

Three-dimensional structure in solution of a wheat lipid-transfer protein from multidimensional ¹H-NMR data

A new folding for lipid carriers

Edith GINCEL¹, Jean-Pierre SIMORRE², Anita CAILLE¹, Didier MARION³, Marius PTAK^{1,4} and Françoise VOVELLE^{1,4}

¹ Centre de Biophysique Moléculaire (CNRS), Orléans, France

² Institut de Biologie Structurale, Grenoble, France

³ Laboratoire de Biochimie et Technologie des Protéines, INRA, Nantes, France

⁴ Université d'Orléans, Orléans, France

(Received July 13/September 5, 1994) – EJB 94 1039/3

Two-dimensional and three-dimensional ¹H-NMR experimental data [Simorre, J. P., Caille, A., Marion, D., Marion, D. & Ptak, M. (1991) *Biochemistry* 30, 11600–11608] were used to build models of the three-dimensional structure of a non-specific wheat lipid-transfer protein (LTP) by using distance geometry, simulated annealing, energy minimization and molecular dynamics techniques. A first set of 881 distance constraints derived from NOE cross-peak intensities was used to generate 74 initial structures. One family of topological mirror images of the protein structure was eliminated by considering helical secondary-structure organization and steric requirements. Back calculations of NOE intensities led us to introduce 535 additional distance constraints. Finally, 21 structures were selected as representative of the structure of the protein. The polypeptide backbone folds into a simple and original right-handed winding. It is composed of a bundle of four helices linked by flexible loops, which is packed against a C-terminal fragment forming a non-standard saxophone-like shape. The folded protein is stabilized by hydrophobic interactions and the four disulfide bridges combined by pairs on each side of the protein. An hydrophobic cleft, formed by residues located in the second half of the protein could be a potential site for the binding of lipids.

Plant non-specific lipid-transfer proteins (ns-LTP) facilitate *in vitro* exchange or transfer of different amphiphilic lipids (phospholipids, fatty acids, glycolipids) between natural and/or between artificial membranes (Kader, 1990), (considering the variety of lipids interacting with this family of transfer proteins, we use henceforth the generic term of lipid-transfer protein (LTP) instead of phospholipid-transfer protein (PLTP). These ubiquitous proteins have been identified in seeds (Campos and Richardson, 1984; Mundy and Rogers, 1986; Takishima et al., 1986; Tchang et al., 1988; Désormeaux et al., 1992; Terras et al., 1992), leaves (Bouillon et al., 1987; Sossountov et al., 1991; Molina et al., 1993), anthers (Koltunow et al., 1990), stems and shoot apex (Torres-Schuman et al., 1992; Fleming et al., 1992). Initially, their transfer activity has been suggested to play a major role in the intracellular traffic of lipids in plant cells. However, the discovery that these ns-LTP are exported through the endoplasmic reticulum pathway (Tchang et al., 1988; Bernhard et al., 1991) and that they are located outside the cell, probably

on the cell wall (Sossountzov et al., 1991; Sterk et al., 1991; Thoma et al., 1993) has led different alternative functions to be ascribed to this class of proteins. More specifically, it was suggested recently that ns-LTP could participate to the plant defence mechanisms against pathogens (Terras et al., 1992; Molina et al., 1993).

The elucidation of the three-dimensional (3D) structure of wheat ns-LTP reported in the present study should contribute to the clarification of the mechanism of lipid binding and lipid transfer. Recently, we demonstrated through NMR experiments (Petit, M., unpublished results) that both wheat and maize ns-LTP have the ability to form well-defined complexes with lipids in aqueous solution. The next step of our studies will be then to solve their structures.

The plant ns-LTP are small proteins of around 9 kDa, with a high isoelectric point (8.8–10.5) and a broad affinity for various hydrophobic molecules and for phospholipids since they are capable of transporting phosphatidylcholine (PtdCho), phosphatidylethanolamine (PtdEtn), phosphatidylinositol (PtdIns). Comparison of their amino acid sequences reveals a rather high degree of similarity, with a consensus sequence and a common pattern of eight cysteines C–C–CC–CXC–C–C forming four disulfide bridges (Désormeaux et al., 1992).

Two-dimensional (2D) and 3D ¹H-NMR studies have been recently carried out (Simorre et al., 1991) on a wheat ns-LTP isolated and purified according to a procedure described by Désormeaux et al. (1992). All the ¹H resonances were assigned sequentially for the 90 residues, IDCGHVDS-

Correspondence to F. Vovelle, Centre de Biophysique Moléculaire (Centre National de la Recherche Scientifique), 1A avenue de la Recherche Scientifique, F-45071 Orléans Cedex 2, France

Fax: +33 38 63 15 17.

Abbreviations. ns-LTP, non-specific lipid-transfer protein; 3D, three-dimensional; 2D, two-dimensional; NOESY, nuclear Overhauser spectroscopy; RMSD, root mean square deviation; DG, distance geometry; lyso-PtdCho, lysophosphatidylcholine; PtdEtn, phosphatidylethanolamine; PtdIns, phosphatidylinositol; FABP, fatty-acid-binding protein.

LVRPCLSYVQGGPGPSG QCCDGVKNLHNQARS QSD-RQSACNCLKGIARGIHNLNEDNARSIPPKCGVNLPLYT-ISLNIDCSR, and the secondary structures were identified. The NOE pattern reveals the existence of four helical regions involving approximately 5–18, 26–36, 41–57 and 63–69 residues and of an irregular and stable C-terminal region. In the present study, we describe the 3D structure derived from these NMR data which shows a particular folding involving the four helical fragments and the C-terminal one which is very different from the folding of already known hydrophobic-molecule transporters (Sacchetti et al., 1989; Andersen and Poulsen, 1992; Hoess et al., 1993).

In addition to their biological interest, ns-LTP have potent biotechnological applications as shown recently in the case of other basic and cysteine-rich wheat seed proteins (Blochet et al., 1993; Wilde et al., 1993).

