^{1}H NMR Assignment and Global Fold of Napin BnIb, a Representative 2S Albumin Seed Protein†

Manuel Rico,*,‡ Marta Bruix,‡ Carlos González,‡ Rafael I. Monsalve,§ and Rosalía Rodríguez§

Instituto de Estructura de la Materia, CSIC, Madrid, Spain, and Departamento de Bioquímica y Biología Molecular, Facultad de Química, Universidad Complutense, Madrid, Spain

Received July 16, 1996; Revised Manuscript Received September 17, 1996⊗

ABSTRACT: Napin BnIb is a representative member of the 2S albumin seed proteins, which consists of two polypeptide chains of 3.8 and 8.4 kDa linked by two disulfide bridges. In this work, a complete assignment of the ^{1}H spectra of napin BnIb has been carried out by two-dimensional NMR sequence-specific methods and its secondary structure determined on the basis of spectral data. A calculation of the tertiary structure has been performed using ∼500 distance constraints derived from unambiguously assigned NOE cross-correlations and distance geometry methods. The resulting global fold consists of five helices and a C-terminal loop arranged in a right-handed spiral. The folded protein is stabilized by two interchain disulfide bridges and two additional ones between cysteine residues in the large chain. The structure of napin BnIb represents a third example of a new and distinctive folding pattern first described for the hydrophobic protein from soybean and nonspecific lipid transfer proteins from wheat and maize. The presence of an internal cavity is not at all evident, which rules out in principle the napin BnIb as a carrier of lipids. The determined structure is compatible with activities attributed to these proteins such as phospholipid vesicle interaction, allergenicity, and calmodulin antagonism. Given the sequence homology of BnIb with other napins and napin-type 2S albumin seed proteins from different species, it is likely that all these proteins share a common architecture. The determined structure will be crucial to establish structure−function relationships and to explore the mechanisms of folding, processing, and deposition of these proteins. It will also provide a firm basis for a rational use of genetic engineering in order to develop improved transgenic plants.

The 2S albumins[1] are storage proteins, widely distributed in dicot seeds, and specially abundant in the *Brassicaceae* family. Due to their amino acid compositions, their high content in the protein bodies of the seeds, and their mobilization during germination, a role as nitrogen and sulfur donor has been proposed for these proteins (Youle & Huang, 1978). However, recent findings of alternate biochemical activities for 2S albumins have again raised the question of their physiological significance. The interest on *Brassicaceae* seeds, whose 2S albumins have been called napins, has increased in the last decade because they represent one of the most important resources for animal nutrition and industrial oils and are susceptible of genetic manipulation. Rapeseed *(Brassica napus),* as well as related species such as *Arabidopsis, Bertollethia,* and *Raphanus,* have also been extensively employed in the study of the gene families of the napin-type 2S albumins (NT2SAs) and their expression. The sequences of NT2SAs, or their genes, have been determined from seeds of a large number of species. (Crouch *et al.*, 1983; Ericson *et al.*, 1986; Krebbers *et al.*, 1988; Raynal *et al.*, 1991). They exhibit a high degree of polymorphism mainly due to the existence of multigene families and proteolytic processing.

These NT2SAs constitute a family of small basic proteins (12.0−15.0 kDa) composed of two different chains (about 4.0 and 9.0 kDa), which derive from a single polypeptide precursor (Cronch *et al.*, 1983). The chains are linked in the mature protein by two disulfide bridges (Ericson *et al.*, 1986). All the NT2SAs show a common pattern of eight cysteine residues, ...C...C.../...CC...CXC...C...C..., which also appears in other small seed proteins nonrelated sequentially with napins, such as nonspecific lipid transfer proteins (ns-LTP) (Takishima *et al.*, 1986) and the hydrophobic protein from soybean (HPS) (Odani *et al.*, 1987). 2S albumins have also been related to the prolamin superfamily of seed storage proteins (Shewry *et al.*, 1995).

Although a specific enzymatic activity has not been demonstrated for the NT2SAs, diverse biochemical functions have been recently described for these proteins or their subunits. A broad spectrum of antifungal activity has been demonstrated for napins obtained from different *Brassicaceae* species, with all their isoforms inhibiting the growth of different plant pathogenic fungi (Terras *et al.*, 1992, 1993). Taking into account the sequence diversity of these antifungal NT2SAs, it is likely that the antimicrobial activity is a general

† This work was supported by DGICYT Grants PB93-0189 and PB92-0195 from the Ministerio de Educación y Ciencia (Spain).

* Author to whom correspondence should be addressed.

‡ Instituto de Estructura de la Materia, CSIC, Madrid, Spain.

§ Departamento de Bioquímica y Biología Molecular, Facultad de Química, Universidad Complutense, Madrid, Spain.

⊗ Abstract published in *Advance ACS Abstracts,* November 1, 1996.

[1] Abbreviations: ASA, accessible surface area; COSY, correlation spectroscopy; $d_{\alpha N}$, sequential NOE connectivity between the $C_\alpha H$ proton and the amide proton resonances; d_{NN}, sequential NOE connectivity between backbone amide proton resonances; $d_{sch}(i,j)$, NOE connectivity between proton resonances involving at least one in the side chain of either residue *i* or *j*; HPS, hydrophobic protein from soybean; NOESY, nuclear Overhauser enhancement spectroscopy; ns-LTP, nonspecific lipid transfer protein; NT2SA, napin-type 2S albumin; 2S, 2 Svedberg units, a measure of the sedimentation coefficient; TOCSY, total correlation spectroscopy; TSP, (trimethylsilyl)propionic acid; WATERGATE, water suppression by gradient-taylored excitation.

S0006-2960(96)01748-5 CCC: $12.00

property of NT2SAs. On the other hand, 2S albumins isolated from castor bean and mustard seeds induce type-I allergy in hypersensitive subjects (Sharief & Li, 1982; Menéndez-Arias *et al.*, 1988; Monsalve *et al.*, 1993). The allergenic NT2SA isolated from yellow mustard seeds, Sin a 1, strongly interacts with phospholipid bilayers (Oñaderra *et al.*, 1994).

The small chains of NT2SAs from radish constitute a family of calmodulin antagonists (Polya *et al.*, 1993). These polypeptides, which inhibit calmodulin-dependent myosin light chain kinase, are also susceptible to phosphorylation by Ca^{2+}-dependent protein kinase in wheat embryos. As germination is associated with changes in both calmodulin levels and calmodulin inhibitory activity (Cocucci & Negrini, 1988), a physiological role has been suggested for the small chains of napins in the control of seed germination (Polya *et al.*, 1993).

In addition to their biochemical interest, the 2S albumins have been used, by means of genetic engineering, as carriers for the synthesis of biologically active peptides (Vandekerckhove *et al.*, 1989), as well as to improve the nutritional properties of grain crops by increasing their content of essential amino acids (Altenbach *et al.*, 1992).

The determination of the three-dimensional structure of a 2S albumin constitutes a fundamental first step in the direction of establishing meaningful relationships between the structure and the activities described for this broad family of proteins: nitrogen and sulfur suppliers, antifungal capability, allergenicity, lipid vesicle interaction, and calmodulin antagonism. Moreover, knowledge of the three-dimensional structure of 2S albumins will provide a firm basis for its genetic manipulation without affecting their biological activities. The determination of the three-dimensional structure of a representative member of the NT2SAs is the main objective of this work. We address it by using 2D-NMR spectroscopic methods. A complete assignment of the proton resonances of napin BnIb (Monsalve *et al.*, 1991), a 106-residue member of NT2SA seed proteins, has been carried out and a preliminary structure, having an unambiguous definition of the global shape and the overall fold, has been computed.

MATERIALS AND METHODS

Protein Isolation. Sequence-grade napin BnIb was isolated from rapeseed as previously described (Monsalve & Rodríguez, 1990; Monsalve *et al.*, 1991). After ion-exchange chromatography in SP-Sephadex C-25, the protein was desalted by gel filtration through a Sephadex G-50 superfine column equilibrated alternatively in 10 mM HCl or 50 mM ammonium bicarbonate.

