# Solution Structure of RicC3, a 2S Albumin Storage Protein from *Ricinus communis*[†,‡]

David Pantoja-Uceda,[§] Marta Bruix,[§] Guillermo Giménez-Gallego,[||] Manuel Rico,[§] and Jorge Santoro*[,§]

*Instituto de Química Física Rocasolano, CSIC, Serrano 119, Madrid 28006, Spain, and Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, Madrid 28040, Spain*

*Received July 10, 2003; Revised Manuscript Received September 22, 2003*

ABSTRACT: The three-dimensional structure in aqueous solution of recombinant $^{15}$N labeled RicC3, a 2S albumin protein from the seeds of castor bean (*Ricinus communis*), has been determined by NMR methods. The computed structures were based on 1564 upper limit distance constraints derived from NOE cross-correlation intensities measured in the 2D-NOESY and 3D-HSQC−NOESY experiments, 70 $\phi$ torsion angle constraints obtained from $^{3}J_{HNH\alpha}$ couplings measured in the HNHA experiment, and 30 $\psi$ torsion angle constraints derived from $^{3}J_{H\alpha Ni+1}$ couplings measured in the HNHB experiment. The computed structures showed a RMSD radius of 0.64 Å for the structural core. The resulting structure consists of five amphipatic helices arranged in a right-handed super helix, a folding motif first observed in nonspecific lipid transfer proteins. Different than the latter, RicC3 does have not an internal cavity, a fact that can be related to the exchange in the pairing of disulfide bridges in the segment ...CXC.... Previous attempts to determine high resolution structures of a 2S albumin protein by either X-ray crystallography or NMR methods failed because of the heterogeneity of the protein prepared from natural sources. Both 2S albumins and nonspecific lipid transfer proteins belong to the prolamine superfamily, some of whose members are food allergens. The solution structure for recombinant RicC3 determined here is a suitable representative structure for the broad family of seed 2S albumin proteins, which may help to establish meaningful relationships between structure and allergenicity. RicC3 is also the peptidic component of the immunomodulator Inmunoferon, a widely used pharmaceutical product, and its structure is expected to help understand its pharmaceutical activity.

The 2S albumins are a family of storage proteins belonging to the prolamin superfamily (*1*) that includes cereal seed inhibitors of α-amylase and/or trypsin, puroindolines, nonspecific lipid transfer proteins (nsLTPs),[1] the hydrophobic protein from soybean, and some cell wall glycoproteins. They are small proteins (12−15 kDa) generally composed of two different polypeptide chains linked by two disulfide bridges. Like other members of the superfamily, the 2S albumin proteins are widely distributed in plant seeds and show a conserved motif of eight cysteine residues spaced as follows ...C...C.../...CC...CXC...C...C.... In addition to the two interchain disulfide bridges, they have two other intrachain disulfide bonds, which make these proteins very stable and compact. Due to their amino acid composition, their high content in the protein bodies of seeds, and their mobilization during germination, these proteins have been proposed to act as nitrogen and sulfur donors (*2*). However, other functions have been ascribed to them, including antifungal activity (*3*), serine protease inhibition (*4*), and calmodulin antagonism (*5*). In addition to their biochemical interest, 2S albumins have been used as carriers for the synthesis of biologically active peptides (*6*), as well as for improving the nutritional properties of grain crops by increasing their content of essential amino acids (*7*).

Albumin 2S from castor bean (*Ricinus communis*, Euphorbiaceae) is encoded by a single mRNA whose translation yields two proteins, RicC1 (11.2 kDa) and RicC3 (12.0 kDa) (*8*). Each of these proteins is composed of a small and a large subunit linked together by disulfide bridges. The high-yield synthesis of the 2S albumin RicC3 using genetically engineered *Escherichia coli* grown in defined culture media has been recently reported (*9*), and this system has been used to prepare $^{15}$N labeled recombinant RicC3 using a minimal medium with $^{15}NH_4Cl$ as the only nitrogen source. The resulting protein conserves the 16-residue linker between the two polypeptide segments that is lost in the wild-type protein after posttranslational processing by endopeptidases and carboxypeptidases. The biological importance of RicC3 and the availability of homogeneous $^{15}$N labeled recombinant protein prompted us to determine its high-resolution three-dimensional structure, to provide a representative structure for the broad family of the 2S albumin proteins. Moreover, natural RicC3 constitutes the peptidic component of the immunomodulator Inmunoferon, a widely used drug (*10*).

[†] This work has been supported by projects PB98-0667 of the DGESIC of the Spanish Ministry of Education and BIO2002-00720 of the Spanish Ministry of Science and Technology to M.R. and J.S., respectively, and partially by project FIS C03/14 and an I+D Industrial Farmacéutica Cantabria S.A. contract to G.G.-G. D.P.-U. was supported by a predoctoral grant from the Ministry of Science and Technology.

[‡] Coordinates for the minimized 20 structures of RicC3 have been deposited in the RCSB Protein Data Bank (accession code 1PSY).

* To whom correspondence should be addressed. E-mail: jsantoro@iqfr.csic.es.

[§] Instituto de Química Física Rocasolano.

[||] Centro de Investigaciones Biológicas.

[1] Abbreviations: nsLTP, nonspecific lipid transfer protein; rRicC3, recombinant RicC3; RMSD, root-mean-square deviation.

10.1021/bi0352217 CCC: $25.00 
Published on Web 11/08/2003

The determination of its three-dimensional structure represents a fundamental first step toward understanding its biological activity at a molecular level, as well as improving its pharmacological properties by genetic engineering.

The prolamin superfamily of proteins has been the subject of recent interest because of the prevalence of some of its members as allergens (*11*) in a much greater level that can be accounted for on the basis of their abundance in plant proteins. Increasing emphasis has been placed on the role of plant proteins as allergens in relation to novel and transgenic foods. The available three-dimensional structures of prolamin superfamily members show that they have a similar fold, that has been described as a four-helix bundle arranged in a right-handed superhelix (*12*). However, the determination of a representative structure for the 2S albumins family by crystallographic or NMR methods has been very difficult due to the heterogeneity of the proteins isolated from natural sources. In fact, there is only a single structure determination of a 2S albumin, that of napin BnIb (*13*) isolated from rapeseed, in which just the global fold of the protein was determined. Thus, there is a pressing need for a high-resolution structure of at least one member of the 2S albumin family of allergens. Here we report on the three-dimensional structure of recombinant RicC3 as determined by NMR methods. Previous reports have shown that recombinant and native RicC3 are indistinguishable in stability and structure, according to far UV CD spectroscopy (*9*, *10*). The structure reported here should serve as the structural foundation for establishing meaningful relationships between sequence, structure, and allergenicity for the members of the prolamine superfamily.

## MATERIALS AND METHODS

*Protein Sample.* The recombinant RicC3 is a 125-residue protein containing the short (residues 8−39) and large (residues 56−120) chains of the mature protein, attached by a 16-residue linker (40−55), as well as extensions at the N- and C-termini of 7 and 5 residues, respectively. The recombinant RicC3 from *R. communis*, rRicC3, was produced by using a new system for high expression of heterologous natively folded proteins using minimal medium cultures of *E. coli* (*9*). $^{15}NH_4Cl$ was employed as the only nitrogen source to produce protein uniformly labeled with $^{15}N$. rRicC3 was purified, at a semipreparative scale, from the supernatant of the cultures, concentrated ∼10 times by high performance ultrafiltration (Pall Filtron) and purified by the subsequent treatments: (i) heat precipitation of the bulk non-RicC3 protein (90 °C, 30 min); (ii) ethanol (80% v/v) precipitation of the remaining soluble proteins; (iii) SP-Sepharose (Amersham-Pharmacia) chromatography of the precipitate fraction soluble in distilled water; (iv) reversed-phase HPLC using a volatile buffer (trifluoroacetic acid 10 mM) and acetonitrile as elution solvent. The purified protein was lyophilized and stored at −20 °C until use.

