Volume 306, number 1, 80–84 FEBS 11235 July 1992

A TIM barrel protein without enzymatic activity?

Crystal-structure of narbonin at 1.8 Å resolution

Michael Hennig[a,b], Bernhard Schlesier[c], Zbigniew Dauter[a], Sabine Pfeffer[a], Christian Betzel[a], Wolfgang E. Höhne[b] and Keith S. Wilson[a]

[a]*European Molecular Biology Laboratory (EMBL) c/o DESY, Notkestrasse 85, W-2000 Hamburg 52, Germany,* [b]*Institut für Biochemie der Humboldt-Universität zu Berlin, Hessische Str. 3-4, O-1040 Berlin, Germany* and [c]*Institut für Pflanzengenetik und Kulturpflanzenforschung, Correnstrasse 3, O-4325 Gatersleben, Germany*

Received 21 April 1992; revised version received 20 May 1992

The major protein component in seeds is storage protein. These have no known enzymatic activity and act to provide amino acids as a source of metabolites in the developing seedling. We report here the first three dimensional crystal structure of a seed storage globulin at high resolution. The molecule of the 2S globulin, narbonin, from *Vicia narbonensis* L., consists of an eight-stranded parallel α/β barrel structure similar to that observed in triose phosphate isomerase (TIM). Narbonin is the first protein with this topology possessing no known enzymatic activity. Because of the lack of sequence information most of the primary structure was determined directly from the electron density.

Seed storage protein; *Vicia narbonensis*; X-ray structure; TIM barrel; Primary structure

1. INTRODUCTION

Seeds of legumes and their byproducts make up a major part of the human diet, but legumes are uniformly limited in methionine and cysteine because their major storage proteins, the globulins, are low in these amino acids [1].

Storage globulins from legumes are generally classified with respect to their molecular mass as 2S, 7S (vicilin) or 11S (legumin) [2]. The vicilin globulins have a trimeric organization of single polypeptide chains. The three dimensional structure of the 7S globulin phaseolin has been determined at medium (3 Å) resolution [3]. The 11S or legumin globulins have six subunits. Each monomer consists of two different disulphide-linked polypeptide chains. The hexamers contain a random combination of different monomers as described in [2].

Beside these differences the storage proteins from legumes have several common properties. The proteins are synthesized during the development of the seeds and laid down in protein bodies for degradation on germination. Although there are significant differences in the amino acid composition between narbonin and the other globulins, they all contain a high proportion of glutamic, aspartic and basic amino acids; sulphur containing amino acids are rare [2]. The strong salting-in property of the proteins, which seems to be common for storage globulins of legumes, is used in their purification. To understand the synthesis, transport and proteolytic processing of these proteins and to allow the design of genetic engineering strategies, detailed structural information is required.

The 2S globulin narbonin is a monomer with a single polypeptide chain and a molecular mass of about 33,000 [4,5]. Crystals of narbonin [6] have been used to determine its three dimensional structure at atomic resolution and to determine the primary structure during the crystallographic refinement.

2. MATERIALS AND METHODS

Narbonin was isolated as described previously [4,5]. The monoclinic crystals are in space group $P2_1$ with cell dimensions a = 46.9 Å, b = 75.5 Å, c = 50.9 Å and β = 120.5°. The asymmetric unit contains one protein molecule and 41% solvent by volume [6]. All X-ray data were collected on an imaging plate scanner at the EMBL outstation at DESY with synchrotron radiation or a sealed tube generator. The mercury acetate derivative produced very clear anomalous and isomorphous difference Patterson maps. Three binding sites were located from the Harker section. The remaining site of this derivative and all sites of the other derivatives were found by cross-phased difference Fourier syntheses (Table I). The electron density map at 2.2 Å resolution was calculated with phases from multiple isomorphous replacement (MIR) slightly improved by solvent flattening [7] (mean figure of merit, 0.71). Because of the lack of sequence information only alanine or glycine residues were built into well-defined parts of the electron density. The initial model contained 55% of the protein atoms and was refined using the program PROLSQ (CCP4 package [8]) to an R factor of 46% for all data between 2.2 and 10 Å. After 9 steps

Coordinates of narbonin will be deposited in the Brookhaven Protein Data Bank.