NMR Spectroscopy. NMR samples were prepared by dissolving the lyophilized protein in either H_2O/D_2O (9:1) or D_2O at a concentration of ~1 mM. Spectra in D_2O show only about half the expected number of amide resonances, which correspond to those that do not exchange with solvent deuterons over the time ranging the entire series of 2D experiments (~36 h) at 293 K. This reduces signal overlap as compared to the spectrum in H_2O and consequently facilitates the assignment process. An additional sample was prepared in which the protein was incubated in D_2O at high temperature (330 K) and pH 7.0, when even the slowest exchanging amide protons were replaced by deuterons. After lyophilization under native conditions, the protein was dissolved in H_2O/D_2O (9:1), with only the fast-exchanging protons appearing in the spectra. Data were collected for three different samples at pH 3.0 (not corrected for isotope effects) and two temperatures, 293 and 303 K. Sodium [2,2′,3,3′-$^2H_4$]-3-(trimethylsilyl)propionate was used as the internal reference ($\delta = 0.00$ ppm).

NMR spectra were recorded on a Bruker AMX-600 spectrometer operating at 600 MHz for the proton. All two-dimensional spectra were acquired in the phase-sensitive mode using the time-proportional phase incrementation technique (Redfield & Kuntz, 1975; Marion & Wüthrich, 1983). Water suppression was achieved either by selective presaturation or by including a WATERGATE module (Piotto *et al.*, 1992) in the original pulse sequences prior to acquisition. Conventional pulse sequences and phase-cycling procedures were used for COSY (Aue *et al.*, 1976) and NOESY (Kumar *et al.*, 1990). Mixing times of 75 and 150 ms were used in NOESY spectra. TOCSY spectra (Bax & Davis, 1985) were acquired using the standard MLEV17 spin-lock sequence and a 60-ms mixing time.

The size of the acquisition data matrix was 2048 × 512 words in f_2 and f_1, respectively. Prior to Fourier transformation, the 2D data matrix was zero-filled up to 4096 × 1024 words. Resolution enhancement methods used included sine-bell or square-sine-bell windows with optimized shifts for every spectrum.

Distance Constraints. After the assignment of most of the proton resonances of napin BnIb, belonging to both the main chain and the side chains, the NOESY spectra obtained with a mixing time of 75 ms were analyzed in order to assign as many dipolar cross-correlations as possible. Those nonoverlapping and unambiguously assigned to a given pair of protons were translated into upper limit distance constraints after visual inspection of their intensity using the following qualitative criteria: high, medium, and low intensity correspond with upper limits of 3.0, 4.0, and 5.0 Å, respectively. No stereospecific assignment of prochiral protons was performed at this stage, and the usual correction for pseudoatoms (Wüthrich, 1986) was added in all cases. In the latest steps of the calculation, hydrogen bond distance constraints were included in segments with well-defined secondary structure for those amide protons showing low exchange rates with solvent deuterons.

Structure Calculation. Distance geometry calculations were performed with the program DIANA (Günter *et al.*, 1991) version 2. The REDAC strategy (Günter & Wüthrich, 1991) was used throughout. The program DIANA works in the torsional space, and the version we have used was not prepared to deal with more than one separate chain. To avoid this limitation, a long segment of 25 pseudoresidues was inserted between the C- and N-termini of the small and large fragments, respectively. These pseudoresidues are modified glycines composed of atoms with null van der Waals radii. The resulting 20 best converged structures were energy minimized with the GROMOS package (van Gunsteren & Berendsen, 1987). All calculations were performed on a Silicon Graphics workstation.

RESULTS

Assignment of the ^1H NMR Spectra. The process of assigning the ^1H NMR resonances of napin BnIb was

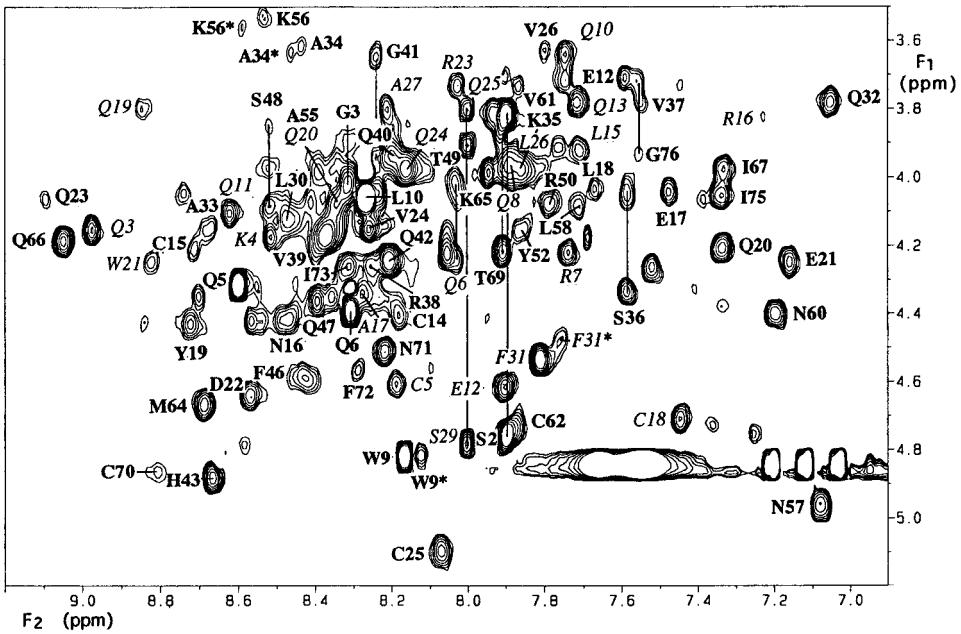

FIGURE 1: Fingerprint region of the 2D-TOCSY spectrum of napin BnIb in H_2O solution, pH 3.0, 20 °C, with assignments marked. Cross-correlations from the small chain are marked in italics. Asterisks mark satellite signals arising from *cis*-Pro forms or other heterogenities.

performed by following the standard sequence-specific method developed by Wüthrich and co-workers (Wüthrich, 1986) on the basis of spectra obtained in H_2O and D_2O and the back-exchanged one in H_2O. Two sets of spectra, one at 20 °C and another at 30 °C, were analyzed in parallel to solve for assignment ambiguities arising from signal overlapping and to detect cross-peaks otherwise occluded by the residual water signal. COSY and TOCSY spectra were first examined in order to detect scalar coupled protons belonging to the same residue. Napin BnIb contains 106 residues, including six Gly and eight Pro, so that a maximum of 103 COSY cross-peaks are expected in the fingerprint region involving the backbone NH−C_αH resonances. In the COSY spectrum in H_2O, 90 backbone cross-peaks were observed. Additional NH−C_αH cross-correlations were observed in the TOCSY spectrum, the fingerprint region of which is shown in Figure 1. In both the COSY and the TOCSY spectra, it could be observed that some intense cross-peaks have a weaker partner cross-peaks of similar chemical shifts (see for instance the cross-correlations for Lys 56 and Ala 34 in the region around $f_1 \sim 3.6$ and $f_2 \sim 8.6-8.4$ in Figure 1). Moreover, the two members of each pair of signals have a very similar pattern of side-chain resonances in the aliphatic region of the TOCSY spectrum, indicating that a double set of signals is present for some residues. In some cases, these signals can be obviously attributed to the small accompanying population of *cis*-Pro isomers, like in the cases of Phe *31* and Trp 9 in the short and long chains, respectively, both of which are found in random-coiled terminal regions (numbers in *italics* will refer throughout to the small chain; the sequence is given in Figure 3). Residues showing a double set of signals are concentrated in the segments Leu 30−Ala 34 and Phe 46−Leu 58. Monsalve *et al.* (1991) have reported the presence of some residue heterogeneities in their determination of the primary sequence of napins BnIa and BnIb. Residues Gln 32 and Ser 48 both in the large chain appear in napins BnIa and BnIb instead of Gly and Ile, which are common in those positions to all other napins. Partner cross-peaks for Gln 32 and Ser 48 have not been found in the proximities of the main signals, a fact that would be expected if a different residue would enter into the sequence. Thus, we think that the species with minor population detected from the spectral analysis may correspond to a residue heterogenity most probably located on residues 32 and 48 or in one of those positions if, as will be shown later on, these segments are spatially proximate. From here on we will concentrate exclusively on the species with a major population.