*NMR Spectroscopy.* rRicC3 samples were prepared for NMR experiments at ∼2.0 mM concentration in 95% $H_2O$/5% $D_2O$ or in $D_2O$ solution containing 10 mM sodium phosphate and TSP at pH 3.5 in 5 mm NMR tubes. All NMR spectra were carried out at 308 K on a Bruker DMX 600 NMR spectrometer operating at 600.13 MHz for the proton equipped with a triple-resonance probe and an active shielded z-gradient coil. Proton chemical shifts were referenced to internal trimethyl silyl propionate (TSP) and $^{15}N$ chemical shifts were referenced indirectly using the gyromagnetic ratios of $^{15}N$:$^{1}H$ (*14*). 2D $^{1}H$-$^{15}N$ HSQC, and 3D $^{15}N$-NOESY−HSQC (80 ms mixing time), $^{15}N$-TOCSY−HSQC (65 ms mixing time), HNHA and HNHB experiments were performed using the $^{15}N$ labeled protein. On the unlabeled sample, 2D COSY, TOCSY (65 ms mixing time), and NOESY (80 ms mixing time) spectra were acquired. Water suppression was achieved either by selective presaturation or by including a WATERGATE module (*15*) in the original pulse sequences prior to acquisition. Conventional pulse sequences and phase-cycling procedures were used for all experiments except HNHA (*16*) and HNHB (*17*) for which the pulse sequences were modified by inclusion of a gradient-PEP module (*18*, *19*).

The size of the acquisition data matrix for the 2D spectra was 2048 × 512 data points in f2 and f1, respectively. For $^{15}N$-NOESY−HSQC and $^{15}N$-TOCSY−HSQC was 2048 × 64 × 240 data points in f3 ($^{1}H$), f2 ($^{15}N$), f1 (HN), respectively, and for HNHA and HNHB was 2048 × 80 × 128 data points in f3 ($^{1}H$), f2 ($^{15}N$), f1 (HN). The data matrixes were zero filled up to the duplicate number of data points in all dimensions. The resolution enhancement methods used included sine-bell or square-sine-bell windows with optimized shifts for every spectrum.

The processing of the 2D and 3D spectra was carried out by using the programs XWINNMR (Bruker Biospin, Karlsruhe, Germany) and NMRPipe (*20*), respectively. The spectra analysis, assignment, and cross-peak volume calculations were performed with NMRView (*21*).

The assignments of $^{1}H$ and $^{15}N$ resonances (*22*) were obtained from a combination of 2D and 3D experiments following a standard strategy and have been deposited in the BioMagResBank database (*23*) under accession number BMRB-5374.

*Distance and Angle Constraints.* After assignment of most of the main chain and side chain proton resonances of rRicC3, the 2D and 3D NOESY spectra were analyzed to assign as many dipolar cross-correlations as possible. Those unambiguously assigned to a given pair of protons were converted, after internal calibration, into upper distance limits by the CALIBA program (*24*). The limits were restricted within a 2.7−6.0 Å range. The strongest sequential $HN_i$/$HN_{i+1}$ connectivity in helical region was taken as the reference for the 2.8 Å distance. Finally, after DYANA filtering, 1564 upper limit distance constraints, distributed as 435 intraresidual, 469 sequential, 389 short range, and 271 long range, were obtained. Additionally, as RicC3 contains eight cysteine residues, four sets of standard upper and lower limits for each identified disulfide bond, 2.1/2.0 Å for S$\gamma$(i)−S$\gamma$(j) and 3.1/3.0 Å for C$\beta$(i)−S$\gamma$(j) and S$\gamma$(i)−C$\beta$(j), were introduced during the rounds of calculations. The distance constraints were supplemented with 70 $\phi$ torsion angle constraints derived from the $^{3}J_{HNH\alpha}$ coupling constants measured from the HNHA experiment and 30 $\psi$ torsion angle constraints derived from the $^{3}J_{H\alpha Ni+1}$ coupling constants from the HNHB experiment. No stereospecific assignments of prochiral protons were introduced in the initial calculations, and the usual correction for pseudoatoms (*25*) was added in all cases. In the final steps, the stereospecific assignments of H$\beta$ of residues Glu11, Gln16, Gln36, Leu63, Cys79, Glu80 and H$\alpha$ of Gly97 and Gly108 were generated by the

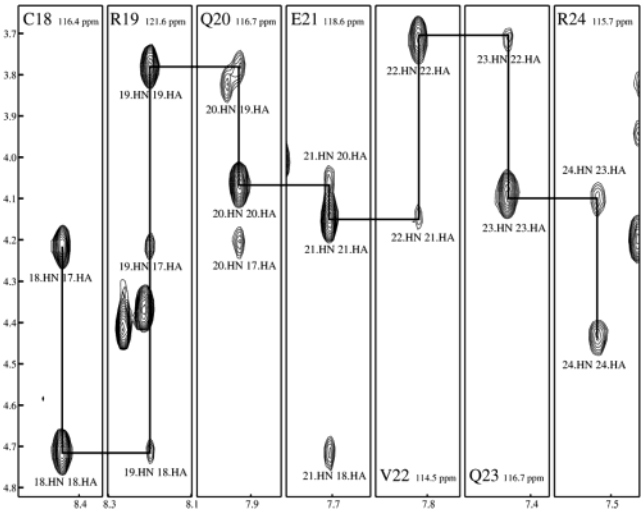


FIGURE 1: Strips from a $^{15}$N NOESY−HSQC spectrum of rRicC3 in 19:1 $H_2O/D_2O$ at 308 K with 80 ms mixing time, showing the assignment pathway for helix Ia.

program GLOMSA (*24*) on the basis of the set of 20 protein structures and interproton distance upper limits.

*Structure Calculation.* The three-dimensional solution structure of rRicC3 was determined by simulated annealing combined with molecular dynamics in torsion angle space followed by a final minimization. On the basis of the observed $^1$H and $^{15}$N chemical shifts, a small number of isolated and unambiguous NOE peaks was selected and used to calculate a set of preliminary rRicC3 solution structures with the program DYANA 1.5 (*26*). On the basis of their consistency with these initial structures, an additional set of NOEs could then be assigned and subsequently used to calculate a new set of structures. Several such cycles of structure calculation and NOE assignment, using the automatic peak assignment programs NOAH (*27*) and CANDID (*28*), were applied to assign the maximum number of NOE peaks possible. With these distance constraints and angular constraints, a set of 50 structures was calculated by using the program DYANA 1.5. All His, Lys, and Arg side chains were treated as positively charged and all Glu and Asp side chains were considered to be negatively charged. The 20 structures with the lowest DYANA target function were selected for further analysis. The 20 structures were refined by energy minimization in vacuo (200 steps) in the presence of NMR constraints using the program AMBER7 (*29*). All DYANA and AMBER calculations were performed on a cluster of 8 Intel computers Pentium II using the RedHat7.2 operating system. PROCHECK-NMR (*30*) was used to analyze the quality of the refined structures. MOLMOL (*31*) was used to prepare the figures of the molecules.

