FEBS Letters 412 (1997) 190–196 FEBS 18825

Amino acid sequence and three-dimensional structure of the Tn-specific isolectin B4 from *Vicia villosa*

Eduardo Osinaga[a], Diana Tello[b], Carlos Batthyany[a], Mario Bianchet[c], Gisele Tavares[1,b], Rosario Durán[d], Carlos Cerveñansky[d], Luc Camoin[e], Alberto Roseto[f], Pedro M. Alzari[b,*]

[a]*Dept de Bioquímica, Facultad de Medicina, Av. Gral. Flores 2125, Montevideo, Uruguay*
[b]*Unité d'Immunologie Structurale, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris, France*
[c]*Dept. of Biophysics, Johns Hopkins Medical School, 725 N Wolfe St., Baltimore, MD 21205, USA*
[d]*Instituto de Investigaciones Biológicas Clemente Estable, Av. Italia, Montevideo, Uruguay*
[e]*Institut Cochin de Génétique Moléculaire, 22 rue Mechain, 74014 Paris, France*
[f]*DICA, Université de Technologie de Compiegne, 60206 Compiegne, France*

Received 10 March 1997; revised version received 26 May 1997

Abstract The partial amino acid sequence of the tetrameric isolectin B4 from *Vicia villosa* seeds has been determined by peptide analysis, and its three-dimensional structure solved by molecular replacement techniques and refined at 2.9 Å resolution to a crystallographic *R*-factor of 21%. Each subunit displays the thirteen-stranded β-barrel topology characteristic of legume lectins. The amino acid residues involved in metal- and sugar-binding are similar to those of other GalNAc-specific lectins, indicating that residues outside the carbohydrate-binding pocket modulate the affinity for the Tn glycopeptide. Isolectin B4 displays an unusual quaternary structure, probably due to protein glycosylation.

Key words: Plant lectin; Tn antigen; *Vicia villosa*; X-ray crystallography

1. Introduction

Lectins are carbohydrate-binding proteins of non-immune origin which display a wide diversity of sugar binding specificity in different biological systems [1]. Lectins have served as structural models for the analysis of protein-carbohydrate interactions [2] and have received considerable attention as biochemical tools to detect subtle variations in carbohydrate structures [3]. Crystallographic studies have been carried out on lectins of viral [4] and animal [5–7] origin, but plant lectins have been by far the most extensively characterized. Known three-dimensional structures of plant lectins include Concanavalin A [8], pea lectin [9], favin [10], *Lathyrus ochrus* isolectin I [11], *Griffonia simplicifolia* lectin 4 [12], *Erythrina corallodendron* lectin [13], lentil lectin [14], soybean agglutinin [15], and peanut lectin [16]. The lectin subunit is a 25–30-kDa protein containing a single carbohydrate binding site, and normally occurs in the form of dimers or tetramers, an essential property for agglutinating cells or precipitating complex carbohydrates. Although lectins differ in glycosylation patterns and in the number of monomers per molecule, they share a common tertiary structure, as expected from a significant sequence homology [17,18].

The seeds of *Vicia villosa*, a legume of the pea family (*Papilionaceae*) commonly known as hairy winter vetch, contain a family of tetrameric lectins composed of two different subunits, designated A and B, with different carbohydrate-binding specificities. The homotetrameric isolectin B4, which is the predominant lectin in *V. villosa* seeds, is a glycosylated protein that specifically binds the tumor-associated Tn glycopeptide. The Tn determinant (GalNAcα-*O*-Ser/Thr), normally a cryptic structure in the peptide core of *O*-glycoproteins, was first discovered at the red blood cell surface of a patient with hemolytic anemia and found to be responsible for the polyagglutination of erythrocytes caused by anti-Tn antibodies which are universally present in human sera (Tn syndrome) [19]. It is now known that this antigen, expressed in an unmasked form in about 90% of human carcinomas [20], is one of the most specific human carcinoma-associated structures. The isolectin B4 from *Vicia villosa* seeds, which specifically agglutinate Tn-exposed red blood cells [21], has recently been used to develop an immuno-lectin enzymatic assay for the detection of Tn soluble antigen in sera of cancer patients [22]. As part of our structural studies of Tn-binding proteins, we report here the partial amino acid sequence and the crystal structure of VVLB4.

