Crystal structure of the peanut lectin – T-antigen complex. Carbohydrate specificity generated by water bridges

R. Ravishankar, M. Ravindran, K. Suguna, A. Surolia and M. Vijayan

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Peanut lectin binds with high specificity to the tumour-associated disaccharide Galβ1–3GalNAc, generally known as T-antigen. The crystal structure of the complex of the lectin with the disaccharide has been determined at 2.5 Å resolution. Comparison of the structure with that of the corresponding complex with lactose reveals that the specificity of the lectin for T-antigen is generated primarily by two specific water-mediated interactions, probably the first instance where water-bridges have been demonstrated to be responsible for generating specificity in protein–carbohydrate interactions. The elucidation of the structure of peanut lectin–T-antigen complex also provides a framework for exploring peanut lectin-based prognosis and diagnosis of certain types of carcinoma.

THOMSEN-FRIEDENREICH antigen, generally known as T-antigen (Galβ1–3GalNAc), is a chemically well-defined tumour-associated antigen of non-oncofetal origin with a well-documented link to malignancy in man. This structure is generally expressed as O-linked glycans (Galβ1–3GalNAc–αSer/Thr), prominently so in more than 85% of human carcinomas such as those in colon, breast, bladder, buccal cavity and prostate, as well as on poorly differentiated cells[1–5]. It is, however, cryptic or absent in normal cells. Among the proteins that recognize T-antigen, peanut (*Arachis hypogaea*) agglutinin (PNA) is most widely used. Unlike other anti-T probes such as amranthin and jacalin, it does not bind to the more abundant cryptic T- and Tn-antigens (GalNAcα-O-Ser/Thr) which are the sialylated derivatives of the T- and Tn-antigens respectively[6–8]. It is for this exclusive specificity of PNA and its consequent usefulness as a diagnostic tool, that its importance has not diminished even after the advent of other T-antigen specific proteins including monoclonal antibodies[9]. The specificity of PNA is exploited widely for monitoring the differential expression of T-antigen for both the prognosis and the diagnosis of malignancies[10–13]. PNA has also found extensive use in the early detection of T-polyagglutinability which has proven to be life saving in numerous instances. Ability of PNA to differentiate between immature thymocytes and their mature counterparts has found applications in bone-marrow transplantations[14].

PNA is a 110 kDa homotetrameric non-glycosylated legume lectin with an unusual quaternary structure[15]. We report here the 2.5 Å crystal structure of the lectin complexed with T-antigen, refined to an *R*-value of 17.5%. The complex provides a detailed description of the geometrical features of PNA–T-antigen recognition and illustrates, perhaps for the first time, how water molecules could be largely responsible for carbohydrate specificity[16]. It also provides a framework for designing better tools for monitoring the expression of T-antigen.

Methods

Experimental

The protein was prepared by affinity chromatography on cross-linked arabinogalactan[17]. Crystals of the complex were grown from a hanging drop of 5 mg/ml protein in 0.05 M sodium phosphate buffer, pH 7.0, containing 0.2 M sodium chloride, 0.02% sodium azide, 1.5 mM Galβ1–3GalNAc and 12% (w/v) PEG 8000, equilibrated against 40% (w/v) PEG 8000 in the same buffer. A 2.5 Å data set was collected from a single crystal (space group $P2_12_12$, $a = 129.892$ Å, $b = 126.676$ Å, $c = 76.516$ Å) on a Siemens-Nicolet area detector mounted on a GX20 Marconi Avionics Rotating anode X-ray generator. Data were processed using XENGEN[18] (Table 1).

Refinement

The structure of the PNA–lactose complex refined to 2.25 Å resolution was used as the starting model for refinement after removing water and sugar molecules from it. Refinement was initially carried out using 3 Å resolution data employing PROLSQ[19] in the CCP4[20] suite of programs. Data in the higher shells were introduced step by step. Restraints for non-crystallographic symmetry were used in all but the final cycles. A difference map calculated after a few cycles of refinement showed clear density for the T-antigen in all the four monomers. All maps were inspected using

Table 1. Crystallographic data and refinement statistics

Data collection statistics	
Resolution	2.5 Å
Number of observations	65934
Number of unique reflections	33952
Data completeness	82%
R-merge*	0.128
Refinement statistics	
Resolution range	10–2.5 Å
Number of protein atoms	6976
Number of carbohydrate atoms	104
Number of solvent atoms	517
Other ions	4 Mg^{2+}, 4 Ca^{2+}
Total number of atoms	7605
R-cryst ($F > 2\sigma F$)	0.175
R-free	0.251
Root-mean-square deviation from ideal geometry	
Bond lengths	0.009 Å
Bond angles	1.7°

*$R\text{-merge} = \Sigma \mid I - \langle I \rangle / \Sigma \langle I \rangle \mid$.