Correspondence address: M. Hennig, European Molecular Biology Laboratory (EMBL), Notkestrasse 85, W-2000 Hamburg 52, Germany. Fax: (49) (40) 89 08 01 49.

Published by Elsevier Science Publishers B.V.

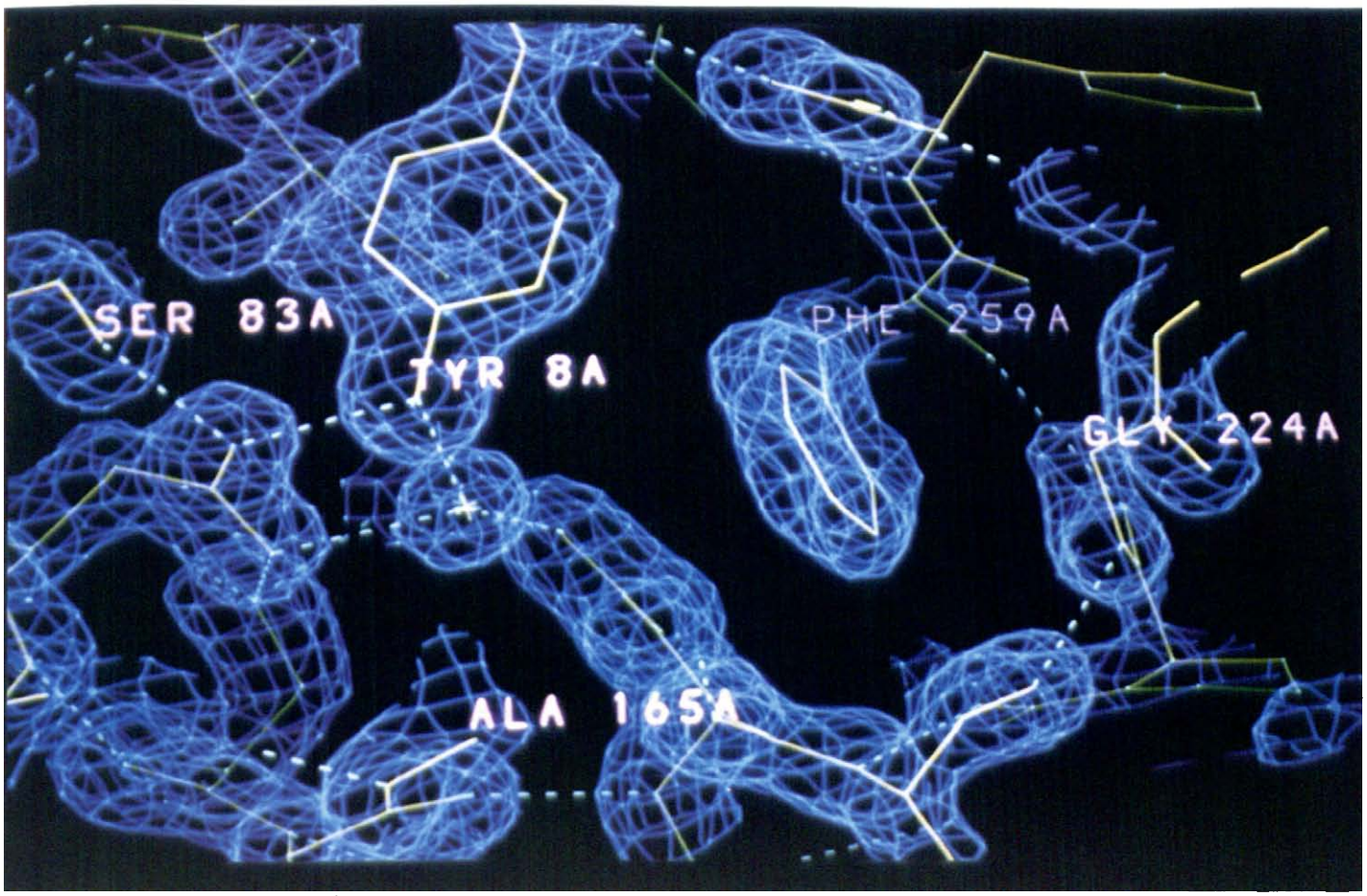

Fig. 1. A view looking down into the hydrophobic part of the top layer of the β-barrel. The electron density is contoured at the 1σ level of the final ($2F_o - F_c$) synthesis, with the final model superimposed.