2. Materials and methods

2.1. Protein digestion, peptide purification and sequencing

Purified VVLB4 was obtained as a lyophilized powder from Sigma (product No. L7513). The preparation was resuspended in water and its purity checked by SDS-PAGE analysis. In order to obtain peptides for sequence analysis, the protein was digested with TPCK-treated trypsin, endoproteinase Lys-C and endoproteinase Glu-C using standard procedures, and treated with 0.01 M HCl for 1 h at 100°C to achieve limited hydrolysis. In addition, Asp-Pro bonds were specifically cleaved by mild acid hydrolysis [23], Trp-X bonds by oxidation of the sample with DMSO/HCl prior to exposure to cyanogen bromide [24], and Met-X bonds by treatment with cyanogen bromide. Peptides from the reaction mixtures were purified by RP-HPLC on a C18 Vydac column (150 mm×2.1 mm) eluted with a linear gradient of acetonitrile in water containing 0.1% trifluoroacetic acid at a flow rate of 0.25 ml/min. Detection was performed at 220 nm. Large peptide fragments obtained by chemical cleavage at Asp-Pro, Trp-X and Met-X bonds were recovered after SDS-PAGE followed by electroblotting onto a polyvinylidene difluoride membrane. For amino acid composition studies, hydrolysis was carried out for 24 h at 110°C in

*Corresponding author. Fax: (33) (1) 45 68 86 39.
E-mail: alzari@pasteur.fr

[1]*Permanent address:* Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, SP, Brazil.

Abbreviations: EcorL, *Erythrina corallodendron* lectin; Gal, Galactose; GS4, *Griffonia simplicifolia* lectin 4; VVLB4, isolectin B4 from *Vicia villosa*; GalNAc, *N*-acetylgalactosamine; GlcNAc, *N*-acetylglucosamine

PII S0014-5793(97)00677-7

6 M HCl containing 10 mg/ml phenol. Microsequence analyses were performed using an Applied Biosystems 475 A sequenator. Sequence comparisons were carried out with the GCG computer software package.

2.2. Crystallization and diffraction data collection

VVLB4 was crystallized as reported [25]. The VVLB4 crystals grow to a large size (~1 mm), but are very unstable under the X-ray beam and display different (albeit related) crystal forms. The crystals studied here belong to the orthorhombic space group $P2_12_12_1$, with $a=83.4$ Å, $b=91.5$ Å, $c=147.8$ Å, and contain four lectin subunits in the asymmetric unit. Diffraction data at 2.9 Å resolution were collected at cryogenic temperature using synchrotron radiation from the DW32 beamline at LURE (Paris, France) and a MARresearch Image Plate scanner. Data reduction (Table 1) was carried out with the programs DENZO and SCALEPACK [26].

2.3. Structure determination and refinement

The structure was solved by molecular replacement using the program AMoRe [27]. Atomic coordinates of the pea lectin dimer and the EcorL monomer obtained from the Brookhaven Protein Data Bank [28] (PDB entry codes 1lte and 2ltn, respectively) were used as search probes. Rotation and translation function maps using the pea lectin dimer as a search probe model showed clear peaks for the position of two independent dimers in the asymmetric unit. The structure could be solved independently using the EcorL monomer as the search molecule, thus confirming the overall correctness of the crystal structure determination. With the latter, however, a significantly larger number of peaks of the rotation function map had to be explored to find the four correct orientations.

Model building was performed with the program O [29], and crystallographic refinement was carried out with the program X-PLOR [30] using the potential parameters proposed by Engh and Huber [31]. The EcorL monomer (with most amino acid residues replaced by alanine) was used as the initial model for refinement. A few initial cycles of simulated annealing refinement with XPLOR, in which the non-crystallographic 222 symmetry was imposed, and inspection of electron density maps reduced the crystallographic *R*-factor to 28% at 3 Å resolution. At this stage, newly available VVLB4 sequence data were incorporated into the model, and residues at positions of unknown sequence were changed to either the consensus between several plant lectins (when consistent with the electron density maps), or to alanine. Further cycles of model building and refinement were then carried out as described above. During the last cycles, the non-crystallographic constraint was released and the whole tetramer in the asymmetric unit was allowed to refine independently. The final refinement parameters are presented in Table 1.

Table 1
Data collection and refinement statistics

Data collection	
Resolution range (Å)	10–2.9
Observed reflections	162 042
Unique reflections	22 571
R-merge (%)[a]	10.4
Completeness (%)	89.4
Refinement	
Resolution range (Å)	10–2.9
Observed reflections	22 571
Sigma cutoff	none
R-factor (%)[b]	21
Free *R*-factor(%)[c]	31
Protein atoms	8462
Water molecules	none
Mean B protein atoms (Å)[b]	20.7
rms bond lengths (Å)	0.01
rms bond angles (degrees)	1.7
rms B bonded atoms (Å)[b]	2.0

Notes:
[a] $R\text{-merge}=\sum_{hkl}|I-\langle I\rangle|/\sum_{hkl}I$
[b] $R\text{-factor}=\sum|F_{calc}-F_{obs}|/\sum|F_{obs}|$
[c] Free *R*-factor defined in Brünger [39].