Resonances belonging to the six alanines, three threonines, and five valines were all easily classified, as well as those from the four serines and nearly all AMX systems (two His, four Phe, two Tyr, two Trp, eight Cys, four Asn, one Asp). Arginine side-chain resonances were identified by the detection (in the TOCSY spectrum) of cross-correlation between the C_αH proton resonances and the NHϵ or C_δHH′ resonances and likewise for three of the five lysines. Most residues of the type Glu, Gln, and Met could also be readily classified at this stage, leaving a few to be identified in the next stage of sequence-specific assignment. Only half of the leucines and isoleucines could be assigned at this step, and all glycines and prolines were assigned in the last stages of the assignment process.

Sequence-specific assignment was based on the NOESY spectra. In Figure 2, the amide−amide region of the NOESY spectrum is shown and the assignment of two helical segments is illustrated. All NN(i,i+1) sequential NOE cross-correlations were observed in the segment *3*−*27* of the small chain with the exception of the one between Arg *16*−Ala *17* and Trp *21*−Ile *22*, in this latter case because of close chemical shifts of their amide protons. The chemical shift of Leu *15* amide proton coincides accidentally within experimental error with that of Gln *13*, which creates a difficulty for the continuity of the assignment. Sequential cross-correlations of the type αN(i,i+1) contributed to overcome this difficulty as well as to assign Arg *16*, to link Trp *21*-Ile *22* and assign the N- and C-terminal pairs Pro *2*-Gln *3* and Pro *30*-Phe *31*. Gln *1* and Ser *29* were assigned from the observation of NOE cross-correlations of their C_αH

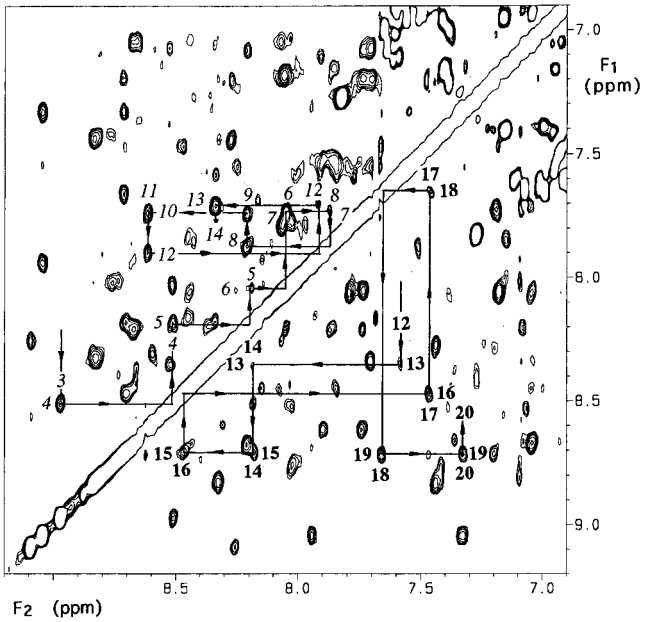

FIGURE 2: Amide−amide resonance region of the NOESY spectrum of napin BnIb in H_2O solution, pH 3.0, 20 °C, with marked sequence-specific assignments. Residues in helix I of the small chain are shown above the diagonal while those of helix II in the large chain are shown below the diagonal.

resonance and the $C_\delta HH'$ resonances of Pro residues next in the sequence. The assignment of Gly *28* is only tentative. It is of note that the His *14* $C_\alpha H$ resonance is upfield shifted by ~3 ppm, an unusually large shift, possibly attributable to its location in the near proximity of the hexad symmetry axis of the phenyl ring plane of Phe *9*.

Assignment of proton resonances of the long chain is hindered because of the large number of Pro residues (six) in its sequence. Detection of NOE cross-correlations with the previous residue, of the type $H_\alpha H_\delta$ ($i-1,i$) an indication of X-Pro *trans* peptide bonds, is not always possible because of severe signal overlapping in that region. They were clearly observed for the pairs Cys 27-Pro 28 and Cys 70-Pro 71. The detection of intense NOE cross-correlations between Pro 59 $C_\alpha H$-Leu 58 NH and Leu 58 NH-Asn 60 NH indicates the probable presence of the peptide bond Leu 58−Pro 59 in a *cis* conformation. No sequential connections were found for Gly 41 and Gln 53. A set of signals corresponding to a glycine residue in which the amide chemical shift is close to those of Gln 40 and Gln 42 was tentatively assigned to Gly 41. Gln 53 was assigned on indirect evidences.

In summary, an almost complete assignment has been performed for the ^1H resonances of napin BnIb. Out of the 106 residues, 99 have been assigned unambigously. The assignments of two of them (Gly 41 and Gln 53 in the large chain) are only tentative and those of Gly *28* (small chain) and Gly 2, Gly 7, Arg 11, and Gln 40 (large chain) though being also tentative are in the same way highly probable. A table, included in the Supporting Information, lists the performed assignments.

Secondary Structure Analysis. Secondary structure was delineated from the analysis of observed NOE patterns together with NH-exchange data and a qualitative interpretation of $C_\alpha H$ conformational shifts (Jiménez *et al.*, 1987; Wishart *et al.*, 1991). Figure 3 shows a summary of observed sequential and short-range NOEs with an indication of their

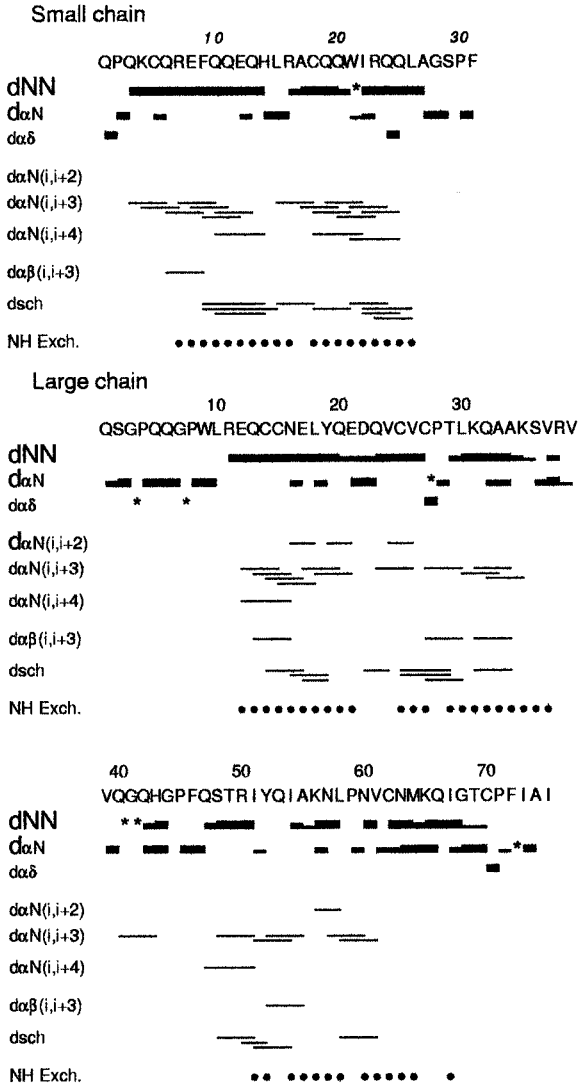

FIGURE 3: Summary of observed sequential and short-range NOEs. The thickness of the sequential NOEs is a qualitative indication of their intensity. Filled circles mean high protection of backbone amide against exchange with solvent deuterons. Asterisks mean unobservable cross-peaks because of signal overlapping. For the definition of $d_{\alpha N}$, d_{NN}, etc., see Abbreviations.