of model building, refinement and phase combination of the MIR and model phases the R factor dropped to 30.1% (for all data between 1.8 to 10.0 Å with a model including 85% of the protein atoms). At this stage of the refinement it was possible to identify the N-terminal and two further parts of the polypeptide chain, for which the amino acid sequence were chemically determined (about 17% of the total sequence). A new phase refinement [9] procedure considerably improved the electron density and most of the side chains could be located unambiguously. The current model contains 288 amino acid residues and 242 water molecules, has an R factor of 16.5% and r.m.s. (1–2) bond length deviations of 0.011 Å. A representative region of the final electron density map is shown in Fig. 1.

3. RESULTS AND DISCUSSION

The narbonin molecule is very compact and has dimensions of about 45×45×30 Å. There is no evidence for metal binding sites or disulphide bridges from the electron density and biochemical studies. Narbonin exhibits a single domain folded in an α/β barrel, similar to that found in triose phosphate isomerase (TIM) [10,11] and many other enzymes. The main feature of the secondary structure is an eight stranded parallel β-barrel making up the core of the molecule with α- helices packed

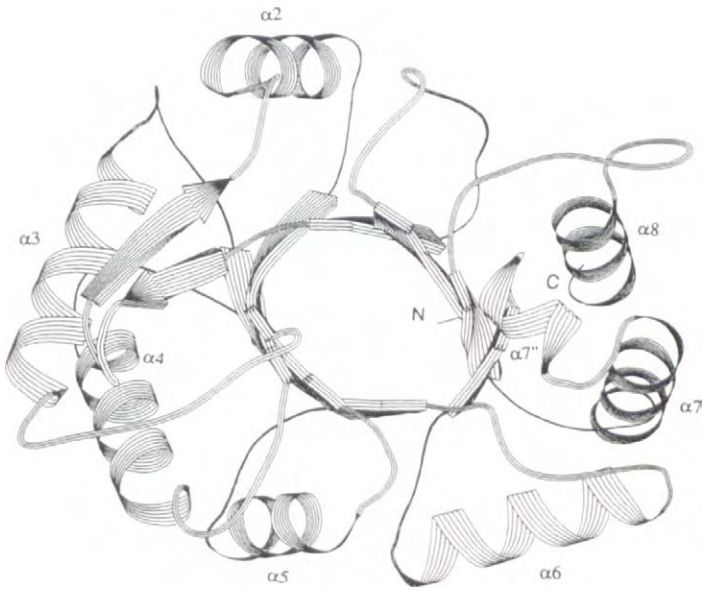

Fig. 2. Ribbon plot [10] illustrating the overall fold of narbonin. The view is from the top of the barrel. The external α-helices are numbered sequentially from the N-terminus and the first TIM-helix is replaced by a loop.

Table I

Data collection and analysis

Data set	Native	$Hg(CH_3COO)_2$	$UO_2(NO_3)_2$	$SmCl_3$	K_2PtCl_4
X-ray source	X11	X31	X12	sealed tube	sealed tube
Wavelength used (Å)	0.96	0.95	0.96	1.54	1.54
No. of measured reflections	124,224	39,570	5,530	87,122	26,307
No. of unique reflections	27,227	14,975	1,782	12,627	10,336
Completeness (%)	95.5	97.4	67.5	93.1	68.5
Maximum resolution (Å)	1.8	2.2	4.0	2.3	2.2
R_{sym} § (%)	5.6	5.3	4.8	4.3	5.2
R_{isom} & (%)	–	26.6	30.4	12.8	23.8
K_{emp} #	–	6.3	–	4.4	4.3
R_c $ (%)	–	29.2	38.4	63.7	51.1
No. of heavy atom sites	–	4	5	1	3
Phasing power*	–	5.1	2.0	0.9	2.5