3. Results and discussion

The amino acid composition of VVLB4 is shown in Table 2. Our results are similar to those previously reported by Tollefsen and Kornfeld [32], indicating that VVLB4 is rich in acidic (Asp, Glu) and hydroxylic (Ser, Thr) amino acids. In contrast with previous results, however, we have found that VVLB4 contains methionine residues, which was further con-

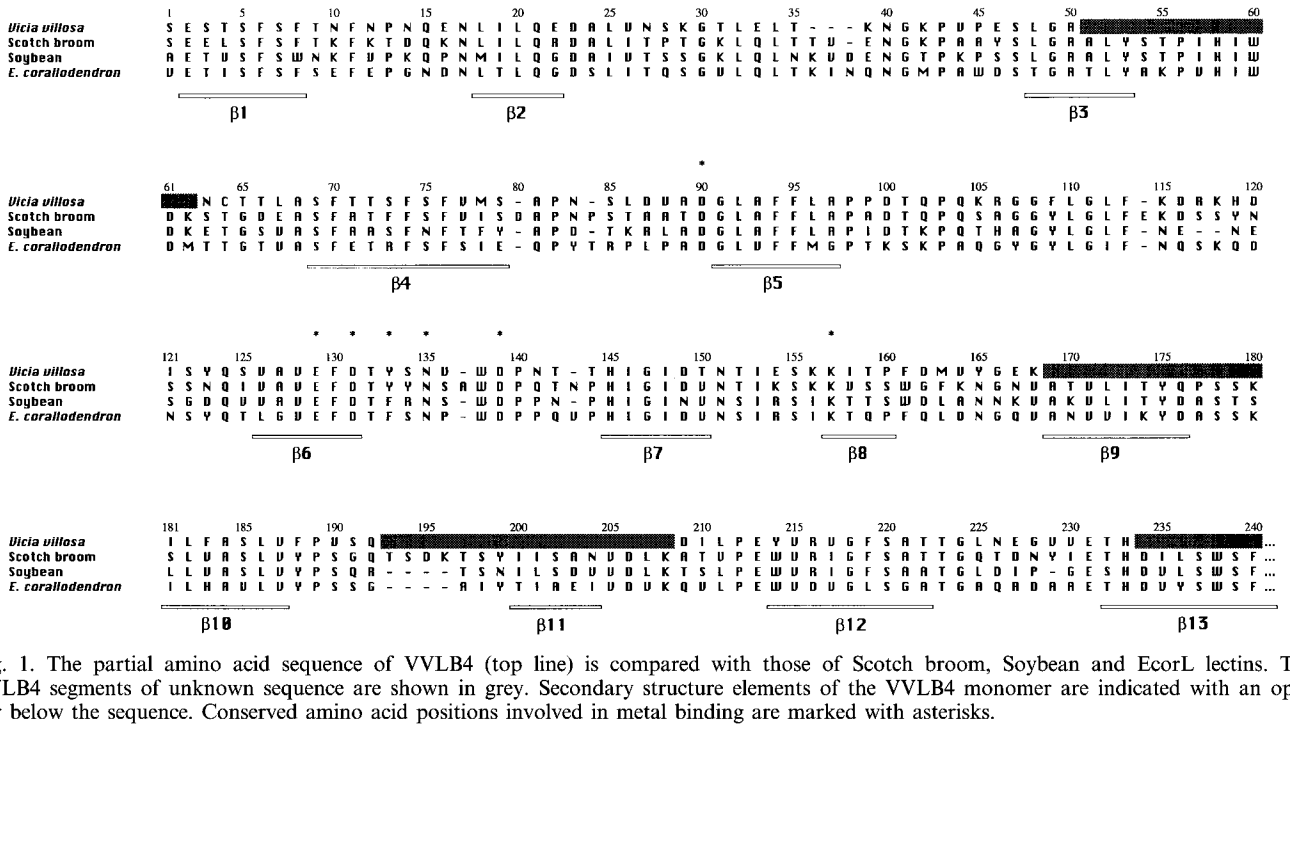

Fig. 1. The partial amino acid sequence of VVLB4 (top line) is compared with those of Scotch broom, Soybean and EcorL lectins. The VVLB4 segments of unknown sequence are shown in grey. Secondary structure elements of the VVLB4 monomer are indicated with an open bar below the sequence. Conserved amino acid positions involved in metal binding are marked with asterisks.

