

The crystal structure of a plant lectin in complex with the Tn antigen

Alvaro Babino^a, Diana Tello^b, Adriana Rojas^{b,1}, Sylvie Bay^c, Eduardo Osinaga^a,
Pedro M. Alzari^{b,*}

^aDepartamento de Bioquímica, Facultad de Medicina, Av. Gral Flores 2125, Montevideo, Uruguay

^bUnité de Biochimie Structurale (CNRS URA 2185), Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cedex 15, France

^cUnité de Chimie Organique (CNRS URA 2128), Institut Pasteur, 28 rue du Dr. Roux, 75724 Paris Cedex 15, France

Received 6 November 2002; revised 4 January 2003; accepted 7 January 2003

First published online 17 January 2003

Edited by Hans Eklund

Abstract The structure of the tetrameric *Vicia villosa* isolectin B4 (VVLB4) in complex with a cancer antigen, the Tn glycopeptide (GalNAc-O-Ser), was determined at 2.7 Å resolution. The *N*-acetylgalactoside moiety of the ligand binds to the primary combining site of VVLB4 in a similar way as observed for other Gal/GalNAc-specific plant lectins. The amino acid moiety of the Tn antigen is largely exposed to the solvent and makes few contacts with the protein. The structure of the complex provides a framework to understand the differences in the strength of VVLB4 binding to different sugars and emphasizes the role of a single protein residue, Tyr127, as a structural determinant of Tn-binding specificity.

© 2003 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.

Key words: X-ray structure; Cancer antigen; Tn-binding protein; Carbohydrate recognition; *Vicia villosa* lectin

1. Introduction

The Tn determinant (GalNAc- α -O-Ser/Thr) is one of the most specific human tumor-associated structures [1]. Expressed in mucin-type *O*-glycans of cancer cells [2], Tn is a useful biomarker because it is expressed early in transformed cells, both in human [3,4] and in animal carcinogenesis [5], and a direct correlation has been shown between carcinoma aggressiveness and the density of this antigen [6]. Although Tn is a glycan structure, it has been reported that some Tn-glycosylated peptides may be capable of binding to proteins of the major histocompatibility complex and inducing T-cell lymphocyte activation [7], thus providing a potentially important therapeutic target [8]. The Tn determinant has also been implicated in organotropic metastasis of tumor cells [9] and HIV infection [10], further stressing the relevance of understanding the molecular basis of Tn-binding specificity.

The Tn antigen was characterized using both monoclonal antibodies and lectins. Among plant lectins the best known are the homotetrameric isolectin B4 from *Vicia villosa* (VVLB4) seeds [11], the lectin from *Salvia sclarea* [12] and

isolectin A4 from *Griffonia simplicifolia* lectin I [13]. While all Tn-binding proteins recognize the GalNAc-O-Ser/Thr determinant, they show subtle differences when used as biochemical or histochemical markers for cancerous tissues. Using synthetic Tn glycopeptide models, we have determined that VVLB4 and *S. sclarea* lectin bind to one Tn determinant, whereas specific anti-Tn monoclonal antibodies required at least two Tn residues for recognition [12,14]. These results suggest that Tn residues organized in clusters are essential for antibody recognition. Moreover, it has been reported that Tn clusters are also important for the recognition by the Tn-specific human macrophage C-type lectin, which reacts poorly with glycopeptides bearing mono-Tn motifs [15].

Understanding the molecular basis of the specificity of Tn-binding lectins is of considerable interest both for elucidating the mechanism of specific carbohydrate recognition and for engineering novel binding activities in legume lectins. VVLB4 belongs to the family of legume lectins with overall specificity for Gal/GalNAc and was shown to specifically agglutinate Tn-exposing red blood cells [16,17]. We have previously reported the partial protein sequence and the unliganded crystal structure of VVLB4 [18]. The tertiary structure of the VVLB4 subunit displays the characteristic legume lectin fold consisting of a β -sandwich formed by a flat six-stranded β -sheet and a curved seven-stranded β -sheet interconnected by loops of variable length [19]. Here we report the crystal structure of VVLB4 in complex with the Tn antigen (GalNAc-O-Ser) at 2.7 Å resolution. The structure reveals the atomic details of the carbohydrate-binding site and provides the first structural data for understanding the molecular basis of Tn specificity in plant lectins.