§ $R_{sym} = \Sigma_i |I_i - \langle I \rangle| / \Sigma \langle I \rangle$; where $\langle I \rangle$ is the mean of the intensity measurements I_i. & $R_{isom} = |\Sigma (F_{PH} - F_P)| / \Sigma F_P$; where F_{PH} is the derivative and F_P the native structure factor amplitude. # $K_{emp} = \Sigma |F_{PH} - F_P| / \Sigma |F_{PH+} - F_{PH-}|$. $ $R_c = \Sigma | F_{PHcalc} - F_{PHobs} | / \Sigma F_{PHobs}$; the Cullis R factor for centric reflections. *The phasing power defined as F_H/E; where F_H is the derivative structure factor amplitude and E the r.m.s. residual lack of closure.

around it (Fig. 2). There are overall ten β-strands and eight helices in the molecule. Only in the N-terminus of the molecule are there deviations from the strict α-β-α folding pattern characteristic of other TIM-like structures. The N-terminal region of the chain begins with four β-strands where only two (the first and third) are involved in the barrel structure. What would be the first helix in TIM is replaced by a loop composed of two β-turns and there are two extra strands (numbers three and four) forming an additional β-sheet motif compared to the common folding pattern of TIM-barrel structures. The presence of only seven α-helices in the nar-

Table II

Primary stucture obtained from the electron density

```
    * * * * * * * * * * * * * 9 9 9 9 9 5 8 9 9 8 5 9 9 5 7 5 7
  1 P K P I F R E Y I G V K P B S T T L K B F P T A I I B T T T

    9 7 9 9 9 9 9 9 9 9 6 9 9 9 9 1 3 9 9 9 9 9 9 9 5 7 9 9 8 5
 31 L Z F H Y I L G F A I Z S Y Y S R G B G T G T F A E S W B T

    2 7 9 9 9 1 1 4 6 7 8 5 5 7 7 9 8 8 6 8 8 9 9 9 9 9 9 9 8 8
 61 Z L F G P A S V T B L K R S H P E T K V T I S I G G R G T B

    8 9 9 9 9 7 8 8 9 8 9 8 8 7 9 8 5 9 8 9 5 9 9 8 9 9 5 8 5 5
 91 T P F B P A E E B V W T T B A T Z S L K Z I I Z K Y Z K T S

    9 7 8 9 7 3 5 8 8 9 9 8 8 9 3 3 7 7 9 9 8 8 8 5 9 9 8 9 8 9
121 G B L I C G I B V H Y Z H I G T B E P F A T L Z G Z L I T Z

    8 7 6 8 5 6 8 6 9 8 8 8 9 9 9 9 9 9 4 8 9 9 7 9 8 8 5 9 6 6
151 L K A D B D L Z I B T V S I A P S Z S B S S B Y Z K L Y B A

    7 * * * * * * * * * * * * * * * * * * * * * * * 5 9 9 5
181 K K D Y I N W V D Y E F S N E E K P V S T D D A F V A I F K

    7 1 9 7 8 9 9 9 7 7 8 8 9 9 9 9 8 8 9 5 8 8 5 8 9 9 9 8 9 9
211 S Z Z A B Y H P H K V L P G F S T B P A D T A H B K I T R B

    8 9 9 9 9 9 8 9 * * * * * * * * * * 9 9 8 9 8 8 9 7 4 7 2
241 I F I G G C T R L V E T F S L P G V F F W B A B D S T S P K

    2 1 7 9 4 9 9 7 7 9 8 8 8 9 1 2 7 2
271 S G G D K P F T V Z L V L Z K I B A
```

The sequence is printed in the standard one-letter code. The number above each residue indicates our rough probability estimate of the certainty of correct identification for this residue, based on the temperature factor and subjective estimation of the quality of the electron density. 1 indicates substantial disorder in the electron density and the 9 indicates certainty. *Indicates those parts with biochemically determined amino acid sequence.

