherson, personal communication). Concanavalin B is expected to bind NADPH (Morrison, DeLozier, Robinson & McPherson, 1984). Concanavalin B (Hennig & Schlesier, unpublished results), narbonin from *Vicia pannonica* and *Vicia narbonensis* can be purified in a similar way and crystallized by pH shift at low ionic strength, giving crystals suitable for structural analysis at atomic resolution (Hennig & Schlesier, 1994). Detailed comparison of the three structures may help in the elucidation of the functional properties of these seed proteins.*

The high-resolution X-ray structure of narbonin may serve for a better understanding of molecular processes. Subsequently, it can be used as a model for the investigation of biosynthesis, intercellular transport, processing, deposition and degradation of seed proteins. The use of narbonin for improvement of nutritional and technological quality of seeds by genetic engeneering (Lumen, 1990; Sun & Larkins, 1993) is limited, because it is accumulated in smaller amounts compared with vicilin and legumin. The amino-acid composition and the time of its biosynthesis and degradation support the classification as a storage protein, but its three-dimensional molecular structure arose suspicion to it having biological activity.

We thank Dr W. E. Höhne, Humboldt University, Berlin, for support. We gratefully acknowledge Professors J. N. Jansonius and K. Kirschner, Biozentrum University of Basel, for interest and advice.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1NAR, R1NARSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: GR0370). A list of deposited data is given at the end of this issue. The sequence has also been deposited with the EMBL sequence library (Reference: Z25536).

References

BANNER, D. W., BLOOMER, A. C., PETSKO, G. A., PHILLIPS, D. C., POGSON, C. I., WILSON, I. A., CORRAN, P. H., FURTH, P. H., MILMAN, A. J., OFFORD, R. E., PRIDDLE, J. D. & WALEY, S. G. (1975). *Nature*, **255**, 609–614.

BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832–847.

COLLABORATIVE COMPUTATIONAL PROJECT, NUMBER 4 (1994). *Acta Cryst.* D**50**, 760–763.

COLMAN, P. M., JANSONIUS, J. N. & MATTHEWS, B. W. (1972). *J. Mol. Biol.* **70**, 701–724.

DERBYSHIRE, E., WRIGHT, D. J. & BOULTER, D. (1976). *Phytochemistry*, **15**, 3–24.

EDWARDS, M. S., STERNBERG, M. J. E. & THORNTON, J. M. (1987). *Protein Eng.* **1**, 173–181.

FARBER, G. K. (1993). *Curr. Opin. Struct. Biol.* **3**, 409–412.

FARBER, G. K. & PETSKO, G. A. (1990). *Trends Biol. Sci.* **15**, 228–234.

GOLDMAN, A., OLLIS, D. L. & STEITZ, T. A. (1987). *J. Mol. Biol.* **194**, 143–153.

HENDRICKSON, W. A. & KONNERT, J. H. (1980). In *Computing in Crystallography*, pp. 13.01–13.23.

HENNIG, M. & SCHLESIER, B. (1993). *Food Proteins, Structure and Functionality*, edited by H. K. D. SCHWENKE & H. R. MOTHES, pp. 109–117. Weinheim: VCH Publishers.

HENNIG, M. & SCHLESIER, B. (1994). *Acta Cryst.* D**50**, 627–631.

HENNIG, M., SCHLESIER, B., DAUTER, Z., PFEFFER, S., BETZEL, C., HÖHNE, W. E. & WILSON, K. S. (1992). *FEBS Lett.* **306**, 80–84.

HENNIG, M., SCHLESIER, B., PFEFFER, S. & HÖHNE, W. E. (1990). *J. Mol. Biol.* **215**, 339–340.

HERRIOTT, J. R., WATENPAUGH, K. D., SIEKER, L. C. & JENSEN, L. H. (1973). *J. Mol. Biol.* **80**, 423–432.

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

KABSCH, W. & SANDER, CH. (1983). *Biopolymers*, **22**, 2577–2637.

KO, T.-P., NG, J. D. & MCPHERSON, A. (1993). *Plant Physiol.* **101**, 729–744.

KRAULIS, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

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

LAWRENCE, M. C., SUZUKI, E., VARGHESE, J. N., DAVIS, P. C., VAN DONKELAAR, A., TULLOCH, P. A. & COLMAN P. M. (1990). *EMBO J.* **9**, 9–15.

LESLIE, A. G. W., BRICK, P. & WONACOTT, A. J. (1986). *CCP4 Newslett.* **18**, 33–39.

LOLIS, E., ALBER, T., DAVENPORT, R. C., ROSE, D., HARTMANN, F. C. & PETSKO, G. A. (1990). *Biochemistry*, **29**, 6609–6618.

LUMEN, B. D. (1990). *J. Agric. Food Chem.* **38**, 1779–1788.

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

MCILVAINE, T. C. (1921). *J. Biol. Chem.* **49**, 183–186.

MORRISON, R., DELOZIER, G., ROBINSON, L. & MCPHERSON, A. (1984). *Plant Physiol.* **76**, 175–183.

NEIDHART, D. J., HOWELL, P. L., PETSKO, G. A., POWERS, V. M., LI, R., KENYON, G. L. & GERLT, J. A. (1991). *Biochemistry*, **30**, 9264–9273.

NOBLE, M. E. M., ZEELEN, J. PH., MAINFROID, V., GORAJ, K., GOHIMONT, A.-C., MARTIAL, J. A. & WIERENGA, R. K. (1993). *Acta Cryst.* D**49**, 403–417.

RAMACHANDRAN, G. N. & SASISEKHARAN, V. (1968). *Adv. Protein Chem.* **28**, 283–437.

SCHEERLINCK, J. P., LASTERS, I., CLAESSENS, M., MAEYER, M., PIO, F., DELHAISE, P. & WODAK, S. J. (1992). *Proteins*, **12**, 299–313.

SCHLESIER, B., HENNIG, M., KRAFT, R. & HORSTMANN, CH. (1993). *Food Proteins, Structure and Functionality*, edited by K. D. SCHWENKE & R. MOTHES, pp. 118–120. Weinheim: VCH Publishers.

SCHLESIER, B., MANTEUFFEL, R., RUDOLPH, A. & BEHLKE, J. (1978). *Biochem. Physiol. Pflanzen*, **173**, 420–428.

SCHLESIER, B. & SCHOLZ, G. (1974). *Biochem. Physiol. Pflanzen*, **166**, 367–369.

SUN, S. & LARKINS, B. (1993). In *Transgenic Plants*, edited by S. KUNG & R. WU, pp. 339–368. San Diego: Academic Press.

WANG, B. C. (1985). *Methods Enzymol.* **115**, 90–112.

WILMANNS, M., HYDE, C. C., DAVIES, D. R., KIRSCHNER, K. & JANSONIUS, J. N. (1991). *Biochemistry*, **30**, 9161–9169.

WISTOW, G. & PIATIGORSKY, J. (1987). *Science*, **236**, 1554–1556.