METHODS

NMR constraints

The description of the experimental procedures of data recording and processing of ^1H -NMR spectra and of their assignments has been already published (Simorre et al., 1991). A total of 1416 distance constraints derived from NOE cross-peak intensities on nuclear Overhauser spectroscopy (NOESY) maps recorded with mixing times of 60 ms and 250 ms, were used. This included a set of selected 370 sequential constraints (i –($i+1$), 254 medium range ($\Delta i \leq 5$) and 177 long-range ($\Delta i > 5$) constraints and 80 intraresidual constraints. All these distances, corresponding to non-trivial NOE, were considered as essential for the folding of the polypeptide backbone. 535 additional intraresidual side-chain constraints which can involve important dynamical effects were introduced only after the first steps of the modelling. Interproton distances were classified into the three common ranges: 0.2–0.33 nm, 0.2–0.4 nm and 0.2–0.5 nm, except those corresponding to long-range and side-chain connectivities for which a 0.2–0.5 nm-range was chosen, whatever the cross-peak intensity. The choice of such distance intervals was partly justified by the difficulty in quantification of NOE in rather crowded 2D ^1H spectra. NOE recorded with a mixing time of 250 ms, were essentially used to confirm the assignments based on NOE with a smaller mixing time. For constraints deduced from these connectivities the upper distance limit was raised to 0.6 nm to take into account possible spin-diffusion effects. The α -helical regions were defined according to NMR criteria: $\text{H}_\text{N}(i)$ – $\text{H}_\text{N}(i+1)$, $\text{H}_\alpha(i)$ – $\text{H}_\text{N}(i+3)$ and $\text{H}_\alpha(i)$ – $\text{H}_\beta(i+3)$ NOE connectivities, small $^3J_{\text{N}\alpha}$ coupling constants (≤ 6 –7 Hz, except for some residues located in distorted regions) and slowly exchanging backbone amide protons. Assuming that these last protons were engaged in hydrogen bonds, we initially introduced 25 additional helical hydrogen-bond distance constraints in which carbonyl oxygen and amide nitrogen atoms were constrained within a range of 0.26–0.35 nm. A determination of disulfide pairings in wheat ns-LTP could not be achieved by chemical methods due to the absence of adequate sites of hydrolysis. A tentative distribution of disulfide bridges was proposed (Takishima et al. 1988) for a castor bean protein from the amino acid analysis. Considering the high degree of conservation in the positions of the eight Cys residues in the family of plant ns-LTP, we assumed the following network of disulfide bridges for wheat ns-LTP, 3–50, 13–27, 28–73 and 48–87, which was partly confirmed by long-range con-

nectivities detected for the following pairs: (Cys3) CaH-C β H (Cys50), (Cys27) NH-CaH (Cys13), (Cys 27) CaH-C β H (Pro12) and (Cys73) CaH-C β H (Cys28) (Simorre et al., 1991). In the maize ns-LTP that we are studying now, interproton connectivities detected by NMR confirm such a disulfide pairing (Sodano, P., personal communication). Nevertheless, it should be mentioned that in the soybean hydrophobic protein which exhibits strong structural similarities with ns-LTP, the pairing of the two Cys residues 43 and 45 engaged in the triplet Cys-Xaa-Cys is different, according to the crystal structure recently established by Baud et al. (1993). This pairing is 8–43 and 45–77, whereas in ns-LTP it is 3–50 and 48–87 i.e., the links are inverted for the 43–45 and 50–48 pairs of Cys residues. In the present work, we decided to keep the disulfide pairings that we initially proposed and to introduce distance constraints in the following ranges: 0.19–0.21 nm for S_i – S_j bonds and 0.25–0.35 nm for C_i^β – S_j^γ and C_j^β – S_i^γ bonds. As seen later, this assumption is fully compatible with the structure built from the total set of constraints. Four dihedral constraints were added to maintain the configurations of Xaa-Pro amide bonds (Pro12, Pro21, Pro23 and Pro78). From NMR experiments (connectivities αH_i – αH_{i+1} and NH_i – αH_{i+1}) it appears that the Gly22-Pro23 bond may have an unusual *cis* configuration whereas the other three bonds appear clearly as *trans* (connectivities αH_i – δCH_2_{i+1} and NH_i – δCH_2_{i+1}). The Φ values derived from $^3J_{\text{N}\alpha}$ measurements were not included in the refinement process of the structure but were used for verification of the results. Due to numerous overlaps of the H_α – H_β and H_α – H_γ cross-peaks in the helical part of the protein, we were only able to stereospecifically assign the β protons of three residues, Asp2, Ser15, Ser68, and the γ methyl groups of Val6, Val10, considering $^3J_{\alpha\beta}$ coupling constants and intraresidual NOE (Hyberts et al., 1987; Wagner et al., 1987). This lack of stereospecific assignments was presumably one of the causes of the rather low resolution of the structures issued from our calculations.

In distance-geometry calculations, constraints were applied on pseudo atoms, replacing unassigned side-chain protons, the upper limits being raised by using Wüthrich values (Wüthrich et al., 1983). For simulated annealing and energy minimisation we used an R^{-6} averaging.

Molecular modelling

The 74 initial structures generated using the selected set of 881 distance constraints and the metric matrix DSPACE Fortran program (Hare, 1990) were divided into two classes, A (42 structures) and B (32 structures), both giving the same overall rough agreement with the experimental data. Several criteria discussed below enable us to eliminate class B. From the 42 remaining structures, 21 leading to large violations and steric clashes were also eliminated. The remaining 21 structures were submitted to simulated annealing using the standard XPLOR SA protocol (Brünger, 1986). Non-bonded energy terms were replaced by a simple van der Waals' repulsion term. For distance and dihedral angle constraints, the initial force constants were set to $20000 \text{ kJ mol}^{-1}\text{nm}^{-2}$ and $200 \text{ kJ mol}^{-1}\text{rad}^{-2}$, respectively, and a soft square well potential with a low asymptote slope $c = 0.01$ was used. The van der Waals' force constant for the repulsive term was set to $0.8 \text{ kJ mol}^{-1}\text{nm}^{-2}$. After 200 cycles of Powell energy minimization, initial velocities were randomly assigned at 1000 K according to a Maxwellian distribution and all masses were set to uniform values (10 u). After 10 ps of molecular dy-

namics (MD) simulations at 1000 K, the van der Waals' force constant was gradually increased from $0.8 \text{ kJ mol}^{-1}\text{nm}^{-2}$ to $40 \text{ kJ mol}^{-1}\text{nm}^{-2}$ while the slope was increased from 0.01 to 0.1. This stage consisted in 20 cycles of 0.5-ps dynamics at 1000 K. The system was then cooled down to 300 K in 25-K steps and subjected to 250 steps of Powell energy minimization.

In simulated annealing, the molecular dynamics simulations were performed using the SHAKE algorithm (Ryckaert et al., 1977) to constrain the bond length. During the simulations, the velocities were rescaled in order to maintain the system at a constant temperature, using a time constant of 0.01 ps. The structures were submitted to energy minimization using the standard CHARMM potential-energy function (Brooks et al., 1983), the electrostatic energies were calculated using an ϵ_r distance-dependent function $4 r_{ij}$.

The 21 structures issued from preliminary calculations were used for back calculations of the NOE intensities using the FELIX software (Hare, 1990). The comparison between calculated and experimental NOESY maps showed a very good agreement in the backbone proton region and confirmed the efficiency of the set of constraints initially introduced with the exception of a few hydrogen-bond constraints that were removed. For side-chains the fitting was improved by taking into account all intraresidue constraints. After a last cycle of simulated annealing calculations including the total set of constraints these 21 structures were energetically minimized using a force constant K_{noe} of $10000 \text{ kJ mol}^{-1}\text{nm}^{-2}$.

All calculations were carried out on a Silicon Graphics Crimson and 4D 35 TG work stations. The structures were visualized and analyzed using the TURBO FRODO program (Roussel and Cambillau, 1989).

RESULTS

Distance-geometry calculations

On the basis of their root mean square deviation (RMSD), the 74 initial structures generated by using DSPACE program were divided in two classes, A and B (RMSD ranging from 0.1 nm to 0.4 nm within a class and from 1 nm to 1.5 nm between the two classes). Comparison of two typical A and B structures shows that they are topological mirror images as described by Pastore et al. (1991). The following criteria were used to discriminate between the two classes: the number of distance constraints violations was lower for members of class A; on Ramachandran maps, the number of residues in energetically allowed regions was much higher for class A than for class B. This difference was specially clear in the α_R region where only a few residues of structures B were present, whereas NMR results indicated unambiguously the existence of four helices which should be right-handed as usually found in proteins.