## RESULTS

*Assignment.* The NMR spectra of rRicC3 were assigned based on the following spectra: NOESY, TOCSY, COSY, $^1$H-$^{15}$N HSQC, $^{15}$N-NOESY−HSQC, and $^{15}$N-TOCSY−HSQC. The backbone assignments were further confirmed by the HNHA and HNHB experiments. The $^1$H and $^{15}$N assignments of the backbone atoms are complete, except for Ala1 and Pro51. Some representative strips of the sequential assignment are shown in Figure 1. Further analysis of the TOCSY and NOE data permitted a nearly complete $^1$H side chain assignment, with the only exceptions being some distal side chain resonances in arginines and lysines and side chain $NH_2$ resonances of asparagines and glutamines (*22*), leading to the assignment of 90.5% of all protons. Additional small cross-peaks including Val47, Leu48, Gly52, and Asn55 were observed in the TOCSY spectra. Since all these residues belong to the highly disordered linker (see below), we ascribe these signals to the cis isomer of Pro51, although other explanations are possible. The preferred X−Pro peptide bond isomer is definitely the trans one, as deduced from the intense NOESY Met50(Hα)−Pro51(Hδ*) sequential cross-peaks.

*Disulfide Bonds.* The three-dimensional structure of rRicC3 is stabilized by four disulfide bonds. The individual covalent links between the cysteine residues could be determined by a careful analysis of the NOE spectra since the distance between the H$\beta$ of one Cys and the Hα of the second Cys involved in the same disulfide bond is within the range of detectable NOEs (*32*). Several inter-cysteine NOE connectivities allowed us to identify clearly the four disulfide bonds: SS1, Cys18(H$\beta$2, H$\beta$3)−Cys77(Hα) and Cys18(H$\beta$3)−Cys77(H$\beta$3); SS2, Cys30(HN, Hα)−Cys66(H$\beta$2); SS3, Cys67(Hα)−Cys114(H$\beta$3, H$\beta$2) and Cys67(H$\beta$2, H$\beta$3)−Cys114(H$\beta$3, H$\beta$2); SS4, Cys79(Hα)−Cys118(Hα, HN) and Cys79(H$\beta$2, H$\beta$3)−Cys118(HN). Long-range NOEs between residues in the vicinity of the cysteines confirm the accepted disulfide pattern (see Figure 3). The proposed cysteine pairing is fully compatible with the distance restraints and is in agreement with the disulfide map determined for other members of the 2S albumin family (*10*, *33*, *34*).

*Secondary Structure.* Figure 2 summarizes the Hα conformational shifts, $^3J_{HNH\alpha}$ coupling constants, $^3J_{NH\alpha}$ coupling constants, sequential and short-range NOEs, and NH exchange data. The conformational shifts are calculated as $\Delta\delta_{H\alpha} = \delta_{H\alpha}^{OBS} - \delta_{H\alpha}^{RC}$, where $\delta_{H\alpha}^{RC}$ are the shifts corresponding to the random coil (*35*). The presence of five helical secondary structure elements was first derived from a dense array of characteristic NOEs and from stretches of slowly exchanging backbone amide protons and low values of $^3J_{HNH\alpha}$ couplings. All these data support the presence of five α-helices spanning approximately residues 17−23, 29−36, 61−73, 78−93, and 99−115. Rounds of three-dimensional structure calculations led more precise limits for the helices.

*Tertiary Structure: Backbone.* A total of 1564 interproton upper distance constraints, corresponding to nonoverlapping and unambiguously assigned cross-correlations, were used in the structure calculation. Figure 3a shows a diagonal plot of NOEs identified in the rRicC3 NOESY spectra. The global fold of the polypeptide chain is uniquely defined due to the large number of NOEs in the protein core, Figure 3b.

A set of 50 structures was calculated using the program DYANA 1.5 (*26*) with the distance constraints derived from unambiguous NOE constraints and the dihedral angle constraints derived from the HNHA and HNHB experiments. The 20 structures with the lowest DYANA target function were selected for analysis. This set of structures was subjected to a restrained energy minimization in vacuo with the program AMBER7 (*29*). The resulting structures satisfy the experimental constraints with small deviations from idealized covalent geometry and most of the backbone torsion angles for nonglycine residues lie within the allowed regions in the Ramachandran plot. The statistics and details of quality and precision for the accepted structures are summarized in