2. Materials and methods

The Tn antigen 3-*O*-(acetamido-2-deoxy- α -D-galactopyranosyl)-L-serine was prepared as described [20] by glycosylation of *N*^α-(benzyloxycarbonyl)-L-serine benzyl ester with 3,4,6-tri-*O*-acetyl-2-deoxy- β -D-galactopyranosyl chloride (obtained from tri-*O*-acetyl-D-galactal) using Ag₂CO₃/AgClO₄ as catalysts, followed by the reduction and acetylation of the 2-position [21]. The amino acid protective groups were then removed by hydrogenolysis under Pd on charcoal. The final deprotection of the carbohydrate moiety at pH 11 in sodium methoxide afforded the Tn antigen with the expected analytical data (nuclear magnetic resonance, mass spectra).

Most of the VVLB4 sequence (85%), including all residues defining the monosaccharide-binding site, was obtained by protein digestion, peptide purification and sequencing as described previously [18]. The rest of the amino acid sequence was deduced from inspection of the electron density map and the nucleotide sequence of the VVLB4 gene (manuscript in preparation).

*Corresponding author. Fax: (33)-1-45 68 86 04.
E-mail address: alzari@pasteur.fr (P.M. Alzari).

¹ Present address: Laboratório de Cristalografia de Proteínas, Instituto de Física de São Carlos, Universidade de São Paulo, CP 369, 13560-970 São Carlos, SP, Brazil.

Crystals of the VVLB4–Tn complex were obtained using the hanging drop method under similar (but not identical) conditions as those reported for the unliganded protein [18,22]. The best crystals of the complex were grown from a solution containing 170 mM acetate, pH 4.7, and MPD 18%. The crystals were very unstable under the X-ray beam and difficult to freeze. However, a complete data set at 2.7 Å resolution could be collected using synchrotron radiation (beamline DW32, LURE) from a single crystal mounted on a capillary and held at 4°C during exposure to the X-ray beam. The crystal belongs to space group $P4_3$, with unit cell dimensions $a = b = 85.5$ Å, $c = 153.6$ Å, and contains one tetramer in the asymmetric unit. Data reduction was carried out with the HKL program package [23] and other programs from the CCP4 suite [24]. Data statistics are shown in Table 1.

The structure was solved by molecular replacement techniques with the program AMoRe [25], using the tetrameric structure of the unliganded lectin previously determined at 2.9 Å resolution. Crystallographic refinement was carried out by alternate cycles of maximum likelihood refinement with program REFMAC [26] and manual model building with program O [27]. During the first rounds of refinement, tetrameric non-crystallographic symmetry (NCS) was imposed on the model. From the early stages of refinement, the Tn ligand was clearly visible in difference Fourier maps and omit maps. As refinement progressed, the NCS restraints were released to check whether the different crystal contacts induced structural rearrangements of the individual monomers. A few water molecules were added to the model from the inspection of difference Fourier maps and stereochemical criteria, and accepted if they were stable during refinement. The parameters of the model after the final refinement cycle are listed in Table 1. The atomic coordinates of the complex have been deposited with the Protein Data Bank, code 1N47.

3. Results and discussion

The crystal structure of the VVLB4–Tn complex has been determined at 2.7 Å resolution (Table 1). The crystallographic model of the complex (Fig. 1) includes four independent monomers of 233 amino acid residues each, four Tn residues and four clusters composed of two *N*-acetylglucosamine and one fucose residues (Fig. 2A) covalently attached to the protein at amino acid residue Asn61. The model has a good stereochemistry, with most non-Pro, non-Gly residues displaying main-chain dihedral angles that fall within the most favored or additionally allowed regions of the Ramachandran plot [28]. Only two amino acid residues in each monomer

Table 1
Data collection and refinement statistics

<i>Data collection</i>	
Resolution range (Å)	20–2.7 (2.78–2.7)
Measured reflections	144 637
Unique reflections	30 037
$\langle I/\sigma(I) \rangle$	9.1
R_{merge} (%) ^a	10.7 (46.7)
Completeness (%)	99.8 (99.8)
<i>Refinement</i>	
Resolution range (Å)	20–2.7
Reflections used in refinement	28 662
R_{factor} (%) ^b	18.8
Free R_{factor} (%) ^c	23.8
Protein atoms	7000
Sugar/ligand atoms	236
Water molecules	57
rms bond lengths (Å)	0.014
rms bond angles (degrees)	1.70
Ramachandran outliers (%)	1.1
Mean B value of protein atoms (Å ²)	32.8
Mean B value of ligand atoms (Å ²)	54.0

^a $R_{\text{merge}} = \sum_{hkl} |I - \langle I \rangle| / \sum_{hkl} I$.