From these results, the structures of class B were discarded. Among the 42 structures of class A, 21 were used for further refinement.

Simulated annealing and energy minimisation

Overall conformation

The 21 final structures issued from simulated-annealing and energy-minimization calculations exhibit low RMSD

Table 1. Deviations with respect to the standard geometry.

Value	Bond angles	Improper dihedrals	Bond lengths
	°		pm
Average	2.9	0.35	10
Standard deviation	0.1	0.04	0.05

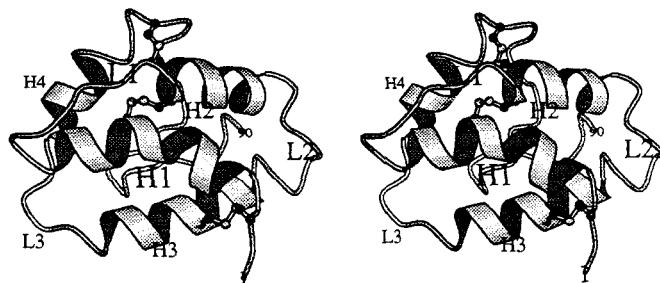


Fig. 1. Stereo view of wheat ns-LTP. The program MolScript (Kraulis, 1991) was used.

Table 2. Average atomic RMS Differences and standard deviations on different segments of the protein. The RMSD are calculated for each pair of structures. Values given are the average values and standard deviation of these RMSD.

Segment	Mean	Standard deviation
	nm	
All atoms	0.28	0.05
Backbone atoms	0.20	0.02
Helices (backbone atoms)		
H1 (5–18)	0.067	0.014
H2 (26–36)	0.054	0.015
H3 (45–56)	0.060	0.020
H4 (62–68)	0.036	0.02
C-terminal (backbone atoms)		
70–80	0.122	0.03
80–90	0.094	0.025
Loops (backbone atoms)		
L1 (19–25)	0.188	0.05
L2 (37–42)	0.154	0.05
L3 (58–61)	0.068	0.03

with respect to the standard covalent geometry (Table 1), low van der Waals' ($-1776 \pm 58 \text{ kJ mol}^{-1}$) and electrostatic ($-1170 \pm 37 \text{ kJ mol}^{-1}$) energies which are indicative of good covalent geometry and good non-bonded interactions. The average deviation from the experimental distance-constraints was $0.005 \pm 0.0009 \text{ nm}$. All residual distance-constraint violations were less than 0.1 nm, and over the 21 structures only 9 violations were larger than 0.05 nm. Fig. 1 shows a schematic drawing of the backbone folding which consists of four helices connected by loops and a non-standard structure for the C-terminal region. All these elements are arranged in a right-handed winding. The conformation of the polypeptide backbone is especially well defined in the helical regions as indicated by the average pairwise RMSD between the indivi-

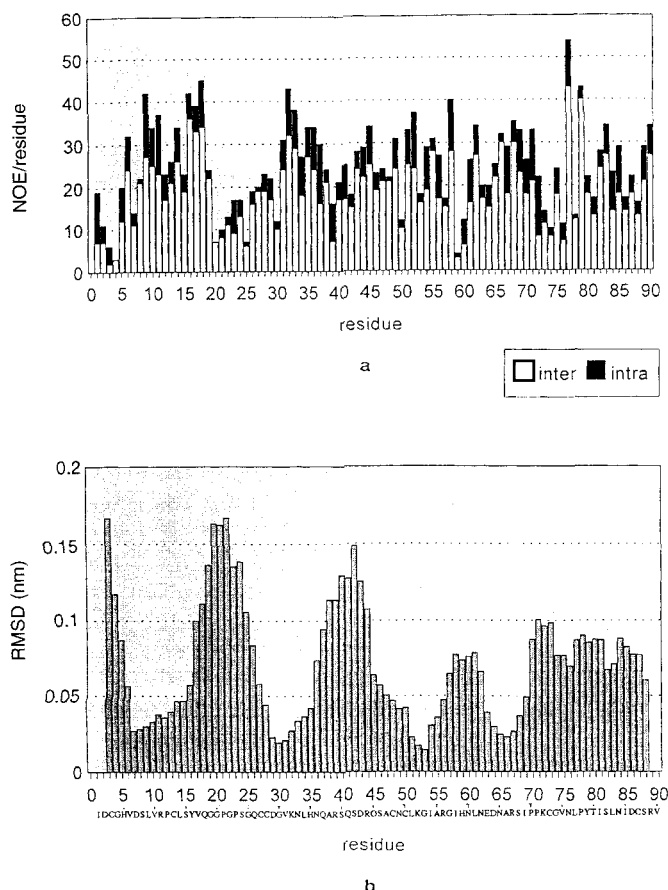


Fig. 2. NOE/residue and mobile RMSD. (a) Number of NOE/residue. (b) Mobile RMSD with a window of 5 residues. The mobile RMSD is a tool developed in our group (Cornet, B., personal communication) which calculates the local RMSD on atomic coordinates for a window of n residues (n being chosen by the user, the window moving along the sequence). The average values of this local RMSD calculated for pairs of structures is plotted versus the residue number.

dual conformers (Table 2). The loops connecting the helices show a much larger variability, specially L1 which includes a succession of Gly and Pro residues.

In the C-terminal region the last ten residues are better defined than the preceding ten (RMSD equal to 0.122 nm for the backbone atoms of residues 70–80 and to 0.095 nm for residues 80–90) (Table 2).

Globally, for the 21 structures, there is a good correlation between the total number of NOE constraints used for each residue and the mobile RMSD (Fig. 2). This mobile RMSD, is defined as the average value of the pairwise RMSD of the backbone heavy atoms calculated on a window of five residues moving along the sequence.

On Ramachandran maps, most residues of these 21 structures fall in energetically favourable regions (see Fig. 3 for a superposition of 21 Ramachandran maps without Gly residues). However, a more detailed analysis of these maps using PROCHECK (Laskowski et al., 1993) shows that about 55% of the residues are in the most favourable region (A, B, L), 40% in the additional allowed regions (a, b, l, p) and 5% in the generously allowed regions ($\sim$ a, $\sim$ b, $\sim$ l, $\sim$ p), the latter corresponding to loop residues. For example Ser24 in loop L1, Arg39 in loop L2, and several residues of the C-terminus are at the limit of the allowed domains. Leu83 exhibits a