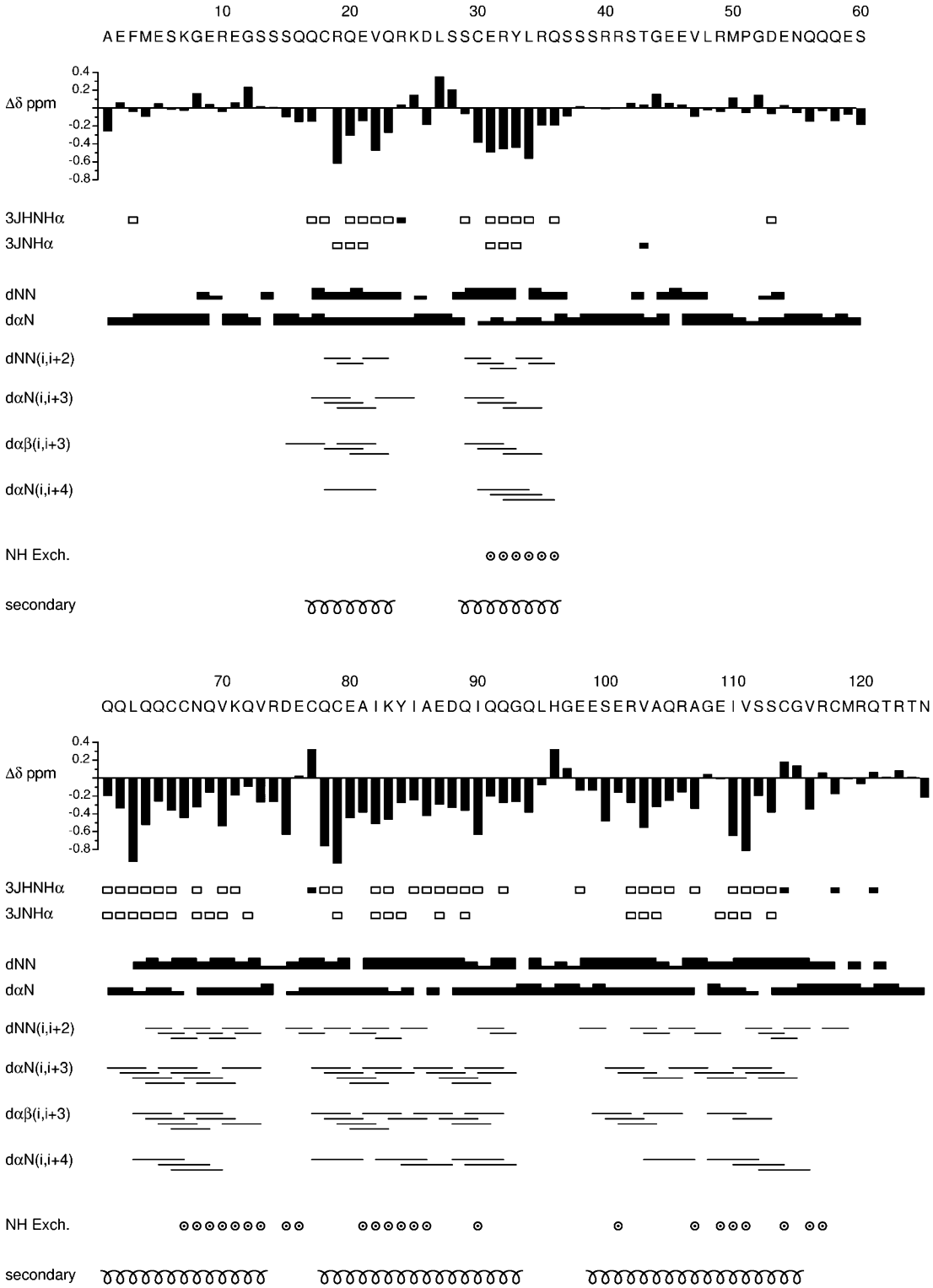


FIGURE 2: Summary of the data on $C^{\alpha}H$ conformational chemical shifts, $^3J_{HNH\alpha}$ (black for $^3J > 8$ Hz, gray for 8 Hz $> {}^3J >$ 6 Hz, and white for 6 Hz $> {}^3J$), $^3J_{NH\alpha}$ (black for $^3J > -1$ Hz and white for $^3J < -1$ Hz), the short-range NOEs, and slow exchanging amide protons.

Table 1 and a superposition of the selected structures is shown in Figure 4. The RMSD radius for backbone heavy atoms of the most structured regions is 0.64 Å, excluding unstructured regions at the N- and C-termini (1−17 and 120−125, respectively) and the linker between residues 36 and 59. The computed structures are fully compatible with the resonance and NOE cross-correlation assignments. The maximum violation was 0.44 Å, the average sum of violations for the whole set of distance constraints 6.2 Å, and the number of violations larger than 0.4 Å was two. The constraints are not evenly distributed throughout the molecule and regions of the protein with higher densities of constraints are better defined (Figures 3 and 4).

The global fold consists of five amphipatic helices spanning residues 18−23, 29−36, 62−71, 77−93, and 100−112. The five helices are arranged in a right-handed superhelix, a folding motif first observed in nonspecific lipid transfer proteins (nsLTPs) and detected afterward in 2S albumins. Between the N-terminus and the linker loop there are two helices, Ia and Ib (in blue), running approximately

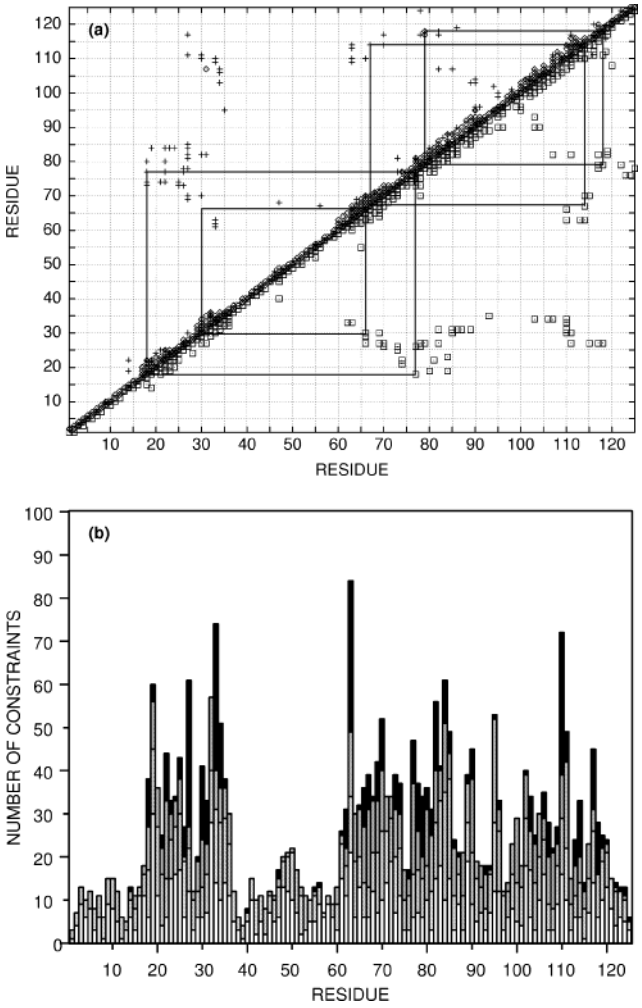


FIGURE 3: (a) Diagonal plot of the interproton NOEs: backbone-backbone (diamond) and side chain−side chain (cross) interactions are plotted above the diagonal; backbone−side chain (square) interactions are shown below the diagonal. Disulfide bridges are shown by lines. (b) Distribution of intraresidual (white), sequential (light gray), short range (dark gray), and long range (black) NOEs by residue.

coaxial and joined by a nonhelical turn (residues 24−26) and an extended strand. This whole segment bends so as to fit its component helices Ia and Ib into the bundle formed by helices III and IV (in green), which runs almost orthogonal to them. Between helix Ib and helix II, there is a long unstructured loop encompassing the residues corresponding to the linker. This unstructured loop does not interact with any other part of the protein, which is evidence that the 3D core structure of the wild-type protein will not change much with respect to rRicC3. Helix II (in green) runs approximately antiparallel to helices Ia and Ib. An analysis of the backbone dihedral angles shows that the helix ends up at residue 71. The next two residues, Gln72 ($\Phi = -95.8$ and $\Psi = 6.2$ averaged over the 20 structures) and Val73 ($\Phi = -68.0$ and $\Psi = 165.0$) are not located in the helical region of the Ramachandran plot, but are involved in regular hydrogen bonds with their respective (i-3) residues (72−69 and 73−70). Helix II is followed by a short loop and helix III. In helix III, the hydrogen bonding network is somewhat irregular, showing bifurcated hydrogen bonds (79 to 82 and 83 as well as 81 to 84 and 85) and several hydrogen bonds are missing. Between helix III and helix IV there is a relatively short segment (in yellow), named

Table 1: Structural Statistics of the 20 Best Structures of rRicC3

| Constraints | | | |
|---|---|---|---|
| distances | | angles | |
| intraresidual | 435 | $\phi$ | 70 |
| sequential | 469 | $\psi$ | 55 |
| medium range: $(i-j) < 5$ | 389 | | |
| long range: $(i-j) \geq 5$ | 271 | | |

| DYANA | | | |
|---|---|---|---|
| | average | minimal | maximal |
| target function | 1.47 | 1.07 | 1.76 |
| maximal violation | 0.27 | 0.18 | 0.44 |
| sum of violations | 6.20 | 4.80 | 7.00 |

| RMSD Radius | | |
|---|---|---|
| | backbone | heavy atoms |
| global (18−37, 60−119) | 0.64 | 1.12 |
| helix Ia | 0.16 | 0.97 |
| helix Ib | 0.13 | 0.83 |
| helix II | 0.15 | 0.82 |
| helix III | 0.48 | 1.00 |
| helix IV | 0.31 | 0.81 |

| AMBER | |
|---|---|
| mean energy (kcal/mol) | −3581 |
| RMS deviations from ideal geometry | |
| bond lengths (Å) | 0.0107 |
| bond angles (deg) | 1.877 |
| Ramachandran | |
| favorable regions | 90.9% |
| additional allowed | 9.1% |
| disallowed | 0% |

the "hypervariable region" of the 2S albumins (*36*−*38*), because of the high variability in length and composition of this segment among the different members of this family. Two $\gamma$-turns stabilized by hydrogen bonds, Leu95(HN) to Gly93(O) and Glu98(HN) to His96(O), are found in this region. Helix IV starts with irregular hydrogen bonds (100 to 103 and 104 and 101 to either 104 or 105) and then extends regularly until residue 112. Residues 114 and 115, immediately after helix IV, although not having helical backbone angles ($\Phi = -152.8$, $\Psi = 39.8$ and $\Phi = -135.0$, $\Psi = 26.8$, respectively), have their NH hydrogen bonded with the (i-4) preceding helical residues (114 to 110 and 115 to 111) in all 20 structures. Finally, after helix IV, there is a long loop devoid of any regular secondary structure, which packs against helix III.