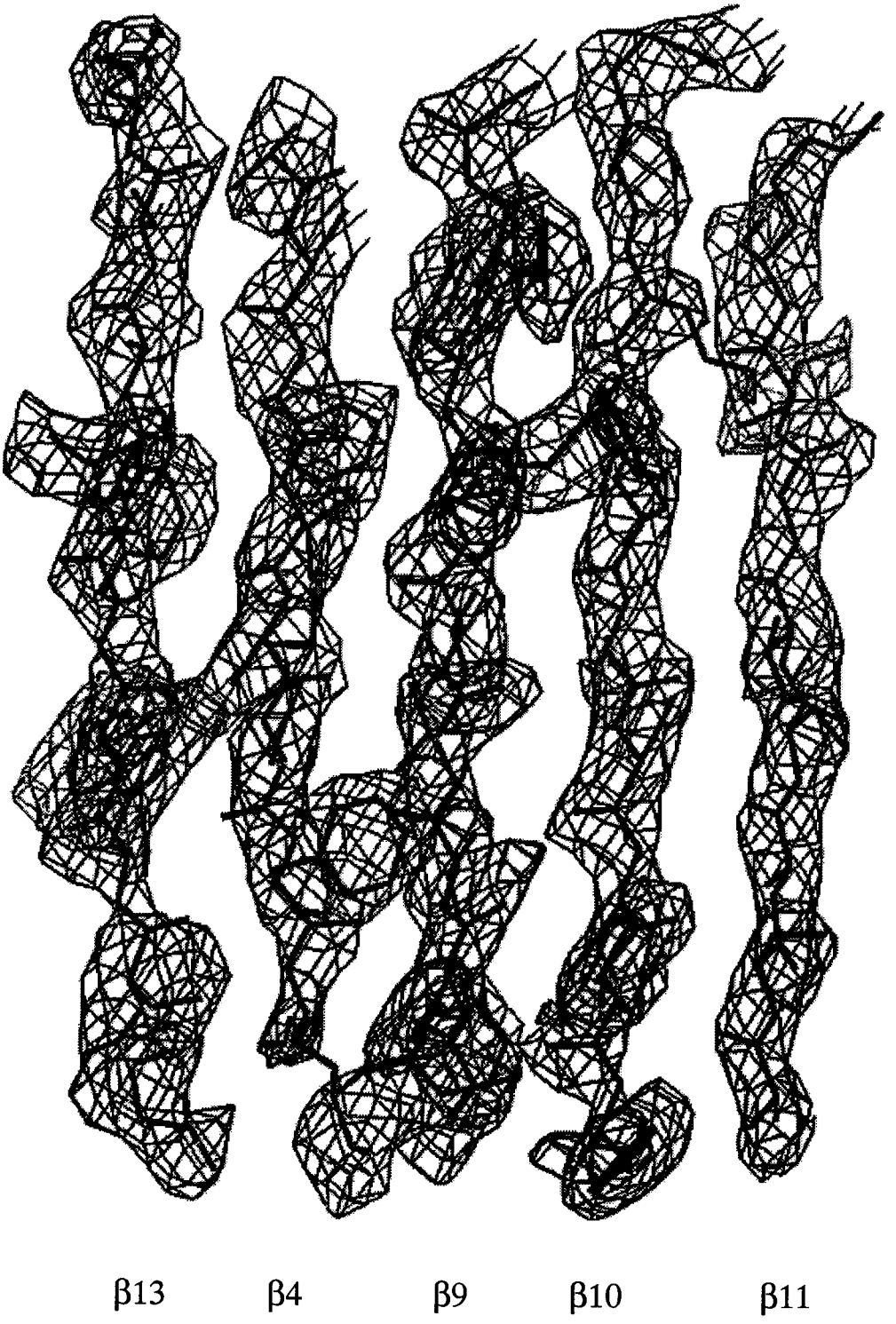

Fig. 2. Electron density map of the β-sheet which forms the internal core of the tetramer. The map, contoured at 1.1 σ, is calculated with Fourier amplitude coefficients ($2F_{obs}-F_{calc}$) and phases from the refined model. β-Strands 9, 11 and 13 for which no peptide sequence could be obtained, probably contain sites of glycosylation.

firmed by hydrolysis with cyanogen bromide and by amino acid sequence analysis (Fig. 1). The observed differences in amino acid composition could be due to the known heterogeneity of proteins obtained from different sources of *Vicia villosa* seeds.

The partial VVLB4 amino acid sequence (Fig. 1) is similar to that of other plant lectins. In particular, VVLB4 has 60% of sequence identity with *Scotch broom* lectin [33], 49% with EcorL [34], and 45% with soybean lectin [35]. The polypeptide segments that form the carbohydrate-binding site in plant lectins have been completely sequenced and, as expected, all critical residues are conserved. These include the essential Asp-90 (numbering as in Fig. 1) and Asn-135, which are invariant in all plant lectins, as well as other residues directly