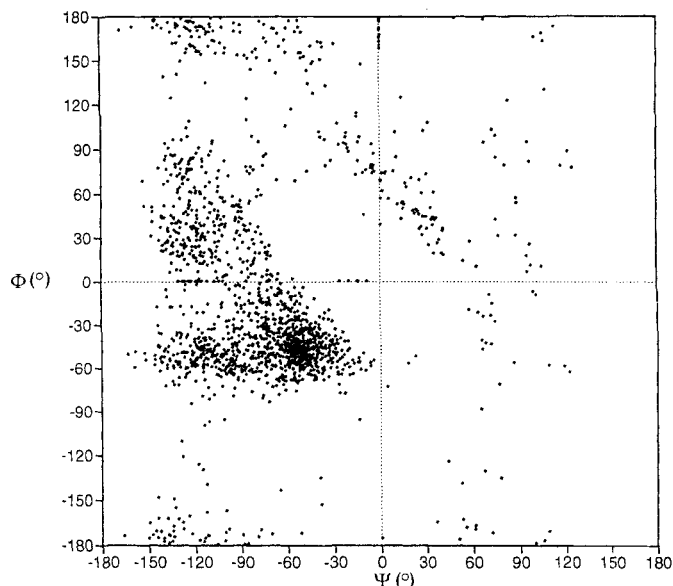


Fig. 3. Ramachandran plots for the 21 structures (without the Gly residues).

particular behaviour: in two structures it is located in the 3_{10} helix area, whereas in others it is found in an unusual area (60° , -60°) corresponding to γ turns (Sibanda and Thornton, 1985, 1991; Topham et al., 1993). According to PROCHECK analysis, such Ramachandran maps correspond to low-resolution structures: 0.3–0.35 nm. However, the resolution is much better, and can be considered as a medium resolution (0.2–0.25 nm) when calculated on other criteria: covalent geometry, bad contacts, hydrogen bonds, side-chain conformations.

Helices

Four helical regions extending approximately from residues 5 to 18 (H1), 26 to 36 (H2), 41 to 57 (H3) and 63 to 69 (H4) were analyzed by using two sets of criteria, (Φ , Ψ) angles (Fig. 4a, b), and hydrogen bonding. The (Φ , Ψ) angle criteria are not very restrictive since the conformational space corresponding to such secondary structures is not limited to the (-57° , -47°) canonical region but can extend through the left part of the Ramachandran map. Topham et al. (1993) considered that the permitted range for main-chain torsion angles is $-100^\circ < \Phi < 0^\circ$ and $-100^\circ < \Psi < 0^\circ$. Okamoto et al. (1991) proposed the following range: $\Phi = -60 \pm 45^\circ$ and $\Psi = -50 \pm 45^\circ$. Wilmot and Thornton (1990) and Rooman et al. (1991) extended the helical area to positive Ψ values, the latter dividing the helical area into A and C domains representing α_R and 3_{10} helices characterized by $C = O_i$, $N-H_{i+4}$ hydrogen bonds, and $C = O_i$, $N-H_{i+3}$, respectively. In our analysis, we used these maps. Hydrogen bonds were considered as identifiable when $O-N$ and $O-H(N)$ distances were less than 0.34 nm and 0.235 nm, respectively, on at least 11 of the 21 structures. The circular variance for each residue as defined by Mc Arthur and Thornton (1993) is represented in Fig. 4c, d. In the helical regions, the dispersion is low since the circular variance is less than 0.05, except for few residues. The 5–18 H1 helix appears as a somewhat distorted α_R helix in which Φ angles are ranging around -90° for residues Val10 and Cys13 involved in a sequence Val-Arg-Pro-Cys and for the terminal

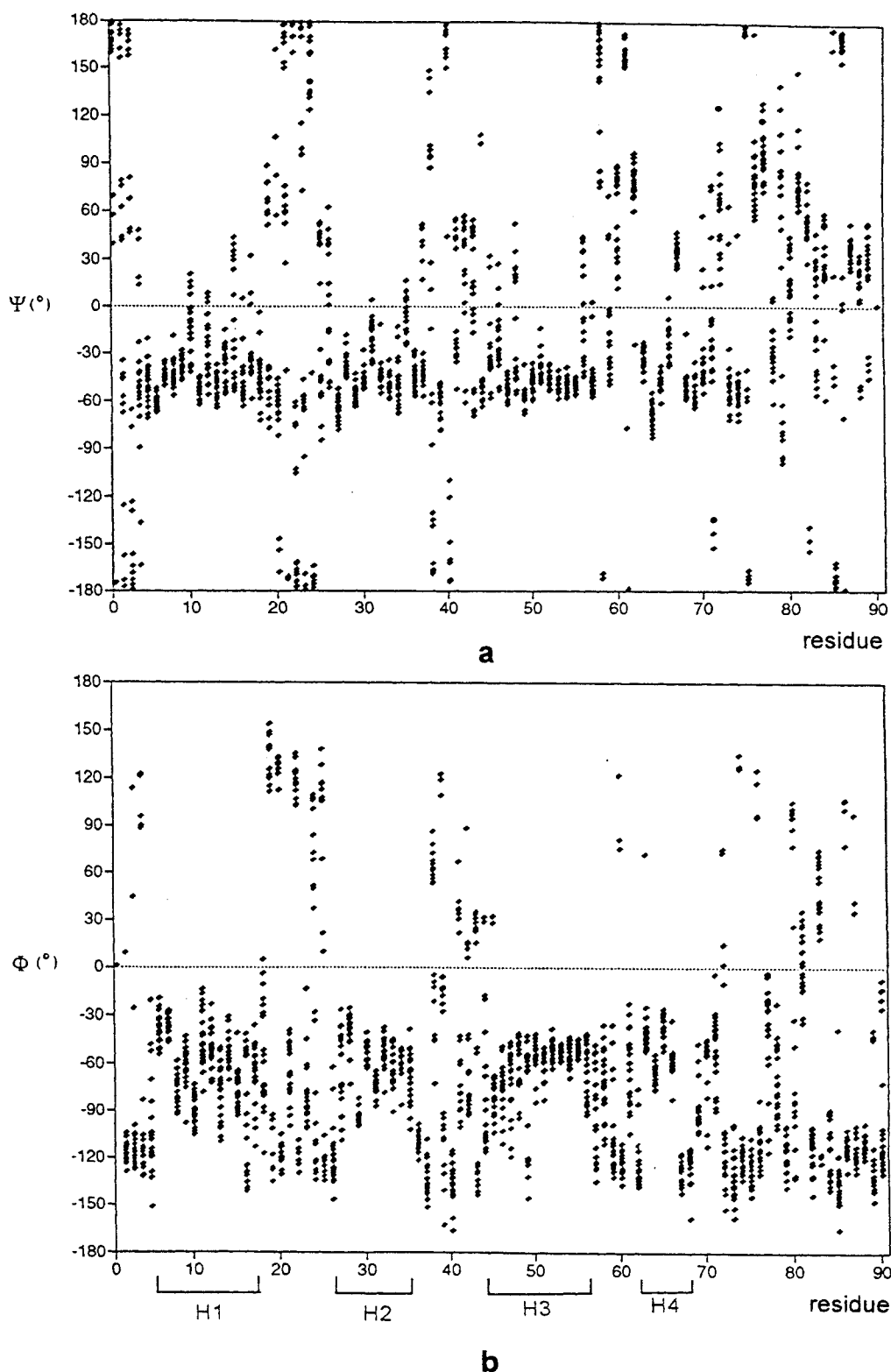


Fig. 4. Variations of Φ , Ψ angles over the 21 structures. (a) Plot of the Φ angles of each residue for the 21 structures. (b) Plot of the Ψ angles of each residue for the 21 structures. (c) Plot of the Φ circular variance/residue. (d) Plot of the Ψ circular variance/residue.