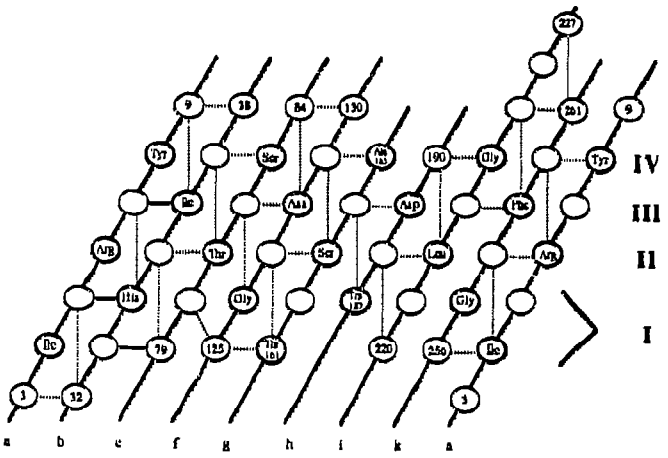

Fig. 3. Topology of the parallel β-sheets in narbonin forming the TIM barrel structure. Each circle represents a residue, broken lines represent hydrogen bonds. Residues forming the interior of the barrel are marked with the residue name. Note that nine strands are shown in such a way that the one at the edge is duplicated.

bonin structure is similar to the TIM-like muconate lactonizing enzyme and mandelate racemase [12]. The usual right handed α-β-α folding pattern follows and the C-terminus lies at the end of an α-helix.

The N- and C-termini are situated on the same side of the barrel, only 13.5 Å away from each other and the N-terminus is on the 'edge' of the barrel very close to the depression formed by the bottom of the barrel. The β-barrel itself is elliptical and has a cross section of 12 Å by 16 Å. The centre of the barrel is filled by side chains of the β-sheet residues. Fig. 3 shows the topology of the β-barrel of narbonin. There are at least four layers of residues. The pattern of residues is chessboard-like. Every second residue points to the centre of the barrel.

The first layer contains the side chains of two levels, Ile-4 and Thr-161 from the bottom level and His-34 and Trp-187 from the next. The space for the big aromatic side chains is compensated by a distortion of the circular shape of the barrel. The other three layers are properly formed by side chains from one level only except Phe-259, which belongs to the fourth layer and fills the cavity caused by Gly-224. There are two internal waters, one in the second layer forming hydrogen bonds to Arg-6 and Ser-163 and the other between the third and fourth layer.

Inspection of space-filling computer models indicates that there is no cavity in the centre of the barrel. Comparison with typical TIM barrel structures [10] shows that the depth of the barrel is greater with four complete layers compared to three in most of the other TIM barrel structures.

The preliminary amino acid sequence, determined directly from the electron density, is in keeping with the common features of the amino acid content of globulins in legume seeds (Table II). The structure contains only two cysteines and no methionine. Comparison of the amino acid composition derived from electron density with the measured amino acid content shows a reasonable agreement (Table III). A search of the sequence data bank showed no significant homology to any other proteins.

So far there is only one other three dimensional structure of a seed storage protein known from X-ray analysis. The trimeric phaseolin from *Phaseolus vulgaris*, a 7S storage protein, has two structurally similar units each with a folding topology similar to the viral coat proteins (antiparallel β-barrel) and the helix-turn-helix motif found in certain DNA-binding proteins [3]. Thus these two representatives of the 2S and 7S globulins, narbonin and phaseolin respectively, have distinctly different folds.

Structures are now known for more than 19 TIM barrel proteins, all of them are enzymes with different functions and non-homologous amino acid sequences [13]. The active sites are located in a pocket formed by the loop regions that connect the carboxy ends of the β-strands with the α-helices. Narbonin is the first member of this topology for which no enzymatic activity is known. Why does narbonin have this folding topology? Comparison of the narbonin structure to the TIM barrel enzymes gave no evidence for a suitable active site.