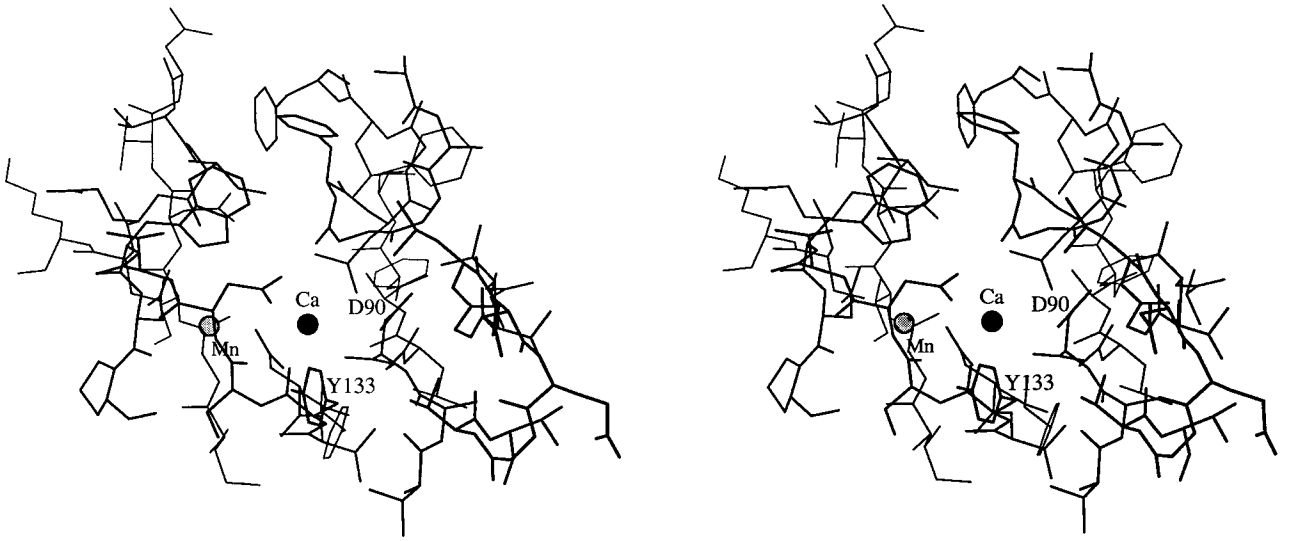

Fig. 3. Stereoview of the carbohydrate-binding site of VVLB4 viewed from the exterior of the molecule. The architecture of the site is very similar to that of other GalNAc-specific lectins.

involved in metal binding: Glu-129, Asp-131, Tyr-133, Asp-139, and Lys-157 (labelled with an asterisk in Fig. 1). Four segments of the VVLB4 polypeptide chain could not be sequenced, even after treatment with several proteolytic and mild acid methods. These segments, which probably contain the protein glycosylation sites (see below), include the C-terminal end of β-strand 3 and the subsequent loop, as well as the entire β-strands 9, 11 and 13 (Fig. 1).

Table 2
Amino acid composition of VVLB4

Residue	VVLB4[a]		VVLB4 (Ref. [32])
	Analysis	Sequence	Analysis
Aspartic acid[b]	34.1	23	35.7
Threonine	17.3	18	17.5
Serine	24.1	18	30.0
Glutamic acid[c]	20.0	17	21.1
Proline	12.1	11	12.3
Glycine	12.6	12	23.1
Alanine	14.0	9	17.1
Half-cystine[d]	0	1	0
Valine	17.6	14	7.9
Methionine	1.8	2	0
Isoleucine	14.4	8	5.7
Leucine	19.5	16	15.1
Tyrosine	7.7	4	7.2
Phenylalanine	14.4	15	11.8
Histidine	5.3	3	6.7
Lysine	10.1	9	10.4
Arginine	5.9	4	5.6
N-acetylglucosamine	2.6	N.D.[e]	3.2
Tryptophan[f]	2.0	1	1.4

Protein analysis was performed as described in Section 2. Data shown in the first column are based on the analysis of duplicate hydrolysates. For comparative purposes, amino acids recovered from partial sequence analysis according to Fig. 1 (second column), and a previously reported amino acid composition of VVLB4 [32] (third column) are included

[a]Data reported as residues/subunit.

[b]Including asparagine.

[c]Including glutamine.

[d]Cystine and methionine were determined in protein samples after performic acid oxidation.

[e]N.D., not determined.

[f]Tryptophan was determined spectrophotometrically.