residue Gln18. These values are in agreement with the large $^3J_{\text{HN}\alpha}$ coupling constants (≈ 9 Hz) found for these three residues (Simorre et al., 1991) and for Val10 with the results of a study on the conformation of residues preceding proline in helices (MacArthur and Thornton, 1991). The hydrogen-

bond network is perturbed by the presence of the Pro12 residue, (i)-($i+4$) bindings, 8–12, 9–13 and 12–16, characteristic of an α_R conformation are missing. The loss of such hydrogen bonds is counterbalanced by two (i), ($i+3$) hydrogen bonds between residues 10–13 and 13–16. In addi-

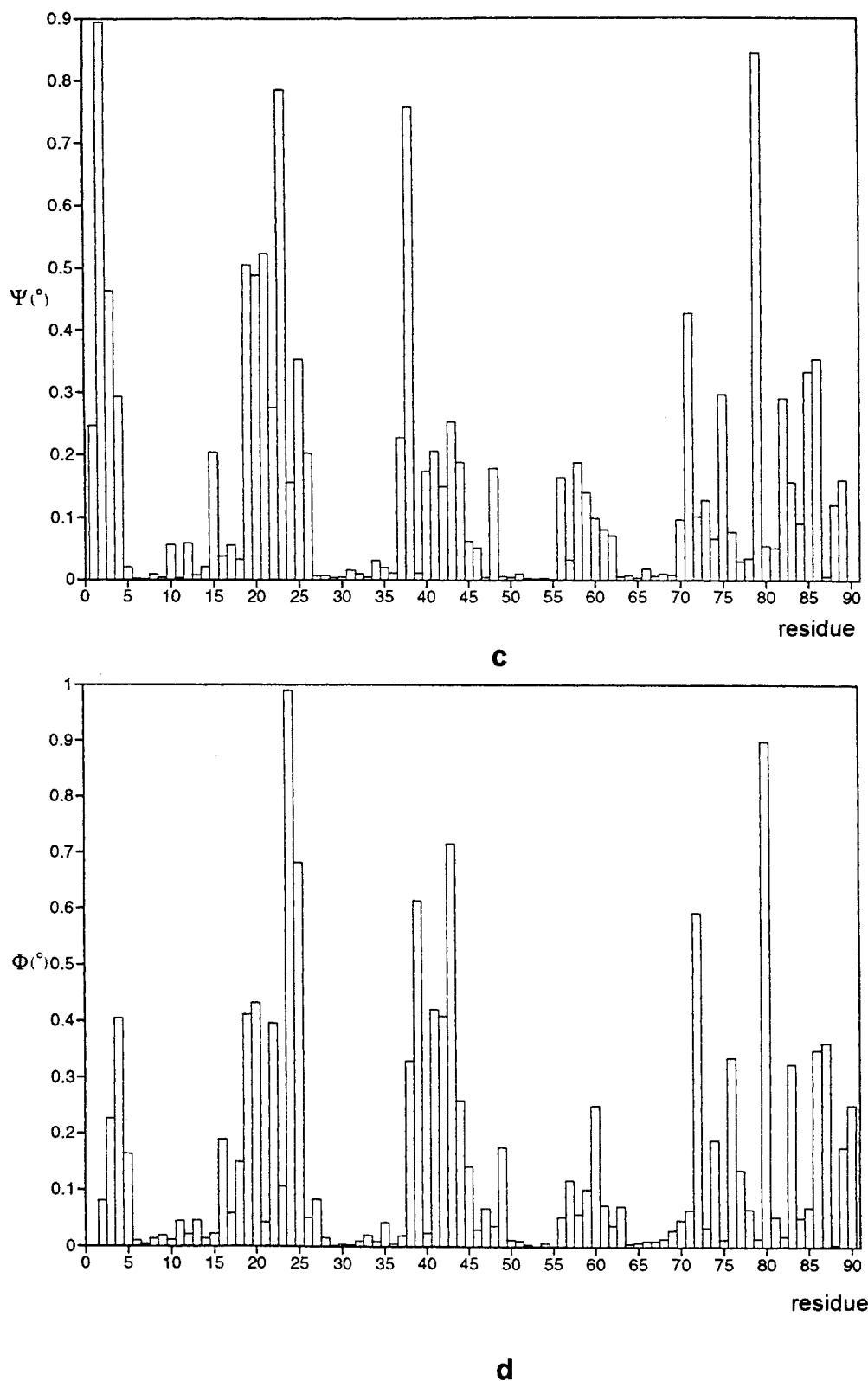


Fig. 4. Continued

tion to these perturbations, in all the calculated structures, the H1 helix is kinked, the kink angles range between 25° and 30° (Fig. 1). This perturbation can be related to the presence of a Pro residue; Pro residues inserted in a helix are known to induce a kink (Barlow and Thornton, 1988). The

Φ , Ψ angles of helix H2 are characteristic of a more regular α_R helix. From residue 27 to 36 the angular dispersion (Fig. 4c and d) is very low. Large negative Φ values are counterbalanced by low negative Ψ values, in agreement with the Ramachandran maps derived from the probability-

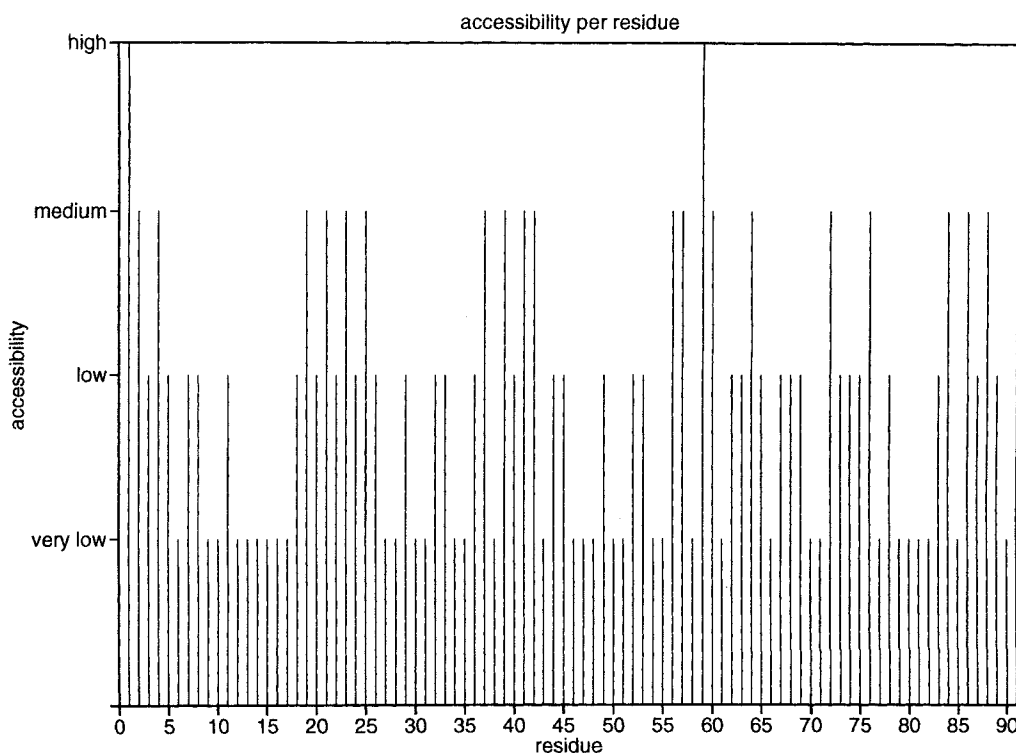


Fig. 5. Accessibility/residue. Very low accessibility, normalized solvent accessibility less than 0.25 for at least 11 of the 21 structures; low, normalized solvent accessibility ranging from 0.25 to 0.5; medium, normalized solvent accessibility ranging from 0.5 to 0.75; high, normalized solvent accessibility larger than 0.75.