^b $R_{\text{factor}} = \sum |F_{\text{calc}} - F_{\text{obs}}| / \sum |F_{\text{obs}}|$.

^c The free R_{factor} as defined by Brünger [38] was calculated for a subset (5%) of reflections that were excluded from the refinement.

Table 2

Protein–ligand hydrogen-bonding interactions (range of values found in the four crystallographically independent monomers)

Protein atom	Tn atom	Distance (Å)
Asp85–Oδ ₁	3–OH	2.6–2.8
Asn129–Nδ ₂	3–OH	2.9–3.2
Asp85–Oδ ₂	4–OH	2.6–2.9
Leu213–N	4–OH	3.0–3.4
Tyr127–OH	NH ₃ (Ser)	2.8–3.4 ^a
Gly103–N	7–OH (<i>N</i> -acetyl)	2.9–3.4 ^a

^a This interaction is not observed in one of the four monomers.

(Glu22 and Asn61) have unfavorable (ϕ , ψ) values. However, these residues show an identical conformation in the four monomers and are well defined in the electron density map, suggesting that this is a genuine feature of the structure. Besides Asn61, a second putative *N*-glycosylation site had been predicted at Asn134 from the amino acid sequence [18]. However, no *N*-glycosylation of this residue is detected in the electron density map and the Asn134 side chain is involved in crystal packing contacts, implying that a sugar residue at this position would interfere with crystal formation.

The oligomeric structure of VVLB4 involves two ‘canonical’ plant lectin dimers packed face to face in a parallel fashion (Fig. 1A). The dimers interact with each other through the outermost strands of their two curved 12-stranded sheets, creating a large channel in the middle of the tetramer (Fig. 1B). The VVLB4 tetramer is similar to that observed in other lectins (such as soybean agglutinin [29] and lectins from *Phaseolus vulgaris* [30], *Dolichos biflorus* [31] and *Robinia pseudoacacia* [32]) and appears to represent a ‘canonical tetramer’ in plant lectins [19]. Although the lectin–Tn complex crystallizes in a different space group (tetragonal) than the unliganded protein (orthorhombic), the overall structure of the tetramer is very similar, indicating that the protein structure is not significantly modified by ligand binding.

The Tn residue (Fig. 2B) binds to a conserved cavity, which is defined by four separate loops of the polypeptide chain and represents the primary carbohydrate-binding site in plant lectins [19]. In this cavity, four conserved sugar-binding residues (Asp85, Gly103, Tyr127 and Asn129) occupy positions similar to those observed in other legume lectins. The Ca²⁺- and Mn²⁺-binding sites (Fig. 1A) also exhibit well conserved coordination spheres. The Ca²⁺ ion is coordinated by the side chains of Asp125, Asn129 and Asp132 and the main-chain carbonyl group of Tyr127, whereas the Mn²⁺ ion is coordinated by the carboxylate groups of Glu123, Asp125, Asp132 and the imidazole group of His137. Water molecules complete the coordination of the cations, but only a few of them could be unambiguously identified and modelled in the present structure, due to the limited resolution of the data.