The final crystallographic *R*-factor of the VVLB4 structure after refinement was 21% at 2.9 Å resolution (Table 1). The model displays good stereochemistry, as evaluated with the programs X-PLOR [30] and PROCHECK [36]. The root mean square deviations from ideal bond lengths and bond angles are 0.01 Å and 1.5°, respectively. Most main-chain dihedral angles (>99%) lie in allowed or generously allowed regions of the Ramachandran plot according to PROCHECK [36], and only six amino acid residues in the entire tetramer are positioned in disallowed regions. Two of these residues (position 59 in two subunits) correspond to a region of unknown amino acid sequence, and all six residues are positioned in solvent-exposed protein loops. The tertiary structure of the VVLB4 subunit displays the characteristic legume lectin fold consisting of a sandwich between a flat six-stranded β-sheet and a curved seven-stranded β-sheet, interconnected by loops of various lengths. Most of the subunit is well defined in the electron density maps (Fig. 2), with no ambiguity at the level of the main-chain protein backbone. However, as observed in other lectin structures, some external loops of the subunit are poorly defined in the density. The four crystallographically independent subunits have similar structures, with rms differences of 0.6–0.9 Å in the α-carbon positions between different pairs of subunits.

As in other legume lectins, each monomer contains a carbohydrate-binding site located in a shallow depression on the protein surface that shares residues with the Ca^{2+}- and Mn^{2+}-binding sites. The amino acid residues forming the sugar-binding pocket are highly conserved, and the architecture of the site is similar to that of other GalNAc/Gal-specific lectins (Fig. 3), consistent with the finding that VVLB4 belongs to this same specificity group [37]. Experimental evidence for GalNAc binding at this site was obtained by soaking VVLB4 crystals with GalNAc; the resulting difference electron density maps displayed a diffuse peak at the combining site (data not shown), although no attempt was made to model the carbohydrate structure based on these data. The alignment of several GalNAc/Gal-binding lectins reveals a remark-