density distribution of Φ , Ψ angles from protein-crystal structures (McArthur and Thornton, 1993). The hydrogen-bond network is formed from residues 26–37, except around the Val31 residue (the 30–34 hydrogen bond is missing) where a distortion of the helical symmetry is attributed to the formation of a salt bridge between the carboxylic group of Asp29 and the amino group of Lys32. An $(i)-(i+3)$ hydrogen bond is found between residue 33 and 36.

The NOE pattern suggests that helix H3 extends from residues 41–57 (Simorre et al., 1991). Fig. 4 shows that only residues 45–57 exhibit Φ and Ψ values characteristic of an α_R helix, in which the circular variance is higher for Ψ_{48} and Φ_{49} and for the last residues. This is confirmed by the hydrogen-bond pattern, since the stretch of $(i)-(i+4)$ hydrogen bonds extends almost without interruption from residue 45–58. However, if the C=O of residue 41 is hydrogen bonded to the N-H of residues 44 and 45, the residues 42 and 43 are not involved in helical hydrogen bonds in any structure. Experimentally, $(i)-(i+3)$ connectivities are clearly observed in the 41–45 region and the $^3J_{N\alpha}$ coupling constant values confirm the existence of a helix, but the amide protons of these residues appear to exchange rapidly with the deuterated solvent which suggests that some distortion or some dynamical effects occur in this region. An examination of the structure (Fig. 1) shows that the global folding of residues 41–45 remains roughly helical but with an unwinding making CO(i)-HN($i+4$) distances too large for hydrogen bonds. As a matter of fact, sequences such as: Q S D R Q S involving only polar and charged residues are known to cause distortions in helices (Bolin et al., 1982).

According to our distance criteria defining hydrogen bonds, helix H4 is rather similar to a 3_{10} helix. The $(i)-(i+3)$ hydrogen bonds are formed on almost each calculated struc-

ture from residues 63–69 except in some structures where $(i)-(i+4)$ hydrogen bonds (62–66 and 65–69) appear. The Ψ values range around -40° and Φ angles of residues 67 and 68 located near the extremity of the helix interrupted by a doublet of Pro residues have unexpected large negative values (-110°) which do not prevent formation of the hydrogen bonds and which are in agreement with $^3J_{N\alpha}$ coupling constant values in the range 6–9 Hz. This result compares well with protein crystallographic data, where 3_{10} helices show a wide spread of (Φ, Ψ) angles and are intermediate between ideal 3_{10} and α_R helices (Barlow and Thornton, 1988).

Loops

Three loops L1(19–25), L2 (37–44) and L3 (58–62) interconnect the four helical fragments previously described. The first loop involving a particular sequence with four Gly and two Pro residues (G G P G P S G) and a 22–23 bond which could be *cis*, is poorly defined; a small number of NOE corresponding to high RMSD (Fig. 2) strongly suggests the existence of important local motions. As discussed previously, the 41–44 sequence connecting the L2 loop to the H3 helix seems to be involved in a conformational equilibrium and the RMSD is at a maximum in this region (Fig. 2). However, the shorter L3 loop 58–62 is better defined in spite of the small number of NOE detected for residues 59 and 60. No hydrogen bonds can be found in these loops using the distance criterion defined previously. The mobile RMSD represented on Fig. 2 using a short window of five residues shows that there is no local convergence in the loop regions.

C-terminal region

The folding of this part of the backbone separated from the H4 helix by a doublet of Pro residues (71–72) does not

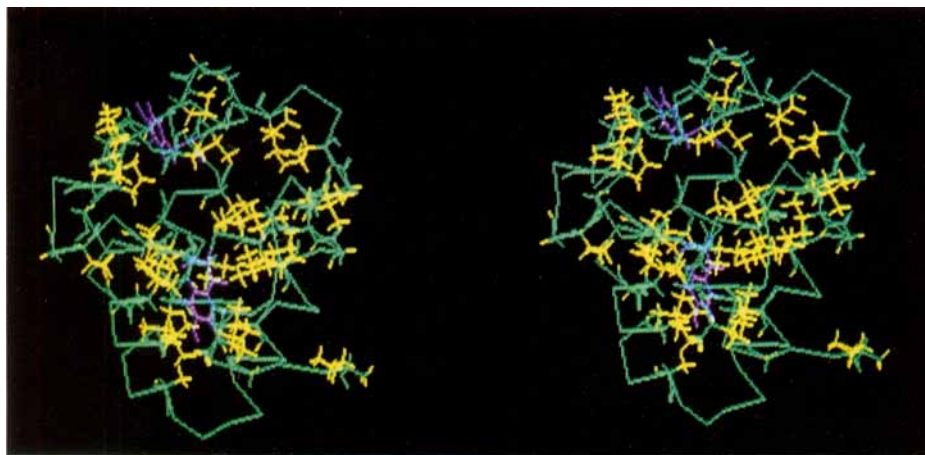


Fig. 6. Stereo view of wheat ns-LTP showing the hydrophobic residues and the backbone.

correspond to any standard structure but adopts a saxophone-like shape formed by two successive turns, the last one being a γ turn, stabilized by a hydrogen bond between the C=O of Ser82 and the NH of Asn84. Another hydrogen bond was detected between the C=O of Ile69 and the NH of Lys72 in spite of the presence of two Pro residues.

Side chains

The χ_1 angle values show a large dispersion (circular variance of 0.34 ± 0.03) which can have three origins: the lack of stereospecific assignments; the large constraint intervals chosen for connectivities involving side-chains protons; the side-chain mobility. However, a PROCHECK analysis shows that over the 21 calculated structures most residues fall in the allowed regions of the χ_1, χ_2 plots.

Solvent accessibility calculations (Lee and Richard, 1977), are summarized on Fig. 5. Most of hydrophobic residues are found inside the structure while polar and charged residues are exposed to the solvent. The residues having the largest accessibility are found in the loop regions and in the C-terminal part, and are either charged, polar or short (Gly) residues. The most accessible residues are Ile1 and also His59 at the beginning of loop L3.