The galactopyranoside ring of the Tn residue is stacked between the aromatic side chain of Tyr127 and the aliphatic side chain of Leu213 (Fig. 3). The sugar hydroxyl groups 3–OH and 4–OH make hydrogen-bonding interactions with the side chains of two polar residues, Asn129 and Asp85 at the bottom of the cavity, and with the main-chain NH group of Gly103 (Table 2). This pattern of interactions between the ligand and the strictly invariant Asp–Asn–Gly triad is conserved in other Gal/GalNAc-binding lectins and appears to be a major determinant of monosaccharide specificity in this lectin family. As a consequence, the sugar moiety of the Tn

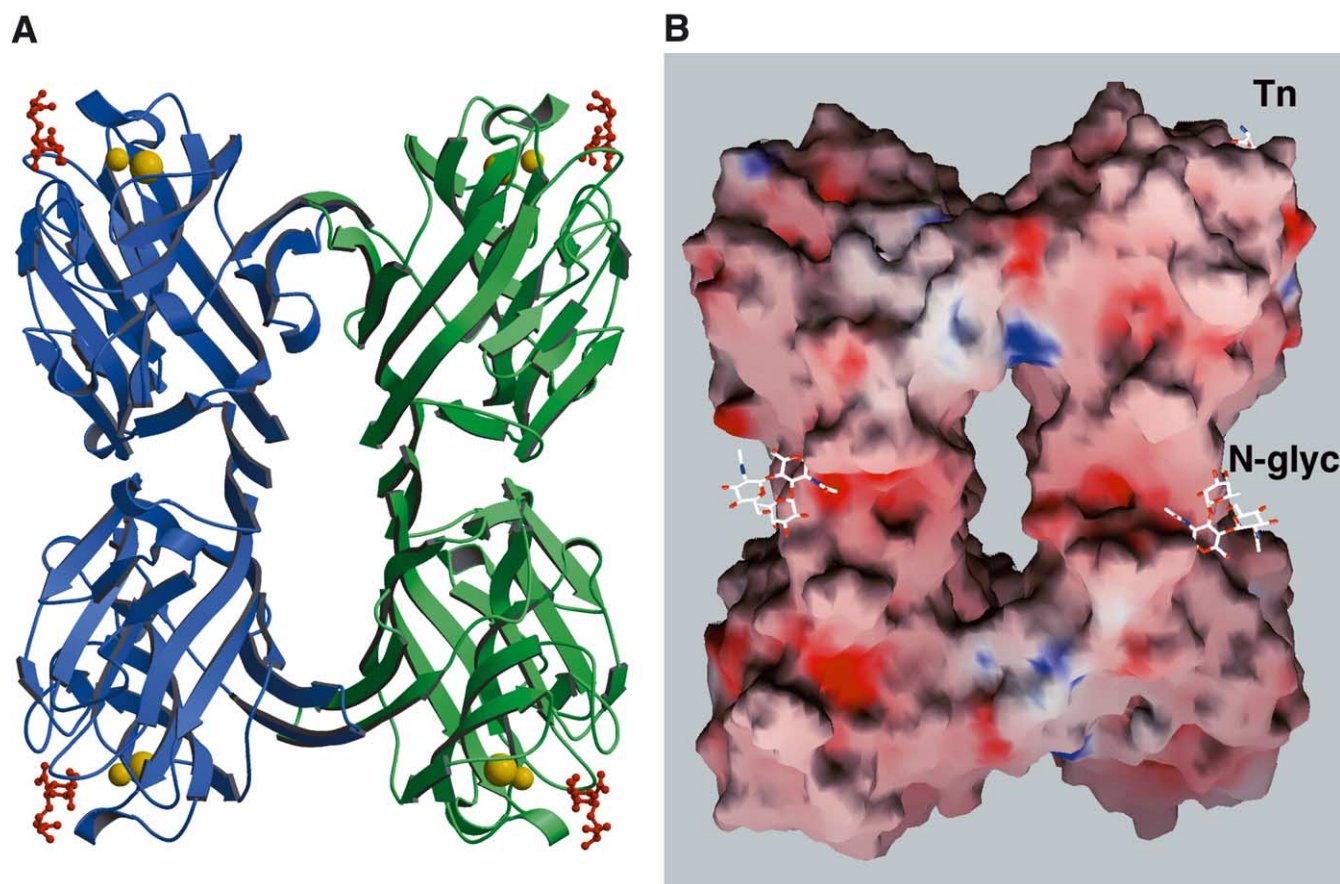


Fig. 1. The overall structure of VVLB4 in complex with Tn. A: The two canonical lectin dimers are shown in different colors. Ca^{2+} and Mn^{2+} are shown as small and large yellow spheres, respectively. The four observed Tn molecules are shown in red. B: Molecular surface of the VVLB4 tetramer. The sugar residues covalently attached to the protein are shown in stick representation, and the amino acid moiety of one Tn residue is seen at the upper right corner. Figure drawn with the programs MOLSCRIPT [39] and GRASP [40].

molecule binds in a similar orientation as seen in other Gal/GalNAc–lectin complexes (Fig. 3). However, small differences are observed. For example, an additional hydrogen bond is observed in the GalNAc–EcorL complex between the side chain of Gln219 of the lectin and the O6 oxygen of the sugar ring [33]. This interaction is not observed in VVLB4 because the Gln residue is substituted by a shorter asparagine residue (Asn214).