Four disulfide bonds stabilize the tertiary structure. Helix II, which contains the Cys−Cys motif, is linked to helix Ib and to the C-terminal end of helix IV. On the other side of the protein, the Cys−X−Cys motif, at the N-terminus of helix III, connects to helix Ia and to the C-terminus of the protein. This arrangement leads to a very compact structure. The volume of the protein as calculated with the program VADAR (*39*) is 11956.4 Å$^3$, a value 30% lower than expected from its molecular weight (16997.6 Å$^3$).

*Tertiary Structure: Side Chains.* Many side chain conformations of residues in the two chains of the mature protein are well defined, with 36 residues having $\chi_1$ circular variances lower than 0.05. In contrast, most side chains of residues in the linker are disordered. Most hydrophobic aliphatic residues of the four amphipatic helices (Val22, Leu27, and Leu34 in helix I, Val70 in helix II, Ile82 and Ile90 in helix III, and Val103, Ile110, and Val111 in helix IV) have well defined

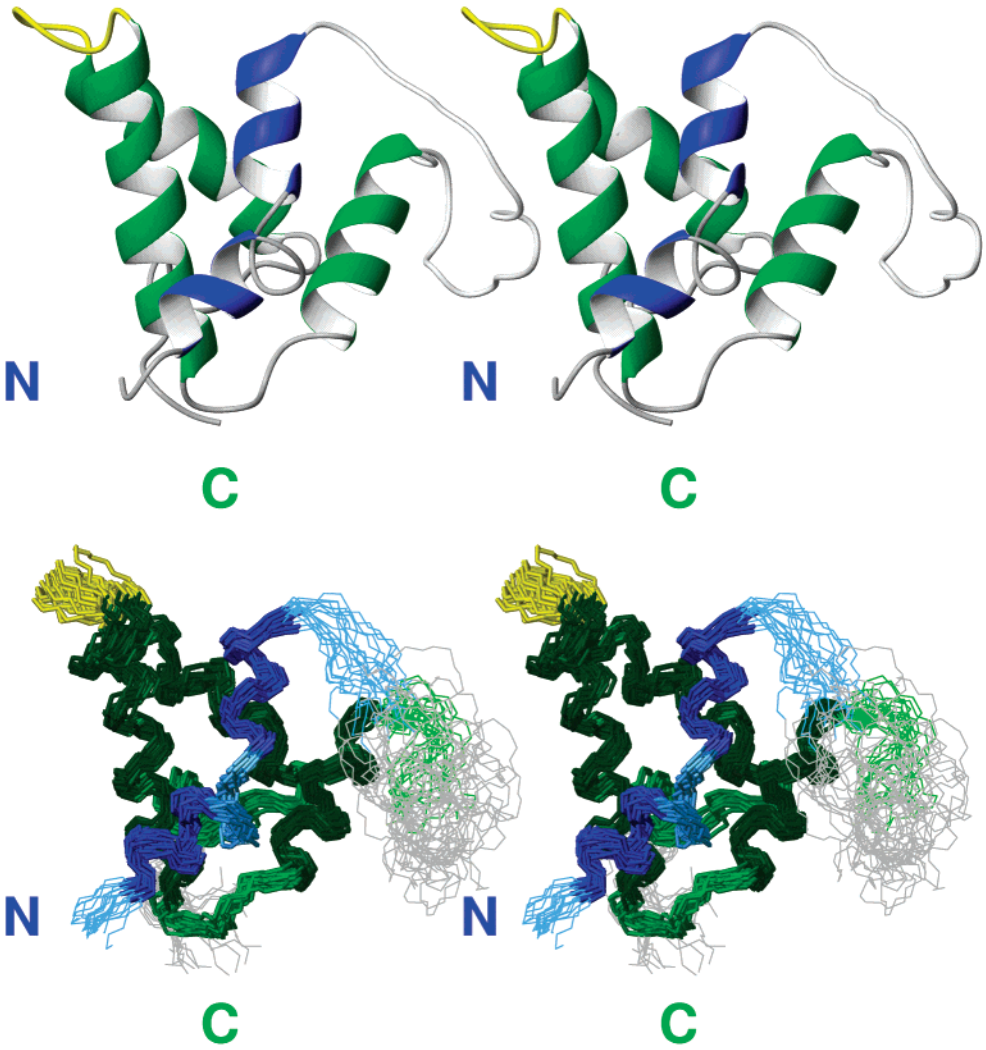


FIGURE 4: Ribbon representation of the solution structure of rRicC3, showing helices (blue and green) and loops (gray, but the hypervariable loop in yellow) (top) and stereoview of the 20 best DYANA structures of rRicC3 after restrained energy minimization (bottom). The first 14 residues, highly disordered, are not represented.

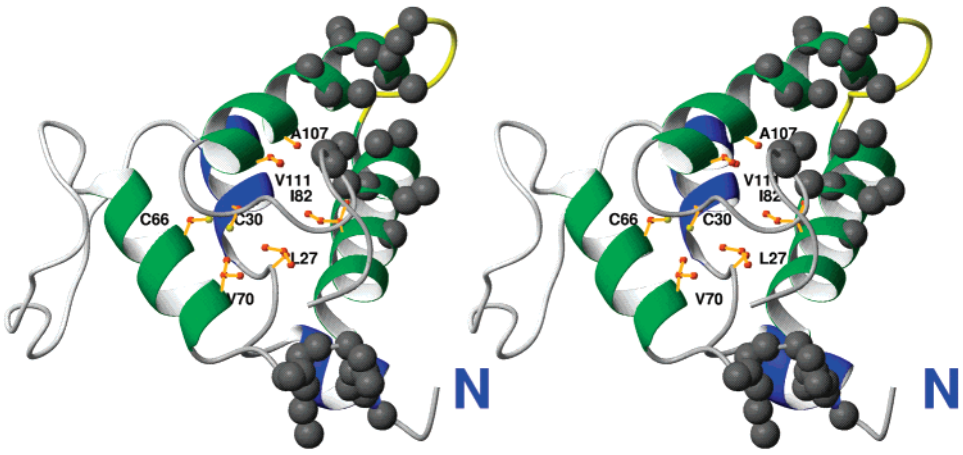


FIGURE 5: Stereoview of rRicC3 showing the side chains of residues belonging to the hydrophobic core and the atoms forming the two hydrophobic patches (gray spheres).

side chains, the only exceptions being Leu63 in helix II and Ile85 in helix III. Tyrosine residues, Tyr33 in helix I and Tyr84 in helix III, also show well-defined side chains despite being partially exposed as deduced from their fractional accessibilities of 0.37 and 0.22.