intensity in the first case. The exchange properties of the backbone amide proton resonance are also illustrated in the figure. The conformational shifts, $\Delta\delta_\alpha = \delta_\alpha{}^{obs} - \delta_\alpha{}^{RC}$, where δ^{RC} are the shifts corresponding to the random coil (Bundi & Wüthrich, 1979), are illustrated in Figure 4. From the consideration of Figures 3 and 4, we can deduce that napin BnIb is fundamentally an all-helix protein with regions involving loops or turnlike conformations. The light chain contains two short helices separated by a few residues. In terms of the intensities of d_{NN} NOE cross-peaks, the two helices would span respectively Gln *3*−His *14* (helix I) and Arg *16*−Ala *27* (helix I′), whereas according to the presence of short-range NOE cross-correlations, the helical segments would be *3*−*13* and *16*−*25*, in agreement with the previous result. From here on, we will consider the segment *3*−*27* as a single helix split into two smaller helices, in order to facilitate the comparison of the structure of napin BnIb with those of other proteins with a similar fold. The location of protected protons as well as the observation of negative conformational shifts support in a general way the presence of two helical conformations in this light chain and show a

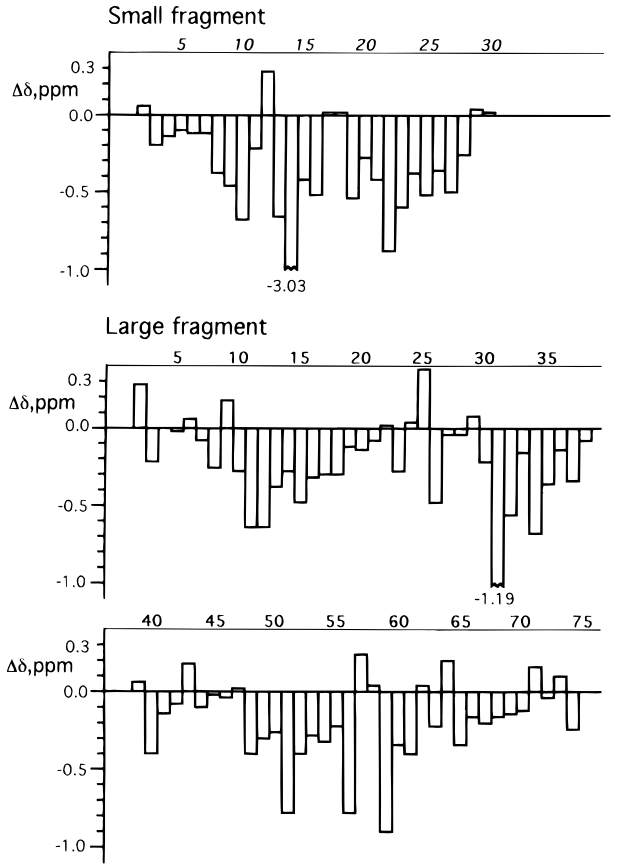

FIGURE 4: Conformational chemical shifts of proton resonances in napin BnIb, $\delta^{conf} = \delta^{obs} - \delta^{RC}$. The δ values of the random coil, δ^{RC}, were taken from Bundi and Wüthrich (1979).

broad agreement with the limits determined above. It is of note that the ring current effects of the aromatic rings of Phe *9*, His *14*, and Trp *21* may affect the magnitude and sign of the conformational shifts of nearby proton resonances (see for instance the abnormally large up-field shift of His *14*).

A similar joint analysis of conformational shifts, exchange data, and sequential and short-range NOEs lead to the conclusion that the first helix in the heavy fragment (helix II) must be placed between residues Arg 11 and Glu 21 (Figures 3 and 4). Similarly, the next helix (helix III) would span Gln 23−Val 37, according mainly to the NOE and chemical exchange data. The chemical shifts of $C_\alpha H$ protons at the begining of this helix must be affected by tertiary effects, since not all of them show the expected negative shifts. A last helix (helix IV), much shorter, appears to be present in the segment Gln 47−Ala 55 as judged by short-range NOEs (Figure 3) and supported by conformational chemical shifts (Figure 4).

Finally, data from the C-terminal segment comprising residues Lys 56−Cys 70 are difficult to attribute to the usual elements of secondary structure. Conformational chemical shifts show different signs and are not homogeneously distributed. Also, a number of *i*, *i* + 2 and *i*, *i* + 3 NOEs are present between residues Lys 56 and Gln 66, a segment in which eight amide protons (57−64) appear as strongly protected presumably by hydrogen bonds. The presence of long-range NOEs between Asn 57 and Thr 69 as well as between Val 61 and Phe 72 makes it impossible for this segment to be in a helical conformation. Furthermore, this chain is fixed to other points in the molecule by disulfide bridges involving Cys 62 and Cys 70. These features may be accounted for if it is supposed that the segment forms a large loop with no periodic structure involving a series of turnlike conformations in the region Asn 57−Gln 66, which would be favored by the *cis* configuration of the Leu 58−Pro 59 peptide bond.

Tertiary Structure Determination. We attempted to calculate an unrefined tertiary structure, from which we could determine the global shape and overall main-chain fold of the protein. A total of ∼500 interproton upper distance constraints (180 intraresidue, 180 sequentials, 60 short range, 80 long range), corresponding to nonoverlapping and unambiguously assigned cross-correlations, were used in the calculation, as described in Materials and Methods. Thirty-five long-range distance constraints correspond to interchain protons. In principle, the arrangement of disulfide bridges determined experimentally for one member of the family of 2S albumins (Lilley & Inglis, 1986; Nirasawa *et al.*, 1993) was accepted. The corresponding restraints (lower/upper limits for $C_{\beta i}-S_j$ and S_i-S_j distances of 3.0/3.1 and 2.0/2.1 Å, respectively) involving the atoms intervening in the four disulfide bridges were then included in the DIANA calculation. We recognize that the number of restraints, particularly those between protons distant in the sequence is rather low for determining a well-defined structure. Nevertheless, we believe that the resulting overall backbone fold is reliable as demonstrated by its reproducibility in independent calculations. A more complete analysis of NOESY spectra involving cycles of NOE assignments and structure computations is in progress, although a more refined structure may be difficult to achieve because of the relatively low number of long-range NOE cross-correlations in all-helix-type proteins.

A family of 25 structures was calculated by using the program DIANA and the REDAC strategy. The average pairwise RMSD for the backbone heavy atoms of these structures is 3.5 Å, excluding residues in random conformations, i.e., those in N- and C-terminal ends like residues *27*−*31* in the small chain and 1−10 and 72−75 in the large one and those in the loop 39−47. The best 10 structures of that set were selected and are shown superimposed in Figure 5, where the chains as well as the entire protein are illustrated separately. The pairwise RMSD for these 10 structures is 2.5 Å. The computed structures are fully compatible with the resonance and NOE cross-correlation assignments. The maximum violation was 1.2 Å, the average sum of violations for the whole set of distance constraints was 24 Å, and the average number of violations larger than 0.35 Å was 4. An average structure was obtained from the 10 selected conformations, which was subjected to restrained energy minimization with the GROMOS package (van Gunsteren & Berendsen, 1987). The final structure is illustrated in Figure 6 (top). Most of the features of secondary structure elements proposed earlier from short-range NOEs, $C_\alpha H$ chemical shifts, and slowly exchanging amide proton patterns can be readily observed in Figure 6. The light chain is split in two helices (helices I and I′) joined by an extended strand in the middle two to four residue long around which the fragment bends to fit the bundle formed by helices III and IV, which runs almost orthogonally to it. The C-terminus of the light chain is relatively close in space to the N-terminus of the heavy chain, to which it may be linked at some stage in the protein expression.