Some lectins of the Gal/GalNAc-specific family have the ability to discriminate between galactose and *N*-acetylgalactosamine. For instance, peanut agglutinin is a Gal-specific lectin which does not bind GalNAc [34], probably because longer loops and bulkier residues in the region surrounding the 2-OH hydroxyl of the galactopyranoside ring sterically prevent the accommodation of the *N*-acetyl group at this position [35]. Other lectins of this family (such as EcorL) display a similar affinity for the *N*-acetyl or the hydroxyl substituent at C2 [33], while still others (such as VVLB4, the *D. biflorus* lectin DBL, or the lectin RPbAI from *R. pseudoacacia*) exhibit a significantly higher affinity for GalNAc than Gal residues [11,31,32]. Comparison of the VVLB4–Tn structure with that of the EcorL–GalNAc complex (Fig. 3) reveals that a hydrogen-bonding interaction between the oxygen atom of the *N*-acetyl group and the main-chain NH group of VVLB4 Gly103 (Table 2) is favored by the presence of a second glycine residue at position 102 (DBL and RPbAI also have a

glycine at the equivalent position). Instead, when a bulkier residue is found at this position (such as Tyr106 in EcorL, see Fig. 3), its C β atom is in contact with the *N*-acetyl group of the ligand. These interactions promote a different orientation of the sugar residue in the binding pocket, weakening the hydrogen-bonding interaction, and could account for the differences in Gal/GalNAc specificity [31].

This ability to discriminate between Gal and GalNAc is crucial for both the biological activity of lectins and their usefulness as tools in biology and medicine. In some cases, subtle changes of the protein may account for its recognition specificity. For instance, the crystal structures of isolectins I-A and I-B of *G. simplicifolia* showed that the modification at a single amino acid position, Ala106 in I-A versus Glu106 in I-B, is the basis for their distinct Gal/GalNAc specificities [36]. Seeds of *V. villosa* also contain other lectins, among which the tetrameric isolectin A4 that specifically recognizes Gal residues [16]. Thus, the *V. villosa* isolectins could provide an excellent model to study the molecular basis of differential monosaccharide binding in plant lectins.

The aglycone (serine) moiety of the Tn molecule is completely exposed to the solvent (Fig. 1) and displays large temperature factors, higher than those of the protein monomers (Table 1). Thus, the Tn serine could easily accommodate additional amino acid residues without affecting protein–ligand interactions, thus explaining why VVLB4 binds Tn with a