The hydrophobic side chains are buried and pack together to form several clusters. The main exceptions are Ile90, only partially buried, and Ala104 whose side chain is exposed. The main hydrophobic core of the protein (Figure 5) is constituted by Leu27, Val70, Ile82, Ala107, Val111, and the Cys30−Cys66 disulfide bond, and is surrounded by helix Ib, helix II, the N-terminal part of helix III and the C-terminal part of helix IV. The Cys67−Cys114 and Cys79−Cys118 disulfide bonds are also located near the hydrophobic core,

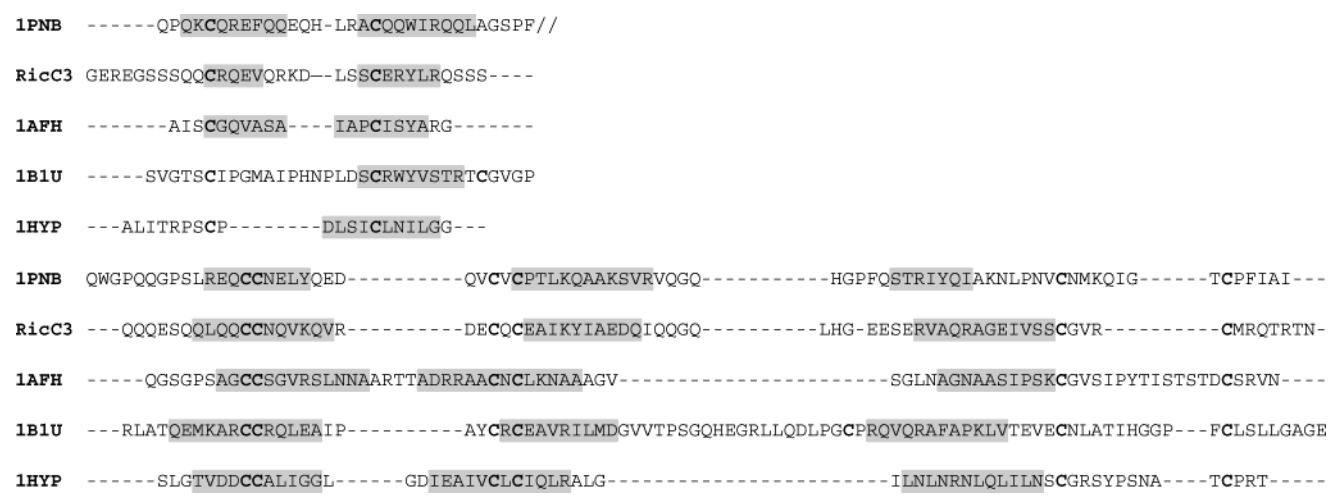


FIGURE 6: Sequence alignment of the prolamin superfamily.

but do not form part of it. On the other side of the Cys30−Cys66 disulfide Tyr33, Leu63, and Ile110 pack together in a partially exposed region. The protein also displays two large hydrophobic patches with solvent-accessible surface areas of 250−300 Å$^2$. One patch extends from the hypervariable loop to C-terminus of the protein through the helices III and IV, while the second one is located around the Cys18−Cys77 disulfide bond and the loop connecting helices II and III.

rRicC3 has 21 negative and 19 positive charged residues. Most of these charged residues are located on the surface of the protein and are highly accessible to solvent. Main exceptions are Asp26 and Glu31 in helix I and Arg117 in the C-terminal loop. The distribution of charged residues on the protein surface is not uniform. A cluster of positive charges is located in the confluence of the C-terminus of helix I (Arg32 and Arg35) and the N-terminus of the helix IV (Arg106), whereas negative charged residues are concentrated in the loop between helices II and III (Asp75, Glu76, and Glu80) and near the hypervariable loop (Glu87, Ap88, Glu98, and Glu99).

## DISCUSSION

*Comparison with 2S Albumins.* The main goal of this research has been to determine the high resolution 3D structure of a 2S albumin family member. Until now, only a single structure of this family, that of napin BnIb purified from rapeseed (*13*), has been described. In that case, the residue heterogeneity precluded the determination of a high-resolution 3D structure, and only the global fold could be advanced. The recently described high-yield synthesis of rRicC3 by recombinant methods yielding a protein with equivalent physicochemical properties as the native protein from castor bean (*1*, *9*, *10*) did make possible its high-resolution 3D structure determination, which is presented here. The high-resolution 3D structure of rRicC3 will serve as a suitable reference structure for the broad family of seed 2S albumin proteins. In particular, this structure may provide insights into similarities and differences between the function of albumins from different species. Moreover, a joint study of the three-dimensional structure of several allergenic 2S albumins, together with other allergenic proteins with similar structures, may serve to establish meaningful relationships between structure and allergenicity. Finally, these results may be of some pharmacological interest since RicC3 is the peptidic component of the drug Inmunoferon. In this regard, it should be pointed out that the pharmacology of the drug is quite complex and not completely known, although it has been in wide use for many years. A comparison of the pharmacological properties of the native and recombinant RicC3 is currently under evaluation. This is not an obvious task, despite the equivalent physicochemical properties of both proteins (*9*, *10*).

The main difference between the structure of napin BnIb and the recombinant RicC3 is that napin BnIb consist of two polypeptide chains, whereas the recombinant RicC3 is a single polypeptide chain. According to the sequence alignment shown in Figure 6 the location of the helices approximately coincides in both proteins and only small differences in the length of the first and the last helix are observed. The global folds are highly similar. A comparison of the two proteins made by superimposing their 3D structures according to the sequence alignment provides a RMSD for the backbone heavy atoms of the helical residues of 3.53 Å. The equivalent residues used in the superposition were 3A−10A, 18A−24A, 11B−19B, 26B−38B, and 48B−61B from napin BnIb (1PNB), and 16−24, 30−36, 63−71, 79−90, and 100−113 from rRicC3. The position of helices Ia, Ib, II, and III is very similar in both proteins. Helix IV, being more orthogonal to helices II and III in rRicC3 than in napin is, however, markedly different. Helix IV was ill defined in the structural study of napin BnIb (*13*). On the contrary, helix IV in rRicC3 is well defined and its orientation is established by a much larger set of long range NOEs. Therefore, the orientation of helix IV in rRicC3 can be definitely considered as more reliable and shall be taken as the standard orientation in 2S albumins.

*Comparison with Other Proteins with a Similar Fold.* RicC3 belongs to a single fold family described in SCOP (*40*) as "4 helices; folded leaf; right-handed superhelix; disulfide-rich". Besides nonspecific lipid transfer proteins, cereal seed inhibitors of α-amylase and/or trypsin, and the hydrophobic protein from soybean, the 2S albumins constitute the prolamin superfamily. The sequence alignment of RicC3 with some representative members of known 3D structures from each family of the prolamin superfamily (albumin 2S from *Brassica napus*, 1PNB (*13*), nsLTP from maize, 1AFH (*41*), α-amylase/trypsin inhibitor from ragi, 1B1U (*42*), and HPS, 1HYP (*43*)) is shown in Figure 6. The

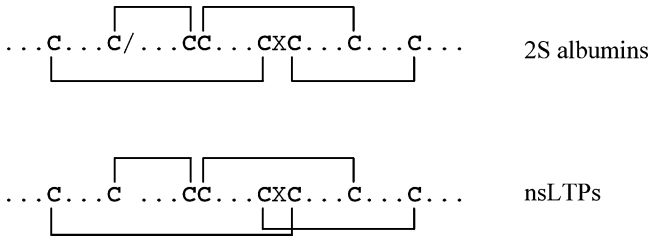


FIGURE 7: Cysteine pairing in 2S albumins and nsLTPs.