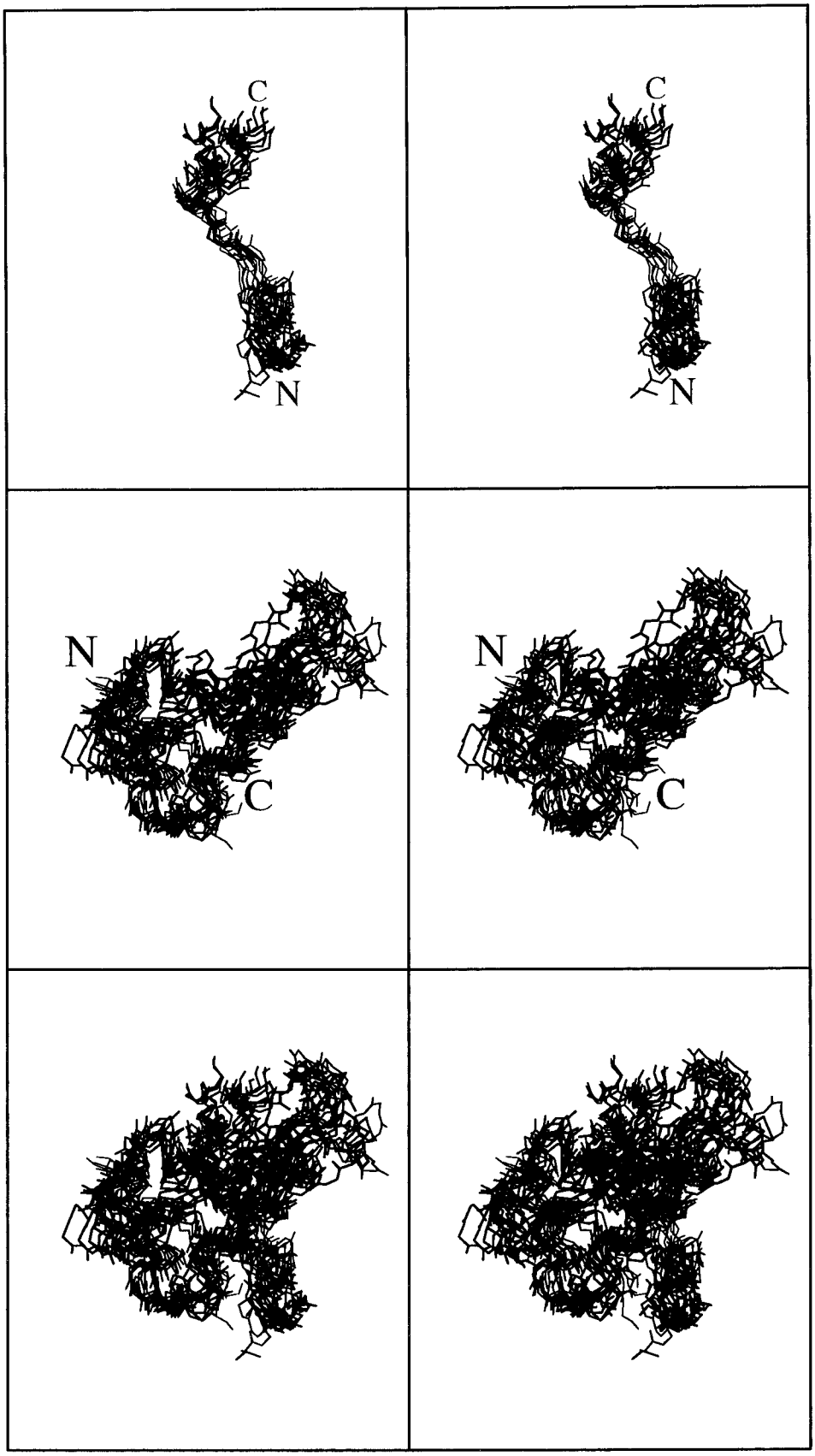

FIGURE 5: Best superposition of 10 converged structures resulting from the application of the DIANA program (Günter *et al.*, 1991). Top, small chain; middle, large chain; bottom, the complete protein.

Helix II, the first helix of the heavy chain, runs antiparallel to helices I and I′ followed by a short loop and helix III, the helix (now spanning residues 26−37) that, in a way, serves as the protein spine since it anchors the N-terminus of the light chain and the C-terminus of the heavy chain through the disulfide bridges (Cys 25−Cys5 and Cys 27 and Cys 70, respectively. Between helix III and helix IV there is a relatively short loop (about eight residues) which is much

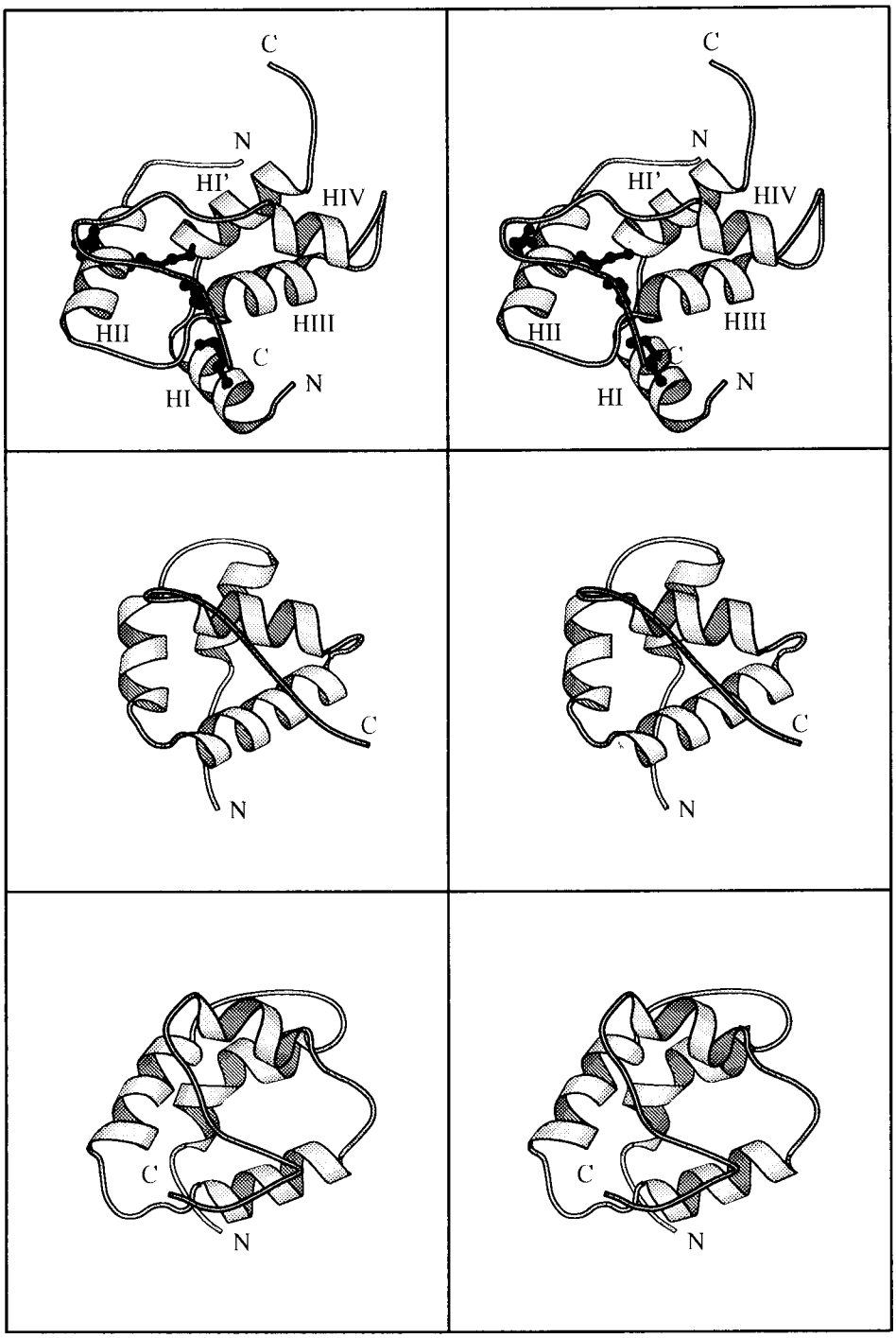

FIGURE 6: Stereoscopic views of (top) napin BnIb (PDB entry 1PNB), (middle) HPS (PDB entry 1hyp), and (bottom) LTP (PDB entry 1lpt). Views have been selected by superimposing the elements of secondary structure.

longer and also very rich in glutamines in napins BnII, BnIII, and BnIV (Monsalve & Rodríguez, 1990). Finally, after helix IV, the shortest helix present, we find a long loop without any standard secondary structure which folds over itself. The additional disulfide bridges join the beginning of helix I′ in the short chain with the middle of helix II (residues Cys *18* and Cys 14, respectively) and the next position to this latter one with the middle position of the C-terminal large loop in the heavy chain (residues Cys 15 and Cys 62). The arrangement of the different consecutive helices and loops corresponds to a right-handed helical disposition.