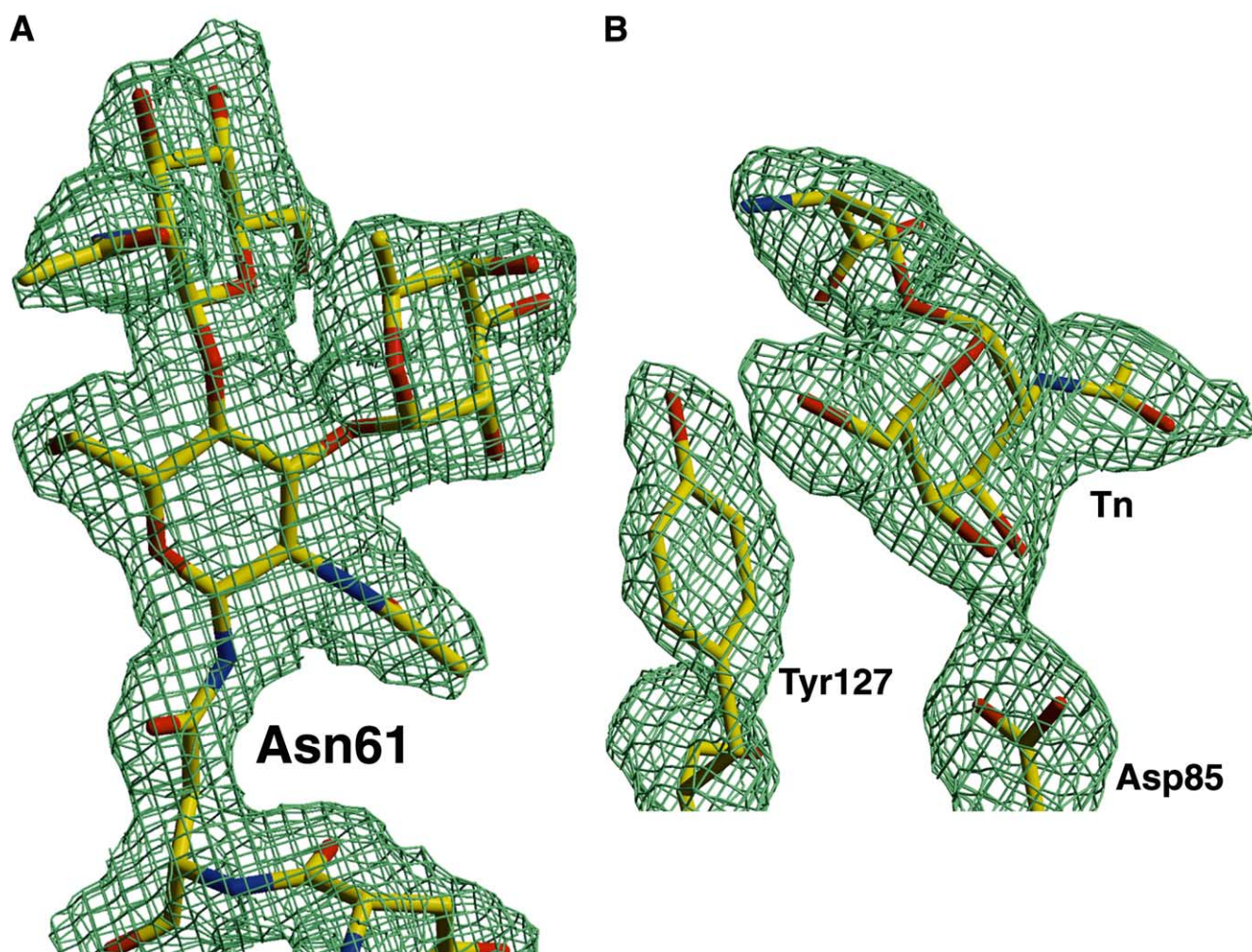


Fig. 2. Electron density ($2F_o - F_c$) maps, contoured at 1σ , showing (A) two GlcNAc and one fucose residues attached to Asn61, and (B) the Tn residue. Figure drawn with the programs XTALVIEW [41] and RASTER3D [42].

similar affinity either as a single molecule or as a Tn cluster [14]. Due to the α -linkage in Tn, the serine moiety of the ligand is brought into contact with residues from the segment Tyr127–Trp131. Protein–ligand interactions in this region should therefore account for Tn specificity, since no other region of the lectin monomer is in contact with the Tn serine. The crystal structure reveals a single hydrogen-bonding inter-

action between the hydroxyl group of Tyr127 and the ligand NH_3 group (Fig. 3, Table 2). This interaction is seen in some (but not all) independent lectin monomers, probably because of the high mobility of the ligand. Indeed, while the protein backbone atoms of the different monomers within the tetramer can be superimposed with rms deviations of 0.21–0.23 Å, these values fall in the range 0.25–0.65 Å for the Tn