routine application of alignment programs such as BLAST (*44*) or CLUSTALW (*45*) do not properly align all cysteine residues. Therefore, a manual reelaboration forcing the alignment of all eight cysteines was performed, which is shown in Figure 6. The shaded segments in each family correspond to a consensus of helix limits for the structures deposited in the PDB for each family. As seen in Figure 6, whereas the location and extension of the helices approximately coincide in the four families, it is however possible to observe some key differences. Only one helix corresponding to helices Ia and Ib in 2S albumins is found in nsLTPs, α-amylase inhibitors and HPS. In α-amylase inhibitors and HPS, this helix is much shorter and corresponds approximately to helix Ib. In nsLTPs, however, the helix is longer and spans to helices Ia and Ib. In fact, in the nsLTP structures (*46*) there is a Pro residue in the middle of helix I, which breaks the hydrogen bond network and induces a bend of about 20°, so that this helix is structurally very similar to the helix pair Ia and Ib found in 2S albumins. The positions of the CC and CXC motifs relative to helix II and helix III, respectively, show striking differences. In nsLTPs, the CC motif is near the N-terminus of helix II, whereas in the other proteins this motif is located more in the middle of the helix. In contrast, the CXC motif is located in the center of the third helix in nsLTPs and in 1HYP (the hydrophobic soybean protein) but is at the N-terminus of helix III in the 2S albumin proteins and in the α-amylase inhibitors.

The most important difference is found in the arrangement of the disulfide bonds formed by the Cys−X−Cys motif in the different families. Whereas in the α-amylase inhibitors, the hydrophobic protein from soybean and the 2S albumins, the first of these cysteines is linked to the N-terminal part of the protein and the second one to the C-terminal region, the reverse arrangement is found in nsLTPs (see Figure 7). The Cys−X−Cys motif in nsLTPs is found in a more shifted position toward the center of helix III than in 2S albumins. These two features seem to be intimately related, and give rise to specific differences in their corresponding three-dimensional structures.

In nsLTPs, a large internal hydrophobic tunnel able to bind one molecule of a lipid is present. This tunnel is parallel to the long axis of the protein and is defined by the almost coplanar helices II and III on one side and helix IV on the other. In rRicC3, however, helix III is rotated and occupies most of the cavity, what can be considered a consequence of the different disulfide pairing. In fact, no internal cavity appears to be present in the rRicC3, according to the program VOIDOO (*47*). A similar displacement of helix III with respect to its position in nsLTPs has been reported for α-amylase inhibitors (*42*, *48*), which also lack the hydrophobic tunnel. These observations strongly suggest that the pattern of cysteine pairing determines whether a hydrophobic cavity is formed in this structural family of proteins.

*RicC3 Is an Allergen.* Previous studies on the aqueous extract of *R. communis* revealed a plethora of antigenic and allergenic components with a broad range of molecular weights (*49*). Among the water-soluble proteins of the castor bean (*50*), the 2S albumin storage protein was demonstrated to be the major allergen. There are two heterodimeric forms of the 2S albumin storage protein of *R. communis*, RicC1 and RicC3, each of which is composed of one short subunit and a long one linked together by disulfide bridges. According to three criteria for isoallergens: similar biological function, molecular size, and sequence similarity in polar residues, RicC1 and RicC3 are isoallergens (*50*). Moreover, a comparison of their sequence with those of 2S albumins from other plants reveals high degrees of similarity and supports the idea that 2S albumins share a common genetic origin (*51*, *52*).

Food allergens may share physicochemical properties that distinguish them from nonallergens. These properties may be used to predict the inherent allergenicity of proteins recently introduced into the food supply by genetic engineering. One candidate property is stability against digestion: important food allergens were found to be stable to digestion in simulated gastric fluid (*53*). The studies of RicC3 upon exposure to acidic pH, protelolytic treatment, and chemical denaturation (*10*) demonstrate that it maintains its molecular integrity and properties during gastrointestinal tract passage. Recombinant RicC3 also maintains its native fold in conditions that simulate passage through the stomach (*9*). These findings also explain the excellent properties of Inmunoferon after oral administration.

By using complementation and competition assays an epitope mapping of the 2S albumin Sin a 1, the major allergen of yellow mustard seeds, was carried out, revealing that one antibody binding site is located in the hypervariable region (*54*, *55*). Recently, another epitope mapping study has identified one immunodominant linear epitope in the Jug r 1, a 2S albumin seed storage protein from the English walnut, which was identified as the major allergen of this nut (*56*). This epitope, RGEEXXE, in which the first four amino acids are critical for IgE binding, is also located in the hypervariable region. A primary sequence comparison of this region with RicC3 revealed a related stretch of amino acids, HGEE, at this position. Since His carries a positive charge at acidic pH, the stretch becomes very similar to the walnut epitope, suggesting that it may play an important role in the IgE−RicC3 interaction.

In rRicC3, the hypervariable region corresponds to an exposed loop, which links two of the α-helices forming the structural core of the molecule. The variability in length and amino acid composition in this loop in different members of the 2S albumin family is good evidence that it does not play any role in determining the folded structure. An empirical equation has been proposed to estimate the NMR $S^2$ order parameters of N−H vectors of the protein backbone (*57*), which can be considered as a measure of local protein flexibility. Application of this equation to rRicC3 indicates that the hypervariable region in rRicC3 shows the highest mobility among all structural elements, with the exception of the linker. Moreover, a computational tool that identifies likely protein−protein binding sites, PASS (*58*), found the

hypervariable loop to be the highest scoring region. All these findings point toward the conclusion that the hypervariable region is the most important antigenic region of the 2S albumins. On the other hand, both hypervariability and the loose three-dimensional structure of this loop suggest that this region of the protein may easily accept its substitution and even the insertion of several residues. Thus, RicC3 may constitute a useful scaffold for the production of synthetic ligands and to produce designed vaccines.

## ACKNOWLEDGMENT

We thank Javier Varela and María I. F. López-Lucendo for the synthesis of the labeled protein, J. P. Pivel and J. L. Alonso for helpful discussions, and D. V. Laurents for the critical reading of the manuscript and revising the English text.