Comparison with Other Proteins with a Similar Fold. The global fold found for napin BnIb is very similar to those determined for the HPS protein (Baud *et al.*, 1993) and ns-LTP from wheat (Gincel *et al.*, 1994) and maize (Shin *et al.*, 1995; Gomar *et al.*, 1996) though they differ in a number of structural features, as discussed below. As said in the previous section, the disulfide bridge organization of napin BnIb has been assumed to be the same as that for other 2S albumins, for which an experimental determination of the location of the disulfide bridges has been carried out (Lilley & Inglis, 1986; Nirasawa *et al.*, 1993). The found organization is identical to that described for HPS (Odani *et al.*, 1987) but slightly different from that shown by the ns-LTPs obtained from plants (Gincel *et al.*, 1994; Shin *et al.*, 1995; Gomar *et al.*, 1996) (see Figure 7). In fact, in this latter protein instead of the bridges corresponding to the ones

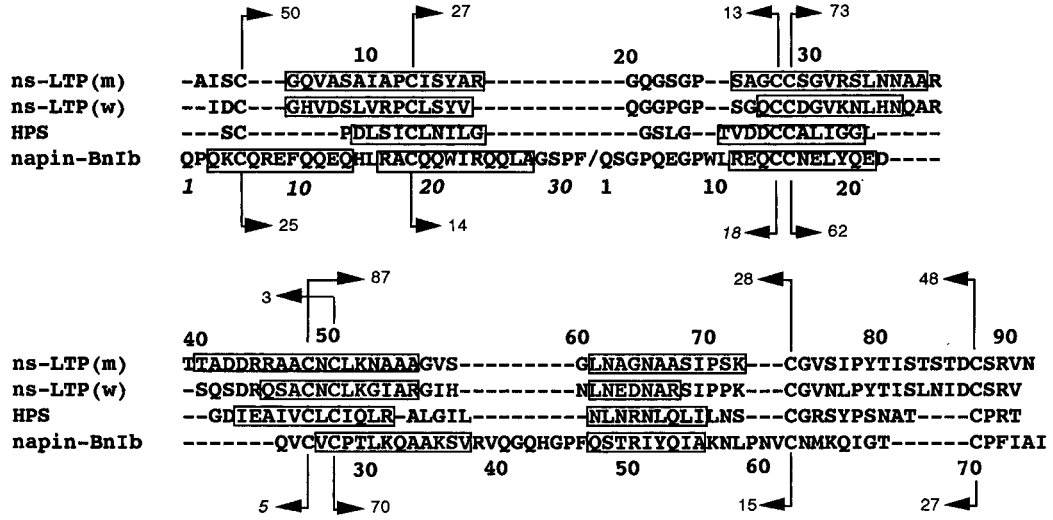

FIGURE 7: Sequence alignment of napin BnIb, HPS, and wheat (w) and maize (m) ns-LTPs. The location of helical segments and cysteine bridges is marked for the different proteins: maize and wheat ns-LTP share the top disulfide organization and HPS and napin BnIb the one in the bottom.

pertinent to napin BnIb (Cys *5*−Cys 25 and Cys 27−Cys 70) the alternative arrangement Cys *5*−Cys 27 and Cys 25−Cys 70 is found. A DIANA calculation was run without introducing the distance restraints corresponding to these two disulfide bridges. The results confirm that the accepted disulfide arrangement is the correct one: distances S−S between Cys *5*−Cys 25 and Cys 27−Cys 70 were 5.4 ± 2.0 and 2.7 ± 1.0, respectively, against 10.9 ± 3.5 and 8.3 ± 2.8 found for the alternative arrangement, Cys *5*−Cys 27 and Cys 25−Cys 70.

In the segment analogous to napin BnIb small chain, the proteins HPS and ns-LTP present a single helix which incorporates the second cysteine residue and shows the typical proline kink in the two ns-LTPs. Napin BnIb shows in this segment two well-defined helices each incorporating a cysteine residue at the N-terminus. In spite of the differences found between napin BnIb and the two other proteins, the two cysteines involved in this segment are conveniently placed so as to form the two interchain disulfide bridges in much the same way.

Helix II and helix III incorporate respectively the motifs −CC− and −CXC− and present a similar length in all four proteins (see Figure 7). In napin BnIb the motif −CXC− is placed at the N-terminal position of helix III, which differs from that found in the other three proteins where the motif appears well in the middle of the helix. Helix IV is the shortest helix in napin BnIb, and its length and location in relation to the seventh cysteine are more like that in wheat ns-LTP than in the other two proteins (Figure 7). As mentioned before, there are large differences in the length of peptide stretches between cysteine residues in napin BnIb and the other three proteins, the differences being least between napin BnIb and HPS. Thus, for instance, there are nine residues in napin BnIb between the motifs −CC− and −CXC−, whereas in HPS and the nsLTPs there are respectively 13 and 19. Between the motif −CXC− and the next cysteine, napin BnIb shows 34 residues against 21 and 22 in HPS and the ns-LTPs, respectively (see Figure 7). This difference increases greatly in the large molecular weight napins where up to 12 additional residues are incorporated in the loop joining helices III and IV corresponding to this segment.

The most outstanding feature in the structures of wheat and maize ns-LTP, both in solution and in the crystal state, is an internal hydrophobic cavity of 270−300 Å^3 running through the whole molecule and able to accommodate the aliphatic chains of fatty acids or lysolecithins. In contrast, there does not appear to be an internal cavity in HPS crystal structure (Baud *et al.*, 1993), although Gomar *et al.* (1996) claimed that, under close examination, a hydrophobic tunnel of roughly the same volume can be observed in HPS. We have not detected an internal cavity in napin BnIb by using the program VOIDOO (Kleywegt & Jones, 1994), but we realize that proper detection of the maximum volume cavity would require extensive rearrangements of the side chains pointing inward in the central core, which we have not attempted. This fact together with the rough quality of our converged structures led us to further examine the possibility of the existence of a cavity in napin BnIb. To that end, a model calculation was carried out in which the backbone atoms of the residues involved in the four main helices of napin BnIb were forced to reach an optimum superposition to those of maize ns-LTP. A more expanded structure was generated in that way, but the pairing of the disulfide bridges was changed to the one corresponding to the ns-LTP proteins. Also, a large number of long-range upper limit distance constraints were violated, thus indicating that the expanded structure does not satisfy experimental observations. Also, the result of this calculation points toward a possible relationship between disulfide pairing and the existence of an internal cavity, which bears an important implication in relation to the functions of these proteins. A more definite answer to this question would have to wait until a more refined three-dimensional structure of napin BnIb is available.

Distribution of Polar and Hydrophobic Residues. Napin BnIb has 15 positive and 8 negative charges at neutral pH, including the two N- and C-termini. Nine positive charged side chains (Arg, Lys, and His residues) and four negative (Asp and Glu residues) are conserved in all known low (BnIa, BnIb) and high (BnII, BnIII, BnIV) molecular weight napins. All conserved charged residues have high values of fractional accessible surface areas (ASA) as calculated with the program VADAR (Wishart & Willard, 1993), with the exception of Lys 31, which is located at the center of the

internal helix III in the large chain and shows a relatively low fractional ASA value of 0.18. This is the case also for His 43, a residue that is not conserved as a positively charged one in the family of napins and that shows an ASA value of 0.13. The distribution of conserved charges is not uniform. Positive charges are mainly located in helices III and IV and in the loop that links them, whereas negative charges cluster in the region of the loop between helices II and III with participation of a further glutamic acid coming out from helix I.

Polar residues including the overabundant glutamines are in general exposed to the solvent with fractional ASA values in the range 0.31−0.88. Main exceptions are Gln *25*, at the C-end of helix I′ in the short chain, Thr 29 in the middle of helix III in the long fragment, which is almost completely buried, and the residues Lys 56, Asn 57, and Asn 60, which are part of the C-terminal loop. Cysteine residues are all buried with an average ASA value of ∼0.09, a fact in keeping with that observed in HPS and ns-LTP proteins. Hydrophobic residues are in general buried in the interior of the protein with fractional ASA values in the range 0.00−0.39 and with a mean value of 0.14.

Thus, although a further refinement of the structure is needed in order to assess whether the found differences are meaningful, it is clear that the large majority of results concerning the relationship between polarity and solvent-exposed areas for the different residues in napin BnIb are in satisfactory agreement with expectation.