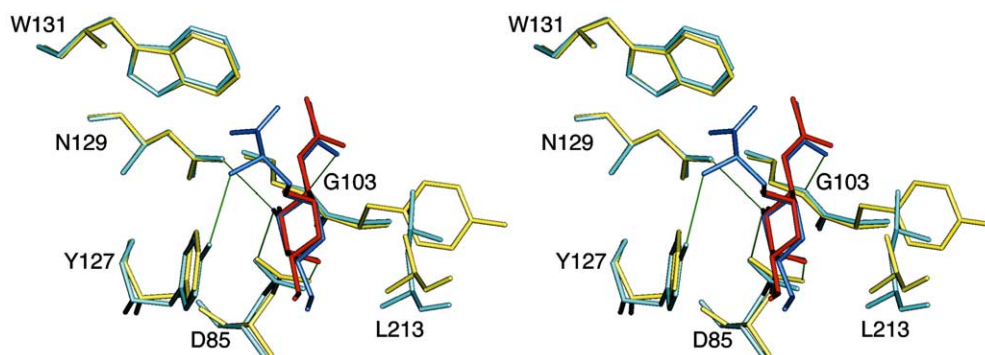


Fig. 3. Stereoview of the carbohydrate-binding region in the VVLB4–Tn complex (shown in light blue and dark blue) superimposed with that of the *Erythrina corallodendron* lectin (EcorL) in complex with lactose (PDB code 1ax0, shown in yellow and red). VVLB4 residues are labeled; selected hydrogen-bonding interactions discussed in the text are also shown.

antigen and most of these differences can be assigned to the highly exposed serine moiety.

Although the limited resolution of this study precludes a full analysis of protein hydration, it is possible that protein–ligand hydrogen-bonding interactions mediated by water molecules could also contribute to the higher affinity for the Tn residue over GalNAc, as seen for instance in the complex between peanut lectin and the T-antigen [37]. Nevertheless, it remains true that the only direct hydrogen bond between protein atoms and the serine moiety of the Tn molecule involves the hydroxyl group of Tyr127, and this interaction must therefore play an important role as determinant of Tn specificity in VVLB4. In order to clarify the functional role of Tyr127 in the recognition of Tn determinant, the molecular cloning of VVLB4 is under way to carry out site-directed mutagenesis studies. Indeed, there are not many Gal/GalNAc-specific lectins having a tyrosine residue at this position, and it would be interesting to investigate whether these proteins, such as the Scotch broom (*Cytisus scoparius*) lectin, are also able to recognize the Tn antigen.

Acknowledgements: This work was partially supported by grants from the Institut Pasteur, La Ligue contre le Cancer (France) and the ECOS France–Uruguay program.