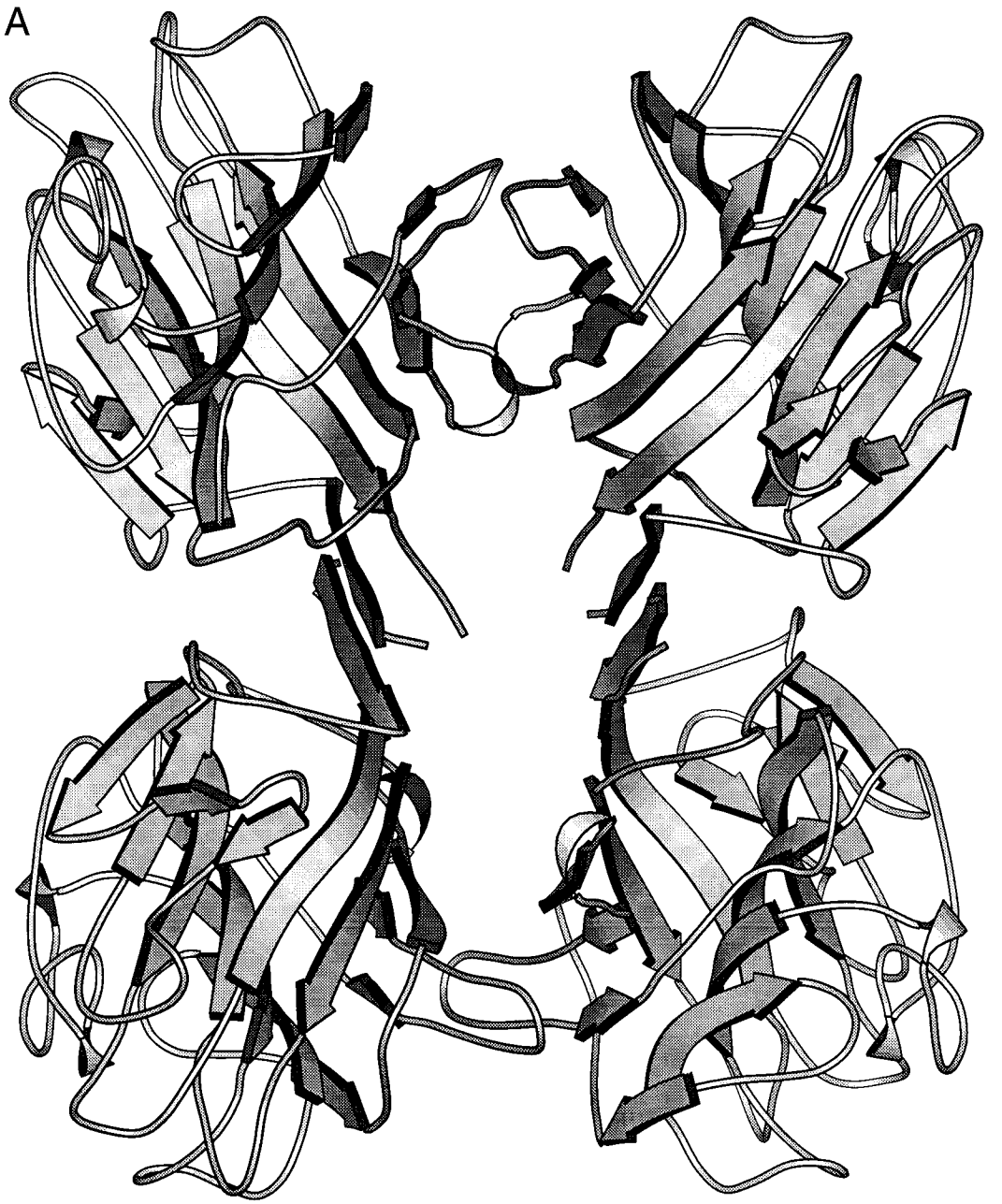

Fig. 4. Subunit association in (A) VVLB4 and (B) concanavalin A. The orientation of the two canonical dimers in VVLB4, with their long axes roughly parallel to each other, gives rise to an internal space not occupied by protein atoms at the center of the tetramer.

ably well-defined spatial relationship between the plane of the sugar ring and an aromatic residue at the primary binding site (Tyr-130 in Fig. 1), as well as conserved hydrogen-bonding interactions between the C4 hydroxyl group of the Gal moiety with Asp-90 [38]. If VVLB4 binds to the GalNAc moiety of the Tn antigen in a similar way, lectin residues outside the primary binding pocket should modulate the affinity for the peptide moiety of the Tn antigen. Indeed, Young and Oomen [18] have found that external positions at the periphery outside the central pocket display high sequence variability and could give lectins their specificity for larger oligosaccharides and glycopeptide structures.

The three isoforms of *Vicia villosa* isolectins (A_4, B_4, A_2B_2) are arranged as tetramers in solution at physiological pH [32]. Crystals of VVLB4 also contain a tetramer in the asymmetric unit (Fig. 4A) that probably corresponds to the tetrameric form in solution. The crystallographic tetramer can be described as a dimer of dimers displaying the 222 symmetry frequently observed in homotetrameric proteins. The primary mode of subunit association involves an anti-parallel side-by-side alignment of the two flat β-sheets leading to the formation of a contiguous 12-stranded β-sheet (see Fig. 4A). This particular arrangement of subunits, referred to as the 'canonical' lectin dimer, is frequently found in legume lectins (including concanavalin A, pea lectin, favin, lentil lectin and soybean agglutinin). The VVLB4 tetramer results from the face-to-face packing of the 12-stranded β-sheets of a pair of 'canonical dimers', with the long axes of the dimers running roughly

Fig. 4. (*Continued*).

parallel to each other (Fig. 4A). Interestingly, concanavalin A and soybean agglutinin also form tetramers by face-to-face packing of the extended β-sheets; however, the interactions between subunits are such that the long axes of the dimers form an angle of about 30°, thus maximizing the contact surface buried upon tetramer formation (Fig. 4B). The mode of oligomerization observed in VVLB4 gives rise to a less compact structure because the curvature of the extended sheets precludes close contact at the center of the exposed surface (Fig. 4A). Inspection of the electron density map in this interface region shows an intermediate layer of density between the two lectin dimers that cannot be accounted for by the protein model. This region is probably occupied, at least in part, by covalently bound sugars which represent approximately 6% of the total molecular weight of VVLB4 [32]. Our results of peptide sequencing reinforce this hypothesis, since the sites of *O*- and/or *N*-glycosylation probably occur within the four segments of the polypeptide chain for which we have no sequence information. As shown in Fig. 1, three of these segments code for β-strands 9, 11 and 13 which form the flat β-sheet of the monomer (Fig. 2), directly exposed to the dimer-dimer interface.

In summary, the partial amino acid sequence and the three-dimensional structure of VVLB4 show that the amino acid residues forming the carbohydrate-binding pocket are largely conserved between GalNAc/Gal-specific lectins. This results indicate that other residues at more external positions are involved in modulating the specificity of VVLB4 for the Tn determinant. We have also shown that, as observed in other studies, glycosylation can modulate the quaternary organiza-

tion of the protein and give rise to novel tetrameric arrangements of the lectin subunits. Further studies of this isolectin in complex with various Tn-containing molecules are in progress, and should provide invaluable information about the specific recognition of this tumor-associated antigen.

Acknowledgements: We are grateful to Frederick Saul for helpful discussions. This work was partially supported by grants from CONICYT-BID, Institut Pasteur (Paris), La Ligue contre le Cancer, ECOS France-Uruguay Program, Comisión Honoraria de Lucha contra el Cancer (Uruguay) and CSIC (Universidad de la República, Uruguay). Gisele Tavares was supported by a grant from the Brazilian National Research Council CNPq.