DISCUSSION

Resonance Assignment, Secondary Structure and Global Fold of Napin BnIb. The wide polymorphism exhibited by NT2SAs has hindered up to the present time the isolation of homogeneous and well-defined isoforms, a prerequisite for the preparation of good diffracting crystals as well as performance of an unambiguous assignment of their NMR spectra. Consequently, a three-dimensional structure of a member of this important and widespread family of seed proteins was lacking. The lowest molecular weight napins, BnI napins (12.0 kDa), do not exhibit such a high polymorphism and can be resolved into two proteins, BnIa and BnIb (Monsalve *et al.*, 1991). The sequence homology among the 12-kDa napins and those of high molecular mass (BnII, BnIII, and BnIV, 14 kDa), ∼55% of identity (Monsalve *et al.*, 1991), and the identical circular dichroism spectra (Menéndez-Arias *et al.*, 1987; Monsalve *et al.*, 1991, 1994), suggest that all napins and homologous 2S albumins have a similar three-dimensional structure. As shown in the previous sections, we have been able to perform an almost complete assignment of the ^{1}H resonances of BnIb, the main component of the low molecular weight 2S albumin napins. The assignment provides, as a first important result, a delineation of the main elements of secondary structure. Also, on the basis of distance constraints translated from unambiguously assigned NOE cross-correlations we were able to determine the overall shape and global fold of the molecule. In spite of the short number of distance constraints used in the determination of the three-dimensional structure of napin BnIb, the essential aspects of the folding have been unquestionably established. Thus, an identical fold results from a number of independent distance geometry calculations, and a very appropriate distribution of polar and hydrophobic residues between the surface and the core results from the calculation.

Comparison of BnIb, HPS, and ns-LTP Structures. The structure determined for napin BnIb represents a third example of a new and distinctive folding pattern first described for HPS (Baud *et al.*, 1993) and found also in wheat (Gincel *et al.*, 1994) and maize (Shin *et al.*, 1995; Gomar *et al.*, 1996) ns-LTPs. Up to the present work, no relationship had been established between plant ns-LTPs or HPS and NT2SAs from seeds. HPS, and wheat and maize ns-LTPs, are proteins that consist of a single polypeptide chain of 80, 90, and 93 amino acid residues, respectively, and their molecular weights are consequently smaller (9000−10 000) than those of napins. As noted in previous sections, the most conspicuous similarity between all these three proteins is the sequence distribution of their eight Cys residues and the identical organization of the four disulfide bridges in napin BnIb and HPS, which in turn differ slightly from that found in the LTPs.

It is difficult to perform an alignment of the sequence of BnIb with those of HPS and ns-LTP because of the different lengths of the segments between the cysteine residues. A routine application of some of the alignment programs provides a very low level of identity (∼20%), which is in the border of significant homology (Doolittle, 1981). And yet the global fold of the three proteins, consisting of four helices and a C-terminal fragment connected by loops and arranged in a right-handed spiral, is very similar as is the location of the four disulfide bridges which stabilize their architecture. This could suggest the evolution of a great number of seed proteins from a common polypeptide ancestor, which diverged to different functional activities but conserved a common and stable conformation.

The main difference between the structure of napin BnIb and those of HPS and ns-LTPs derives from the fact that napin BnIb consists of two polypeptide chains, whereas the two latter proteins are formed by one single polypeptide chain. The three-dimensional structure of BnIb shows the C-terminal end of the small chain and the N-terminal from the large one in spatial proximity, very exposed to the solvent—as Muren *et al.* (1995) had suggested—and without marked conformational restrictions. These features are in close agreement with the proposal that a long loop existing in the precursor accessible to maturation proteolytic enzymes is cleaved during translation (Crouch *et al.*, 1983; Hara-Nishimura *et al.*, 1991; Scott *et al.*, 1991; D'Hont *et al.*, 1993). They would also allow the N- and C-terminal exoproteases of the seed to digest portions of the chains during maturation to render the partly frayed ends of the NT2SAs (Muren *et al.*, 1995; Gehrig *et al.*, 1996).

Structure−Function Relationship. The activity of maize and wheat ns-LTPs as lysolecithin or fatty acid carrier proteins has been sufficiently documented, and even the structure of the complex between maize ns-LTP and palmitate has been determined at 1.8-Å resolution (Shin *et al.*, 1995). Essential structural features for that activity are the presence of an internal cavity able to host the lipid molecules as well as a number of hydrophobic side chains pointing inward and in particular a Tyr residue (Tyr 79 in wheat and Tyr 81 in maize) in the C-terminal loop that binds the carboxylate group of palmitate. HPS, the other protein with a fold similar to that of napin BnIb, also shows the existence of a hydrophobic tunnel (as analyzed by Gomar *et al.* (1996))

and a Tyr residue (Tyr 71) in the C-terminal loop. However, this Tyr residue is exposed to the solvent (ASA value 0.42 as opposed to nearly 0.00 in ns-LTPs) and potentially unable to form a hydrogen bond with the polar head of the lipid molecule. This fact, together with the absence of a polar cluster of charged residues in the loop between helices II and III able to stabilize the lipid−protein complex, rules out HPS as a carrier of lipids. The fact that an internal cavity is not apparent in the structure of napin BnIb as well as the absence of a buried tyrosine in the C-terminal segment point toward a conclusion similar to that arrived at with HPS. As mentioned in the introduction, a number of activities have been demonstrated for 2S albumin proteins. These will be discussed below in terms of the determined three-dimensional structure.

It has been proposed that the antifungal activity of the NT2SAs would involve the permeabilization of the fungal plasmalemma close to the hyphal tip (Terras *et al.*, 1993). As a strong interaction with phospholipid bilayers has been demonstrated for the NT2SA obtained from mustard (Oñaderra *et al.*, 1994), it is reasonable to suggest that the permeabilization process could be promoted by a perturbation of the lipid membrane, induced by the NT2SAs when they interact with the fungal bilayer. It is generally accepted that the first steps in the polypeptide−membrane interactions include electrostatic contacts between basic protein residues and acidic phospholipids. Napins are very basic (pI > 9.5), and as can be deduced from our structural study, BnIb exhibits most of the Lys and Arg side chains projecting toward the solvent. Most of the acidic residues (Glu *8* and Glu *12* of the small chain; Glu 17, Glu 21, and Asp 22 of the large chain) are located very close to each other in the three-dimensional structure of the protein, and this leaves the rest of the molecule with an overall external positive charge aiding in protein−membrane electrostatic interactions.

The 14.5-kDa NT2SAs contain a stretch of high sequence variability (Krebbers *et al.*, 1988; Raynal *et al.*, 1991; Monsalve *et al.*, 1991; Gehrig *et al.*, 1996), between positions 38 and 39 of BnIb, which has been used to insert within it active peptides without disturbing the biological functions of these proteins (Vandekerckhove *et al.*, 1989). The structure determined here shows that this strectch is in fact an extension of the turnlike conformational segment connecting helices III and IV, so that very likely the folding process and the core structure will be independent of its length, as suggested by Krebbers *et al.* (1988).

A calmodulin antagonist function has been found for the free small chain of NT2SAs (Polya *et al.*, 1993). Since both calmodulin and these small chains contain a similar α-helix−hinge−α helix motif in their structure, a competitive inhibitory action by binding to the calmodulin protein targets could be suggested.

Many 2S albumins, and related proteins such as α-amylase and trypsin inhibitors, have been described as allergenic components of the seeds. Knowledge of the three-dimensional structure is an essential tool for the antigenic mapping of allergens, mainly of those containing conformational determinants and disulfide bridges that are not usually amenable for analysis with synthetic peptides. Three-dimensional structures of a few number of allergens, such as mellitin (Terwilliger & Eisenberg, 1982) or Amb t V and Amb a V (Metzler *et al.*, 1992a,b) have been determined. With BnIb, a representative member of the NT2SAs, the spectrum of small allergens with elucidated structures is increased, allowing new studies on the mechanisms for inducing allergy.

Finally, and taking into account the similarity between the BnIb and other NT2SAs from different species, the modeling of a high number of conformational structures of known sequences can now be performed, which will certainly help to better understand their structure−function relationships.