References

- [1] Hakomori, S. (1989) Adv. Cancer Res. 52, 257–331.
- [2] Brockhausen, I. (1999) Biochim. Biophys. Acta 1473, 67–95.
- [3] Itzkowitz, S., Bloom, E., Lau, T. and Kim, Y. (1992) Gut 33, 518–523.
- [4] Terasawa, K., Furumoto, H., Kamada, M. and Aono, T. (1996) Cancer Res. 56, 2229–2232.
- [5] Babino, A., Oppezzo, P., Bianco, S., Barrios, E., Berois, N., Navarrete, H. and Osinaga, E. (2000) Int. J. Cancer 86, 753–759.
- [6] Springer, G. (1995) Crit. Rev. Oncogen. 6, 57–85.
- [7] Galli-Stampino, L., Meinjohanns, E., Frische, K., Meldal, M., Jensen, T., Werdelin, O. and Mouritsen, S. (1997) Cancer Res. 57, 3214–3222.
- [8] Lo-Man, R., Bay, S., Vichier-Guerre, S., Dériaud, E., Cantacuzène, D. and Leclerc, C. (1999) Cancer Res. 59, 1520–1524.
- [9] Schlepper-Schafer, J. and Springer, G. (1989) Biochim. Biophys. Acta 1013, 266–272.
- [10] Hansen, J.E.S., Nielsen, C., Arendrup, M., Olofsson, S., Mathiesen, L., Nielsen, J. and Clausen, H. (1991) J. Virol. 65, 6461–6467.
- [11] Puri, K., Gopalakrishnan, B. and Surolia, A. (1992) FEBS Lett. 312, 208–212.
- [12] Medeiros, A., Bianchi, S., Calvete, J.J., Balter, H., Bay, S., Robles, A., Cantacuzène, D., Nimtz, M., Alzari, P.M. and Osinaga, E. (2000) Eur. J. Biochem. 267, 1434–1440.
- [13] Wu, A., Wu, J., Chen, Y.-Y., Song, S.-C. and Kabat, E. (1999) Glycobiology 9, 1161–1170.
- [14] Osinaga, E., Bay, S., Tello, D., Babino, A., Pritsch, O., Assemet, K., Cantacuzène, D., Nakada, H. and Alzari, P.M. (2000) FEBS Lett. 469, 24–28.
- [15] Iida, S., Yamamoto, K. and Irimura, T. (1999) J. Biol. Chem. 274, 10697–10705.
- [16] Tollefsen, S. and Kornfeld, R. (1983) J. Biol. Chem. 258, 5165–5171.
- [17] Tollefsen, S. and Kornfeld, R. (1983) J. Biol. Chem. 258, 5172–5176.
- [18] Osinaga, E., Tello, D., Batthyany, C., Bianchet, M., Tavares, G., Durán, R., Cerveñansky, C., Camoin, L., Roseto, A. and Alzari, P. (1997) FEBS Lett. 412, 190–196.
- [19] Loris, R., Hamelryck, T.W., Bouckaert, J. and Wyns, L. (1998) Biochim. Biophys. Acta 1383, 9–36.
- [20] Lemieux, R.U. and Ratcliffe, R.M. (1979) Can. J. Chem. 57, 1244–1251.
- [21] Paulsen, H. and Hölck, J.-P. (1982) Carbohydr. Res. 109, 89–107.
- [22] Eisele, J.-L., Tello, D., Osinaga, E., Roseto, A. and Alzari, P.M. (1993) J. Mol. Biol. 230, 670–672.
- [23] Otwinowski, Z. and Minor, W. (1997) Meth. Enzymol. 276, 307–325.
- [24] Collaborative Computational Project 4 (1994) Acta Crystallogr. D50, 760–763.
- [25] Navaza, J. (1994) Acta Crystallogr. A50, 157–163.
- [26] Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S. and Dodson, E.J. (1999) Acta Crystallogr. D55, 247–255.
- [27] Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Acta Crystallogr. A47, 110–119.
- [28] Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) J. Appl. Crystallogr. 26, 283–291.
- [29] Dessen, A., Gupta, S., Dabesan, S., Brewer, C.F. and Sachettini, J.C. (1995) Biochemistry 34, 4933–4942.
- [30] Hamelryck, T.W., Dao-Thi, M.H., Poortmans, F., Chrispeels, M.J., Wyns, V.L. and Loris, R. (1996) J. Biol. Chem. 271, 20479–20485.
- [31] Hamelryck, T., Loris, R., Bouckaert, J., Dao-Thi, M.-H., Strecker, G., Imberty, A., Fernandez, E., Wyns, L. and Etzler, M. (1999) J. Mol. Biol. 286, 1161–1177.
- [32] Rabijns, A., Verboven, C., Rougé, P., Barre, A., Van Damme, E.J.M., Peumans, W.J. and De Ranter, C.J. (2001) Proteins Struct. Funct. Genet. 44, 470–478.
- [33] Elgavish, S. and Shaanan, B. (1998) J. Mol. Biol. 277, 917–932.
- [34] Swamy, M.J., Gupta, D., Mahanta, S.K. and Surolia, A. (1991) Carbohydr. Res. 213, 59–67.
- [35] Sharma, V., Srinivas, V.R., Adhikari, P., Vijayan, M. and Surolia, A. (1998) Glycobiology 8, 1007–1012.
- [36] Lescar, L., Loris, R., Mitchell, E., Gautier, C., Chazalet, V., Cox, V., Wyns, L., Pérez, S., Breton, C. and Imberty, A. (2002) J. Biol. Chem. 277, 6608–6614.
- [37] Ravishankar, R., Ravindran, M., Suguna, K., Surolia, A. and Vijayan, M. (1997) Curr. Sci. 72, 855–861.
- [38] Brünger, A.T. (1992) Nature 355, 472–474.
- [39] Kraulis, P.J. (1991) J. Appl. Crystallogr. 24, 946–950.
- [40] Nicholls, A., Sharp, K. and Honig, B. (1991) Proteins Struct. Funct. Genet. 11, 281–296.
- [41] McRee, D.E. (1992) J. Struct. Biol. 125, 156–165.
- [42] Merritt, E.A. and Bacon, D.J. (1997) Methods Enzymol. 277, 505–524.