References

[1] Lis, H. and Sharon, N. (1986) Annu. Rev. Biochem. 55, 35–67.
[2] Imberty, A. and Pérez, S. (1994) Glycobiology 4, 351–366.
[3] Sharon, N. and Lis, H. (1989) Science 246, 227–234.
[4] Wilson, I.A., Skehel, J.J. and Wiley, D.C. (1981) Nature 289, 366–378.
[5] Weis, W.I., Kahn, R., Fourme, R., Drickamer, K. and Hendrickson, W.A. (1991) Science 253, 1608–1615.
[6] Liao, D.I., Kapadia, G., Ahmed, H., Vasta, G.R. and Herzberg, O. (1994) Proc. Natl. Acad. Sci. USA 91, 1428–1432.
[7] Lobsanov, Y.D., Gitt, M.A., Leffler, H., Barondes, S.H. and Rini, J.M. (1993) J. Biol. Chem. 268, 27034–27038.
[8] Reeke, G.N., Becker, J.W. and Edelman, G.M. (1975) J. Biol. Chem. 250, 1525–1547.
[9] Einspahr, H., Parks, E.H., Suguna, K., Subramanian, E. and Suddath, F.L. (1986) J. Biol. Chem. 261, 16518–16527.
[10] Reeke, G.N. and Becker, J.W. (1986) Science 234, 1108–1111.
[11] Bourne, Y., Abergel, C., Cambillau, C., Frey, M., Rouge, P. and Fontecilla-Camps, J.C. (1990) J. Mol. Biol. 214, 571–584.
[12] Delbaere, L.T.J., Vandonselaar, M., Prasad, L., Quail, J.W., Wilson, K.S. and Dauter, Z. (1993) J. Mol. Biol. 230, 950–965.
[13] Shaanan, B., Lis, H. and Sharon, N. (1991) Science 254, 862–866.
[14] Loris, R., Steyaert, J., Maies, D., Lisgarten, J., Pickersgill, R. and Wyns, L. (1993) Biochemistry 32, 8772–8781.
[15] Dessen, A., Gupta, D., Sabesan, S., Brewer, C.F. and Sacchettini, J.C. (1995) Biochemistry 34, 4933–4942.
[16] Banerjee, R., Mande, S., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S., Suguna, K., Surolia, A. and Vijayan, M. (1994) Proc. Natl. Acad. Sci. USA 91, 227–231.
[17] Sharon, N. and Lis, H. (1990) FASEB J. 4, 3198–3208.
[18] Young, N.M. and Oomen, R.P. (1992) J. Mol. Biol. 228, 924–934.
[19] Dausset, J., Moullec, J. and Bernard, J. (1959) Blood 14, 1079–1093.
[20] Springer, G. (1984) Science 24, 1198–1206.
[21] Tollefsen, S. and Kornfeld, R. (1983) J. Biol. Chem. 258, 5172–5176.
[22] Osinaga, E., Babino, A., Grosclaude, J., Cairoli, E., Batthyany, C., Bianchi, S., Signorelli, S., Varangot, M., Musé, I. and Roseto, A. (1996) Int. J. Oncol. 8, 401–406.
[23] Landon, M. (1977) Meth. Enzymol. 47, 145–149.
[24] Huang, H.V., Bond, M.W., Hunkapillar, M.W. and Hood, L.E. (1983) Meth. Enzymol. 91, 318–324.
[25] Eisele, J.-L., Tello, D., Osinaga, E., Roseto, A. and Alzari, P.M. (1993) J. Mol. Biol. 230, 670–672.
[26] Otwinowski, Z. (1993) in: Proceedings of the CCP4 Study Weekend: Data Collection and Processing (Sawyer, L., Isaacs, N. and Bailey, S., Eds.), pp. 56–62, SERC, Daresbury, England.
[27] Navaza, J. (1994) Acta Crystallogr. A50, 157–163.
[28] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[29] Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) Acta Crystallogr. A47, 110–119.
[30] Brünger, A.T., Kuriyan, J. and Karplus, M. (1987) Science 235, 458–460.
[31] Engh, R.A. and Huber, R. (1991) Acta Crystallogr. 47, 392–400.
[32] Tollefsen, S. and Kornfeld, R. (1983) J. Biol. Chem. 258, 5165–5171.
[33] Konami, Y., Yamamoto, K., Osawa, T. and Irimura, T. (1992) J. Biochemistry 112, 366–375.
[34] Adar, R., Richardson, M., Lis, H. and Sharon, N. (1989) FEBS Lett. 257, 81–85.
[35] Vodkin, L., Rhodes, P. and Goldberg, R. (1983) Cell 34, 1023–1031.
[36] Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) J. Appl. Crystallogr. 26, 283–291.
[37] Puri, K.D., Gopalakrishnan, B. and Surolia, A. (1992) FEBS Lett. 312, 208–212.
[38] Rini, J. (1995) Annu. Rev. Biophys. Biomol. Struct. 24, 551–577.
[39] Brünger, A.T. (1992) Nature 335, 472–474.