## REFERENCES

1. Shewry, P. R., Beaudoin, F., Jenkins, J., Griffits-Jones, S., and Mills, E. N. C. (2002) *Biochem. Soc. Trans. 30*, 906−910.
2. Youle, R. J., and Huang, A. H. (1978) *Plant Physiol. 61*, 13−16.
3. 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., and Broekaert, W. F. (1992) *J. Biol. Chem. 267*, 15301−15309.
4. Genov, N., Goshev, I., Nikolova, D., Giorggieva, D. N., Filippi, B., and Svendsen, I. (1997) *Biochim. Byophys. Acta 1341*, 157−164.
5. Polya, G. M., Chandra, S., and Condron, R. (1993) *Plant Physiol. 101*, 545−551.
6. Vandekerkhove, J., van Damme, J., van Lijsebettens, M., Botterman, J., De Block, M., Vanderwiele, M., De Clercq, A., Leemans, J., van Montagu, M., and Krebbers, E. (1989) *Biotechnology 7*, 929−932.
7. Altenbach, S. B., Kuo, C. C., Staraci, L. C., Pearson, K. W., Wainwright, C., Georgescu, A., and Townsed, J. (1992) *Plant Mol. Biol. 18*, 235−245.
8. Irwin, S. D., Keen, J. N., Findlay, J. B. C., and Lord, J. M. (1990) *Mol. Gen. Genet. 222*, 400−408.
9. Fernández-Tornero, C., Ramón, A., Navarro, M. L., Varela, J., and Giménez-Gallego, G. (2002) *Biotechniques 32*, 1238−1242.
10. Varela, J., Navarro Rico, M. L., Guerrero, A., García, F., Giménez-Gallego, G., and Pivel, J. P. (2002) *Methods Find. Exp. Clin. Pharmacol. 24*, 471−80.
11. Pastorello, E. A., Pompei, C., Pravettoni, V., Brenna, O., Fariolo, L., Trambaioli, C., and Conti, A. (2001) *Allergy 5*, 45−47.
12. Pantoja-Uceda, D., Bruix, M., Santoro, J., Rico, M., Monsalve, R., and Villalba, M. (2002) *Biochem. Soc. Trans. 30*, 919−924.
13. Rico, M., Bruix, M., González, C., Monsalve, R. I., and Rodríguez, R. (1996) *Biochemistry 35*, 15672−15682.
14. Wishart, D. S., Bigam, C. G., Yao, J., Abilgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) *J. Biomol. NMR 6*, 135−140.
15. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR 2*, 6, 661−665.
16. Vuister, G., and Bax, A. (1993) *J. Am. Chem. Soc. 115*, 7772−7777.
17. Düx, P., Whitehead, B., Boelens, R., Kaptein, R., and Vuister, G. W. (1997) *J. Biomol. NMR 10*, 301−306.
18. Palmer, A. G., Cavanagh, J., Wright, P. E., and Rance, M. (1991) *J. Magn. Reson. 93*, 151−170.
19. Kay, L., Keifer, P., and Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10665.
20. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.
21. Johnson, B. A., and Blevins, R. A. (1994) *J. Biomol. NMR 4*, 603−614.
22. Pantoja-Uceda, D., Bruix, M., Varela, J., López-Lucendo, M. I., Giménez-Gallego, G., Rico, M., and Santoro, J. (2002) *J. Biomol. NMR 23*, 331−332.
23. Seavey, B. R., Farr, E. A., Westler, W. M., and Markley, J. L. (1991) *J. Biomol. NMR 1*, 217−236.
24. Güntert, P., Braun, W., and Wüthrich, K. (1991) *J. Mol. Biol. 217*, 517−530.
25. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
26. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) *J. Mol. Biol. 273*, 283−298.
27. Mumenthaler, C., Güntert, P., Braun, W., and Wüthrich, K. (1997) *J. Biomol. NMR 10*, 351−362.
28. Herrmann, T., Güntert, P., and Wüthrich, K. (2002) *J. Mol. Biol. 319*, 209−227.
29. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., III, Wang, J., Ross, W. S., Simmerling, C., Darden, T., Merz, K. M., Stanton, R. V., Cheng, A., Vincent, J. J., Crowley, M., Tsui, V., Gohlke, H., Radmer, R., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P., and Kollman, P. A. (2002) AMBER 7, University of California, San Francisco.
30. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR 8*, 477−486.
31. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graph. 14*, 51−55.
32. Williamson, M. P., Havel, T. F., and Wüthrich, K. (1985) *J. Mol. Biol. 182*, 295−315.
33. Egorov, T. A., Odintsova, T. I., Musolyamov, A. K., Fido, R., Tatham, A. S., and Shewry, P. R. (1996) *FEBS Lett. 396*, 285−288.
34. Alcocer, M. J. C., Murtagh, G. J., Bailey, K., Dumulin, M., Sarabia Meseguer, A., Parker, M. J., and Archer, D. B. (2002) *J. Mol. Biol. 324*, 165−175.
35. Bundi, A., and Wüthrich, K. (1979) *Biopolymers 18*, 285−297.
36. Krebbers, E., Herdies, L., De Clercq, A., Seurinck, J., Leemans, J., Van Damme, J., Segura, M., Gheysen, G., Van Montagu, M., and Vandekerckhove, J. (1988) *Plant Physiol. 87*, 859−866.
37. Monsalve, R. I., Lopez-Otin, C., Villalba, M., and Rodríguez, R. (1991) *FEBS Lett. 295*, 207−210.
38. Raynal, M., Depigny, D., Grellet, F., and Delseny, M. (1991) *Gene 99*, 77−86.
39. Wishart, D., Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R., and Sykes, B. (2003) Program Vadar, Faculty of Pharmacy University of Alberta, Edmonton, Canada.
40. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) *J. Mol. Biol. 247*, 536−540.
41. Gomar, J., Petit, M. C., Sodano, P., Sy, D., Marion, D., Kader, J. C., Vovelle, F., and Ptak, M. (1996) *Protein Sci. 5*, 565−577.
42. Gourinath, S., Alam, N., Srinivasan, A., Betzel, C., and Singh, T. P. (2000) *Acta Crystallogr. D56*, 287−293.
43. Baud, F., Pebay-Peyroula, E., Cohen-Addad, C., Odani, S., and Lehmann, M. S. (1993) *J. Mol. Biol.* 877−887.
44. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res. 25*, 3389−3402.
45. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res. 22*, 4673−4680.
46. Gomar, J., Sodano, P., Sy, D., Shin, D. H., Lee, J. Y., Suh, S. W., Marion, D., Vovelle, F., and Ptak, M. (1998) *Proteins 31*, 160−171.
47. Kleywegt, G. J., and Jones, T. A. (1994) *Acta Crystallogr. D50*, 178−185.
48. Behnke, C. A., Yee, V. C., LeTrong, I., Pedersen, L. C., Stenkamp, R. E., Kim, S., Reek, G. R., and Teller, D. C. (1998) *Biochemistry 37*, 15277−15288.
49. Thorpe, S. C., Kemeny, D. M., Panzani, R., McGurl, B., and Lord, M. (1998) *J. Allergy Clin. Immunol. 82*, 67−72.
50. Bashir, M. E., Hubatsch, I., Leinenbach, H. P., Zeppezauer, M., Panzani, R., and Hussein, I. H. (1998) *Int. Arch. Allergy Immunol. 115*, 73−82.
51. Odani, S., Koide, T., Ono, T., and Ohnishi, K. (1983) *Biochem. J. 213*, 543−545.
52. Ampe, C., Van Damme, J., de Castro, L., Sampaio, M., Van Montagu, M., and Vandekerckhove, J. (1986) *Eur. J. Biochem. 159*, 597−604.
53. Astwood, J. D., Leach, J. N., and Fuchs, R. L. (1996) *Nat. Biotechnol. 14*, 1269−1273.
54. Menéndez-Arias, L., Domínguez, J., Moneo, I., and Rodríguez, R. (1990) *Mol. Immunol. 27*, 143−150.
55. Monsalve, R. I., González de la Peña, M., Lopez-Otin, C., Villalba, M., and Rodríguez, R. (1993) *Biochem. J. 293*, 625−632.
56. Robotham, J. M., Suzanne, S., Teuber, S., Sathe, S. K., and Roux, K. H. (2002) *J. Allergy Clin. Immunol. 109*, 143−149.
57. Zhang, F., and Brüschweiler, R. (2002) *J. Am. Chem. Soc. 124*, 12654−12655.
58. Brady, G. P., Jr, and Stouten, P. F. (2000) *J. Comput.-Aided Mol. Des. 14*, 383−401.

BI0352217