ACKNOWLEDGMENT

We thank Apolo Gómez, Cristina López, and Luis de la Vega for excellent technical assistance. We also thank Domingo García for his valuable help in all performed calculations. We are indebted to Prof. Marius Ptak (Université d'Orléans, Orléans, France) for sending us the Cartesian coordinates of wheat ns-LTP before they were available from the PDB.

SUPPORTING INFORMATION AVAILABLE

Table including the assignment of proton resonances of napin BnIb (4 pages). Ordering information is given on any current masthead page.

REFERENCES

Altenbach, S. B., Kuo, C.-C., Staraci, L. C., Pearson, K. W., Wainwright, C., Georgescu, A., & Townsend, J. (1992) *Plant. Mol. Biol. 18*, 235−245.

Aue, W. B., Bartholdi, E., & Ernst, R. R. (1976) *J. Chem. Phys. 5*, 2229−2246.

Baud, F., Pebay-Peroula, E., Cohen-Addad, C., Odani, S., & Lehmann, M. S. (1993) *J. Mol. Biol. 231*, 877−887.

Bax, A., & Davies, D. G. (1985) *J. Magn. Reson. 65*, 355−360.

Bundi, A., & Wüthrich, K. (1979) *Biopolymers 18*, 285−297.

Cocucci, M., & Negrini, N. (1988) *Plant Physiol. 88*, 910−914.

Crouch, M. L., Tenbarge, K. M., Simon, A. E., & Ferl, R. (1983) *J. Mol. Appl. Genet. 2*, 273−283.

D'Hont, K., van Damme, J., van den Bossche, C., Leejeerajumnean, S., De Rycke, R., Derksen, J. Vanderkerckhove, J., & Krebbers, E. (1993) *Plant Physiol. 102*, 425−433.

Doolittle, R. F. (1981) *Science 214*, 149−159.

Ericson, M. L., Rödin, J., Lenman, M., Glimelius, K., Josefsson, L.-G., & Rask, L. (1986) *J. Biol. Chem. 261*, 14576−14581.

Gehrig, P. M., Krzyzaniak, A., Barciszewski, J., & Biemann, K. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 3647−3652.

Ginzel, E., Simovre, J. P., Caille, A., Marion, D., Ptak, M., & Vovelle, F. (1994) *Eur. J. Biochem. 226*, 413−422.

Gomar, J., Petit, M. C., Sodano, P., Sy, D., Marion, D., Kader, J. C., Vovelle, F. & Ptak, M. (1996) *Protein Sci. 5*, 565−577.

Günter, P., & Wüthrich, K. (1991) *J. Biomol. NMR 1*, 447−456.

Günter, P, Braun, W., & Wüthrich, K. (1991) *J. Mol. Biol. 217*, 517−530.

Hara-Nishimura, I., Inove, K., & Nishimura, M. (1991) *FEBS Lett. 294*, 89−93.

Jiménez, M. A., Nieto, J. L., Herranz, J., Rico, M., & Santoro, J. (1987) *FEBS Lett. 221*, 320−324.

Kleywegt, G. J., & Jones, T. A. (1994) Program VOIDOO, Dept. of Molecular Biology, Biomedical Center, University of Uppsala, Uppsala, Sweden.

Krebbers, E., Herdies, L., De Clercq, A., Seurinck, J., Leemans, J., van Damme, J., Segura, M., Gheysen, G., van Montagu, M., & Vandekerckhove, J. (1988) *Plant Physiol. 87*, 859−866.

Kumar, A., Ernst, R. R., & Wüthrich, K. (1990) *Biochem. Biophys. Res. Commun. 95*, 1−6.

Lilley, G. G., & Inglis, A. S. (1986) *FEBS Lett. 195*, 235−241.

Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.

Menéndez-Arias, L., Monsalve, R. I., Gavilanes, J. G., & Rodríguez, R. (1987) *Int. J. Biochem. 19*, 899−907.

Menéndez-Arias, L., Moneo, I., Domínguez, J., & Rodríguez, R. (1988) *Eur. J. Biochem. 177*, 159−166.
Metzler, W. J., Valentine, K., Roebber, M., Friedrichs, M. S., Marsh, D. G., & Mueller, L. (1992a) *Biochemistry 31*, 5117−5127.
Metzler, W. J., Valentine, K., Roebber, M., Friedrichs, M. S., Marsh, D. G., & Mueller, L. (1992b) *Biochemistry 31*, 8697−8705.
Monsalve, R. I., & Rodríguez, R. (1990) *J. Exp. Bot. 41*, 89−94.
Monsalve, R. I., López-Otín, C., Villalba, M., & Rodríguez, R. (1991) *FEBS Lett. 295*, 207−210.
Monsalve, R. I., González de la Peña, M. A., Menéndez-Arias, L., López-Otín, C., Villalba, M., & Rodríguez, R. (1993) *Biochem. J. 293*, 625−632.
Monsalve, R. I., Menéndez-Arias, L., González de la Peña, M. A. Batanero, E., Villalba, M., & Rodríguez, R. (1994) *J. Exp. Bot. 45*, 1169−1176.
Muren, E., Ek, B., & Rask, L. (1995) *Eur. J. Biochem. 227*, 316−321.
Nirasawa, S., Liu, X., Nishino, T., & Kurihara, Y. (1993) *Biochem. Biophys. Acta 1202*, 277−280.
Odani, S., Takehiko, K., Ono, T., Seto, Y., & Tanaka, T. (1987) *Eur. J. Biochem. 162*, 485−491.
O'Neil, K. T., & DeGrado, W. F. (1989) *Proteins 6*, 284−293.
Oñaderra, M., Monsalve, R. I., Mancheño, J. M., Villalba, M., Martínez del Pozo, A., Gavilanes, J. G., & Rodríguez, R. (1994) *Eur J. Biochem. 225*, 609−615.
Persechini, A., & Kretsinger, R. H. (1988) *J. Biol. Chem. 263*, 12175−12178.
Piotto, M., Saudek, V., & Sklenar, V. (1992) *J. Biomol. NMR 2*, 661−665.
Polya, G. M., Chandra, S., & Condron, R. (1993) *Plant Physiol. 101*, 545−551.
Raynal, M., Depigny, D., Grellet, F., & Delseny, M. (1991) *Gene 99*, 77−86.
Redfield, A. G., & Kuntz, S. D. (1975) *J. Magn. Reson. 19*, 250−259.
Scott, M. P., Jung, R., Muntz, K., & Nielsen, N. (1991) *Proc. Natl. Acad. Sci. U.S.A. 89*, 658−662.
Sharief, S. F., & Li, S. S.-L. (1982) *J. Biol. Chem. 257*, 14753−14759.
Shewry, P. R., Napier, J. A., & Tatham, A. S. (1995) *Plant Cell 7*, 945−956.
Shin, D. H., Lee, J. Y., Hwang, K. Y., Kim, K. K., & Suh, S. W. (1995) *Structure 3*, 189−199.
Takishima, K., Watanabe, S., Yamada, M., & Mamiya, G. (1986) *Biochem. Biophys. Acta 870*, 248−255
Terras, F. R. G., 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.
Terras, F. R. G., Torrekens, S., van Leuven, F., Osborn, R. W., Vanderleyden, J., Cammue, B. P. A., & Broekaert, W. F. (1993) *FEBS Lett. 316*, 233−240.
Terwilliger, T. C., & Eisenberg, D. (1982) *J. Biol. Chem. 257*, 6016−6022.
Vandekerckhove, J., van Damme, J., van Lijsebettens, M., Botterman, J., De Block, M., Vandewiele, M., De Clercq, A., Leemans, J., van Montagu, M., & Krebbers, E. (1989) *Biotechnology 7*, 929−932.
van Gunsteren, W. F., & Berendsen, H. J. C. (1987) Groningen Molecular Simulations (GROMOS) Library Manual, Biomos, Groningen, The Netherlands.
Wishart, D. S., & Willard, L. (1993) Program VADAR, Protein Engineering Network of Centres of Excellence, University of Alberta, Alberta, Canada.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) *J. Mol. Biol. 222*, 311−333.
Wüthrich, K. (1986) in *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
Youle, R. J., & Huang, A. H. (1978) *Plant Physiol. 61*, 13−16.

BI961748Q