The amphiphilic H1, H2, H3 helices and the C-terminal region define an hydrophobic central core in which all the following residues are packed together (Fig. 6): Val6, Leu9, Val10, Leu14 on H1, Val31, Leu34 on H2, Ala38 in loop L2, Ala47, Leu51, Ile54 on H3 and Leu77, Ile81 on the C-terminus. Several more peripheral hydrophobic contacts are found between residues Val17, Ile58 and Ile61, between Leu83, which points outward the C-terminal short γ turn and Ala55 and also between Ile85, Ile81 and Leu51. To determine the possibility of lipid binding inside an hydrophobic cleft, we looked for cavities inside the protein structure by using the Pore program developed by Smart (Smart et al., 1993). The overall surface of the cavity is shown in Fig. 7. At the centre of the protein structure, its radius can be as large as 0.2 nm. On both extremities this cavity is constricted, its dimensions ranging from 0.02 nm between H4 and L1 to 0.035 nm and 0.055 nm between L2 and the C-terminus and the N-terminus, respectively. These results shows that, even using the most favourable route between L2 and the N-terminus, rearrangements would be necessary to allow the passage of the lipid.

Using the same distance criteria as for the helices, we were able to detect several hydrogen bonds involving side-chain atoms. Three side-chain backbone hydrogen bonds were found between C=O (Val17) and H δ (Asn62), between O γ (Ser40) and HN (Gln41) and between C=O (Thr80) and H γ (Ser82). Several hydrogen bonds and salt bridges involved polar and charged side-chain atoms, O δ (Asp86)...H ϵ (Arg89), which closes the short loop formed by the C-terminal part of the protein, O δ (Asp29)...H ϵ (Lys32) stabilizing helix H2 and to conclude, two long-range hydrogen bonds may contribute to maintenance of the folding of the protein; they are situated between O δ (Glu63) in helix H4 and H δ (Asn84) in the C-terminal region, OH (Tyr16) in helix H1 and O δ (Asn65) in helix H4. However, the packing of the amphiphilic helices is mainly stabilized by the four S-S bridges and by hydrophobic interactions.

Disulfide bridges

Considering the small number of specific constraints available for Cys residues forming the four disulfide bridges as well as for neighbouring residues, the conformations of these covalent links are rather poorly defined. Two minima around the $\pm 90^\circ$ values currently found in protein crystal structures (Richardson, 1981) were obtained for the χ_3 torsional angles around the S γ -S γ bonds. For the 21 structures, the conformations of the 13–27 and 28–73 bridges are appreciably less scattered than those of the other two. These S-S bridges are grouped by pairs at two opposite poles of the protein and it could be considered that they possess some flexibility favouring a conformational change induced by the binding of a lipid molecule.

DISCUSSION

The first 3D structure of a ns-LTP has been established by molecular modelling on the basis of 3D and 2D ^1H -NMR data of an aqueous solution of a wheat ns-LTP (Simorre et al., 1991). In spite of the rather high number of NOE taken into account in our calculations, the use of large intervals in the definition of the distance constraints and the small number of stereospecific assignments were at the origin of the rather low resolution of the calculated structures. Nevertheless some essential aspects of the folding were unques-



Fig. 7. Stereo view of the pore through wheat LTP. The backbone is represented as a tube, for clarity only the hydrophobic residues are shown, pink circles show the surface of the cavity and a yellow line the center of each pathway.

tionably established. The backbone folds into a simple and original right-handed winding composed of a bundle of four helices linked together by flexible loops and packed against a C-terminal fragment made of a series of turns forming a non-standard saxophone-like shape. Several Pro residues play an important role in this folding: Pro12 induces a kink in the H1 helix, which favours interactions with the internal hydrophobic core; Pro21 and Pro23 combined with three Gly residues are involved in a loop connecting helices H1 and H2; the doublet Pro70-Pro71 makes the transition between the H4 helix and the C-terminal fragment. The Gly22-Pro23 amide bond could be *cis*, an isomer already observed in turns (type IV) of different proteins (Richardson, 1981). The structural importance of these amino acid residues is emphasized by the fact that Pro12, Pro23 and Pro70 are highly conserved in the sequences of plant ns-LTP (Désormeaux et al., 1992).

Considering the secondary structure data obtained by infrared spectroscopy at protein concentration (millimolar range) and low pH similar to those used in NMR experiments (Désormeaux et al., 1992), the 3D model defined in the present work leads to a higher helix content: about 55% instead of 40%. This discrepancy could be explained by the discontinuities of the C=O–N–H hydrogen-bond network in the helices previously discussed, which is the basis of secondary-structure determination by infrared spectroscopy (Désormeaux et al., 1992).

The two Tyr residues present in the sequence of the wheat protein are highly conserved in the ns-LTP family. Among the nine sequences compared by Désormeaux et al. (1992) small variations are only observed in castor bean isoforms, where Tyr16 is replaced by Phe and Tyr79 by Ile, Val or Phe. These residues are supposed to play an important role in the function of the proteins. Tyr79, belonging to the C-terminal fragment is located at the extremity of the hydrophobic cleft in the vicinity of Leu34, Ala47 and Leu51 and should be perturbed by lipid binding. Tyr16 located at the extremity of the H1 helix is adjacent to Val17 residue which is involved in a small hydrophobic patch (see above) extending the internal hydrophobic pocket. The structure presented herein defines

a new basis for a rational interpretation of fluorescence measurements (Marion, D., unpublished results).

To extend our structure analysis to the other members of the plant ns-LTP family, we undertook recently a ¹H-NMR study of the maize protein. Its secondary-structure elements were identified and compared to those of wheat ns-LTP (Petit et al., 1994). Globally, these elements are located in the same regions and it can be assumed that the foldings of both proteins should be very similar considering the conservation of the eight Cys positions and the existence of about 60% similarity. Nevertheless, a more precise comparison between these two structures will be necessary to understand the differences in the transfer activities since the maize ns-LTP is much more efficient than that of wheat (Petit et al., 1994).

As previously mentioned, we also investigated the interactions of these two ns-LTP with lyso-PtdCho. In the experimental conditions required for NMR experiments (pH ≈ 4.5, approximately millimolar protein concentration in water) two lyso-PtdCho with different aliphatic chain lengths (C₁₆ and C₁₂) bind to the proteins to form complexes. The NMR data show that the lipid aliphatic chains strongly interact with the proteins and are very probably inserted in their hydrophobic pockets, whereas the polar heads are located at the surface, in the vicinity of the polar and charged sites. These preliminary results, strongly suggest that ns-LTP can behave as lipid carriers in aqueous solution. Several 3D structures of animal lipid-binding proteins have been resolved recently. The spatial organisation of fatty-acid-binding proteins (FABP; Sacchettini et al., 1989), acyl-coenzyme-A-binding proteins (Andersen and Poulsen, 1992) as well as of a lipopolysaccharide-binding protein (Hoess et al., 1993) is quite different from the right-handed winding of wheat ns-LTP which can be considered as a new type of organization for small proteins interacting with lipids. The present study is a prerequisite for a further determination of the 3D structures of the protein/lipid complex and for the establishment of structure/activity relationships. Our NMR studies in progress on the 3D structure of maize ns-LTP and on the binding

of different lipids by wheat and maize proteins will be the next step in this direction.

This work was supported by the *Centre National de la Recherche Scientifique* (CNRS), the University of Orléans, the Région Centre, the *Institut National de la Recherche Agronomique* (INRA, AIP 91/4749) and the European Economic Community (ECLAIR AGRE 0052). We are grateful to M. C. Petit for providing unpublished data on the interaction between lyso-PtdCho and wheat ns-LTP, and Dr Jean-Claude Kader (University of Paris VI, France) for his interest and helpful discussions. We are indebted to Professor M. Pézolet (Laval University, Québec, Canada) for stimulating discussions and vibrational spectroscopy. We would like to thank O. Smart for providing us with the Hole program.