Table III

Amino acid composition derived from electron density compared with the measured amino acid content

Amino-acid	g amino acid/100 g protein		Number
	Chemically measured [5]	X-ray sequence	
Asx	13.6	12.8	36
Glx	11.8	11.2	28
Thr	4.8	9.5	30
Val	5.4	4.3	14
Ser	5.6	6.0	22
Pro	5.1	5.1	17
Gly	3.4	3.4	19
Ala	2.2	3.8	17
Ile	7.4	7.8	22
Leu	7.7	6.0	17
Tyr	5.3	5.6	11
Phe	7.1	7.8	17
Lys	9.0	7.6	19
His	3.8	3.0	7
Arg	4.0	2.9	6
Trp	1.8	2.3	4
Cys	3.2	0.6	2
Met	4.0	–	–
Σ	105.3	99.7	288

Compositions are in good agreement except that there are more Threonine residues perhaps arising from disordered Serine, Lysine and Arginine side chains. In addition the lack of the sulphur-containing amino acids Cysteine and Methionine can be seen.

On the other hand there is a pocket formed by the loop regions as well. The pocket is divided into a hydrophobic and a hydrophilic part by a tryptophan side chain. The hydrophilic part is filled by eight well-defined water molecules.

It is possible that during evolution narbonin lost its enzymatic function. But it is more likely that the TIM barrel conformation is a stable building block onto which different functionalities can be selected during divergent evolution or is a stable conformation onto which a number of independently evolved functions have converged [14].

Further biochemical and structural investigations, which are planned for the near future, should provide a better understanding of the role of storage proteins in seeds. Seeds of legumes are an important protein source for human beings and in animal feed, but their nutritional properties do not conform precisely with those demanded by the consumer [15–17]. Biotechnology provides a possibility whereby the properties of seeds might be tailored specifically through site-directed mutagenesis of their storage proteins to modify the storage protein genes to code for more of the nutritionally essential amino acids.

Acknowledgements: We thank Dr. V. Lamzin for his help with his automated refinement program. This study has been funded by the German Federal Ministry of Research and Technology (BMFT/DARA) under Contract No. 50QV90895.

REFERENCES

[1] Jackman, R.L. and Yada, R.Y. (1987) Food Biotechnol. 1, 167–223.
[2] Derbyshire, E., Wright, D.J. and Boulter, D. (1976) Phytochemistry 15, 3–24
[3] Lawrence, M.C. Suzuki, E., Varghese, J.N., Davis, P.C., Van Donkelaar, A., Tulloch, P.A. and Colman, P.M. (1990), EMBO J. 9, 9–15.
[4] Schlesier, B. and Scholz, G. (1974) Bioch. Physiol. Pflanzen 166, 367–369 .
[5] Schlesier, B., Manteuffel, R., Rudolph, A. and Behlke, J. (1978) Biochem. Physiol. Pflanzen 173, 420–428.
[6] Hennig, M., Schlesier, B., Pfeffer, S. and Höhne W.E. (1990) J. Mol. Biol. 215, 339–340.
[7] Wang, B.C. (1985) Methods Enzymol. 115, 90–112.
[8] CCP4, A Suite of Programs for Protein Crystallography (SERC Daresbury Laboratory, Warrington, UK, 1986).
[9] Lamzin, V.S. and Wilson, K.S. Acta Crystallogr. D (1993) in press.
[10] Lesk, A.M. , Bränden, C-I. and Chothia, C. (1989) Proteins 5, 139–148.
[11] 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. and Waley, S.G. (1975), Nature 255, 609–614.
[12] Neidhart D.J., Kenyon, G.L., Gerit, J.A. and Petsko G.A. (1990) Nature 347, 692–694
[13] Lebioda, L. and Stec, B. (1988) Nature 333, 683–685.
[14] Chothia, C. (1988) Nature 333, 598–599.
[15] Eason, P.J., Johnson, R.J. and Castleman, G.H. (1990) Aust. J. Agric. Res. 41, 565–571.
[16] Kim, S. (1990) Protein Eng. 3, 725–730.
[17] Lumen, B.D. (1990) J. Agric. Food Chem. 38, 1779–1788.