REFERENCES

- Andersen, K. V. & Poulsen, F. M. (1992) *J. Mol. Biol.* **226**, 1131–1141.
- Barlow, D. J. & Thornton, J. M. (1988) *J. Mol. Biol.* **201**, 601–619.
- Baud, F., Pebay-Peyroula, E., Cohen-Addad, C., Odani, S. & Lehman, M. S. (1993) *J. Mol. Biol.* **231**, 877–887.
- Bernhard, W. R., Thoma, S., Botella, J. & Somerville, C. R. (1991) *Plant Physiol. (Bethesda)* **95**, 164–170.
- Bloch, J. E., Chevalier, C., Forest, E., Pebay-Peroula, E., Gautier, M. F., Joudrier, P., Pézolet, M. & Marion, D. (1993) *FEBS Lett.* **329**, 336–340.
- Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C. & Kraut, J. (1982) *J. Biol. Chem.* **257**, 13650–13662.
- Bouillon, P., Drischel, C., Vergnolle, C., Duranton, H. & Kader, J. C. (1987) *Eur. J. Biochem.* **166**, 387–391.
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983) *J. Comput. Chem.* **4**, 187–217.
- Brünger, A. T., Clore, G. M., Gronenborn, A. M. & Karplus, M. (1986) *Proc. Natl Acad. Sci. USA* **83**, 187–217.
- Campos, F. A. P. & Richardson, M. (1984) *FEBS Lett.* **167**, 221–225.
- Désormeaux, A., Bloch, J. E., Pézolet, M. & Marion, D. (1992) *Biochim. Biophys. Acta* **1121**, 137–152.
- Fleming, A. J., Mandel, T., Hofman, S., Sterk, P., De Vries, S. C. & Kuhlmeier, C. (1992) *Plant J.* **2**, 855–862.
- Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120.
- Hare, D. (1990) Reference manual, *Hare Research inc.*, Woodinville.
- Hoess, A., Watson, S., Siber, G. R. & Liddington, R. (1993) *EMBO J.* **12**, 3351–3356.
- Hyberts, S. G., Marki, W. & Wagner, G. (1987) *Eur. J. Biochem.* **164**, 625–635.
- Kader, J. C. (1990) in *Subcellular biochemistry* (Hilderson, H. J., ed.) pp. 69–110, Plenum Publishing Corporation, New York.
- Koltunow, A. M., Truettner, J., Cox, K. W., Walroth, M. & Goldberg, R. B. (1990) *Plant Cell* **2**, 1201–1229.
- Kraulis, P. J. (1991) *Acta Crystallogr. Sect. B Struct. Sci.* **48**, 115–116.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283–291.
- McArthur, M. W. & Thornton, J. M. (1991) *J. Mol. Biol.* **218**, 397–412.
- McArthur, M. W. & Thornton, J. M. (1993) *Proteins* **17**, 232–251.
- Molina, A., Segura, A. & Garcia-Olmedo, F. (1993) *FEBS Lett.* **316**, 119–122.
- Mundy, J. & Rogers, J. C. (1986) *Planta (Heidelb.)* **169**, 51–63.
- Okamoto, Y., Fukugita, M., Nakazawa, T. & Kawai, H. (1991) *Protein Eng.* **4**, 639–647.
- Pastore, A., Atkinson, R. A., Daudek, V. & Williams, R. J. P. (1991) *Proteins Struct. Funct. Genet.* **10**, 22–32.
- Petit, M. C., Sodano, P., Marion, D. & Ptak, M. (1994) *Eur. J. Biochem.* **222**, 1047–1054.
- Richardson, J. S. (1981) *Adv. Protein Chem.* **34**, 168–339.
- Rooman, M. J., Kocher, J. P. A. & Wodak, S. J. (1991) *J. Mol. Biol.* **220**, 961–979.
- Roussel, A. & Cambillau, C. (1989) in *Silicon graphics geometry partner* (Silicon Graphics, eds) pp. 77–78, Silicon Graphics, Mountain View.
- Ryckaert, J. P., Ciccoli, G. & Berendsen, H. J. C. (1977) *J. Comput. Chem.* **23**, 327–341.
- Sacchettini, J. C., Gordon, J. I. & Banaszak, L. J. (1989) *J. Mol. Biol.* **208**, 327–339.
- Sibanda, B. L. & Thornton, J. M. (1985) *Nature* **316**, 170–174.
- Sibanda, B. L. & Thornton, J. M. (1991) *Methods Enzymol.* **202**, 59–82.
- Simorre, J. P., Caille, A., Marion, D., Marion, D. & Ptak, M. (1991) *Biochemistry* **30**, 11600–11608.
- Sossountzov, L., Avila, L. R., Vignols, F., Jolliot, A., Arondel, V., Tchang, F., Grosbois, M., Guerbette, F., Miginiac, E., Delseny, M., Puigdomenech, P. & Kader, J. C. (1991) *Plant Cell* **3**, 923–933.
- Smart, O., Goodfellow, J. M. & Wallace, B. A. (1993) *Biophys. J.* **65**, 2455–2460.
- Sterk, P., Booi, H., Schellekens, G. A., van Kammen, A. & De Vries, S. C. (1991) *Plant Cell* **3**, 907–921.
- Takishima, K., Watanabe, S., Yamada, M. & Mamiya, G. (1986) *Biochim. Biophys. Acta* **870**, 248–255.
- Takishima, K., Watanabe, S., Yamada, M., Suga, T. & Mamiya, G. (1988) *Eur. J. Biochem.* **177**, 241–249.
- Tchang, F., This, P., Stiefel, V., Arondel, V., Morch, M.-D., Pages, M., Puigdomenech, P., Grellet, F., Delseny, M., Bouillon, P., Huet, J. C., Guerbette, F., Beauvais-Cante, F., Duranton, H., Pernollet, J. C. & Kader, J. C. (1988) *J. Biol. Chem.* **263**, 16849–16855.
- Terras, F. R. F., Schoofs, H. M. E., De Bolle, M. F. C., van Leuven, F., Rees, S. B., Vanderleyden, J., Cammue, B. P. A. & Broekaert, W. F. (1992) *J. Biol. Chem.* **267**, 15301–15309.
- Topham, C. M., McLeod, A., Eisenmenger, F., Overington, J. P., Johnson, M. S. & Blundell, T. (1993) *J. Mol. Biol.* **229**, 194–220.
- Torres-Schumann, S., Godoy, J. A. & Pintor-Toro, J. A. (1992) *Plant Mol. Biol.* **18**, 749–757.
- Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Go, N. & Wüthrich, K. (1987) *J. Mol. Biol.* **196**, 611–639.
- Wilde, P. J., Clark, D. C. & Marion, D. (1993) *J. Agric. Food Chem.* **41**, 1570–1576.
- Wilmot, C. M. & Thornton, J. M. (1990) *Protein Eng.* **3**, 479–493.
- Wüthrich, K., Billeter, M. & Braun, W. (1983) *J. Mol. Biol.* **169**, 949–961.