FRODO[21]. In the absence of the crystal structure, the initial model of the T-antigen was constructed using the 'carbohydrate builder' routine in the 'biopolymer' module in the Biosym/MSI package, interfaced with InsightII. After fitting the T-antigen, water molecules were added in steps using 2Fo-Fc and Fo-Fc maps. Further refinement was carried out using XPLOR[22]. Engh and Huber parameters were used throughout the refinement. Omit type maps were used to inspect the model[23,24]. The model converged to a final *R*-value of 0.175 and the corresponding *R*-free was 0.251. The relevant refinement statistics are presented in Table 1.

Reliability and geometrical parameters

The stereochemistry of the model was checked using PROCHECK[25]. More than 90% of the residues are in the most favourable region in the Ramachandran map[26] (Figure 1) with no residue in the disallowed region. The Luzzati plot[27] calculated at the end of the refinement showed the coordinate error to be around 0.2 Å. The displacement parameters of the main chain atoms have the maximum, the minimum and the average values of 68.7, 2.0 and 10.9 $Å^2$ respectively. The corresponding values for the side chain atoms are 73.0, 2.0 and 12.9 $Å^2$. Those for the sugar atoms and solvent atoms are 36.8, 2.0 and 17.6 $Å^2$ and 91.3, 2.0 and 25.7 $Å^2$ respectively. Hydrogen bonds and interactions involving water molecules were identified using the criteria employed in the case of the PNA–lactose complex[28].

The coordinates and structure factors have been deposited in the Brookhaven Protein Data Bank, with accession codes, 1TEP and R1TEPSF respectively.

Results and discussion

General features

The four subunits in the tetrameric complex have the same structure. The root-mean-square (r.m.s.) deviations between the α-carbon positions when pairs of sub-units are superposed, range between 0.15 and 0.18 Å The range for all atoms is 0.23 to 0.25 Å while that for the atoms in the disaccharide is 0.15 to 0.24 Å. The tertiary and quaternary structure of the lectin in the T-antigen complex is the same as that in its complex with lactose[28]. The same is true about the location and the geometry of the carbohydrate binding site (Figure 2). Each subunit, 236 amino acids long, has the typical legume lectin fold and is made up of a six-stranded flat β sheet, a seven-stranded curved sheet, a small five-stranded sheet, which has a major role in connecting the other two large sheets, and a number of loops with differing length and conformation. The carbohydrate-binding region is generated by residues in four loops, 91 to 106, 125 to 135, 75 to 83 and 211 to 216, at one edge of the subunit.

PNA–disaccharide interactions

The electron-density for the bound T-antigen is well-defined in all the four subunits and it permits an unambiguous description of the PNA–T-antigen interactions (Figure 3 and Table 2). The four hydrogen bonds found in other legume lectin–galactose complexes occur in the present complex as well. They are Asp 83 OD1–galactose O3, Gly 104 N–galactose O3, Asn 127 ND2–galactose O3 and Asp 83 OD2–galactose O4 (Figure 3 *b*). The other invariant features of legume lectin–carbohydrate interactions, namely, the stacking of an aromatic residue, Tyr 125 in PNA, against the galactose ring and the proximity of Ala 82, are also observed in the complex[16]. These invariant interactions, involving the first three of the four loops mentioned earlier, define the primary common requirements for the binding of the galactose moiety to a legume lectin. The specificity of individual lectins to different sugars is determined mainly by interactions involving the fourth loop (211–216) (refs 29–31).

Comparison with the lactose complex and the structural basis of specificity

The structural basis for the specificity of PNA for T-antigen is best described in relation to the protein–sugar interactions in the lactose complex (Figures 3 and 4). Both the sugars are disaccharides, with galactose as the first residue. The second hexapyranose is β1-4 linked

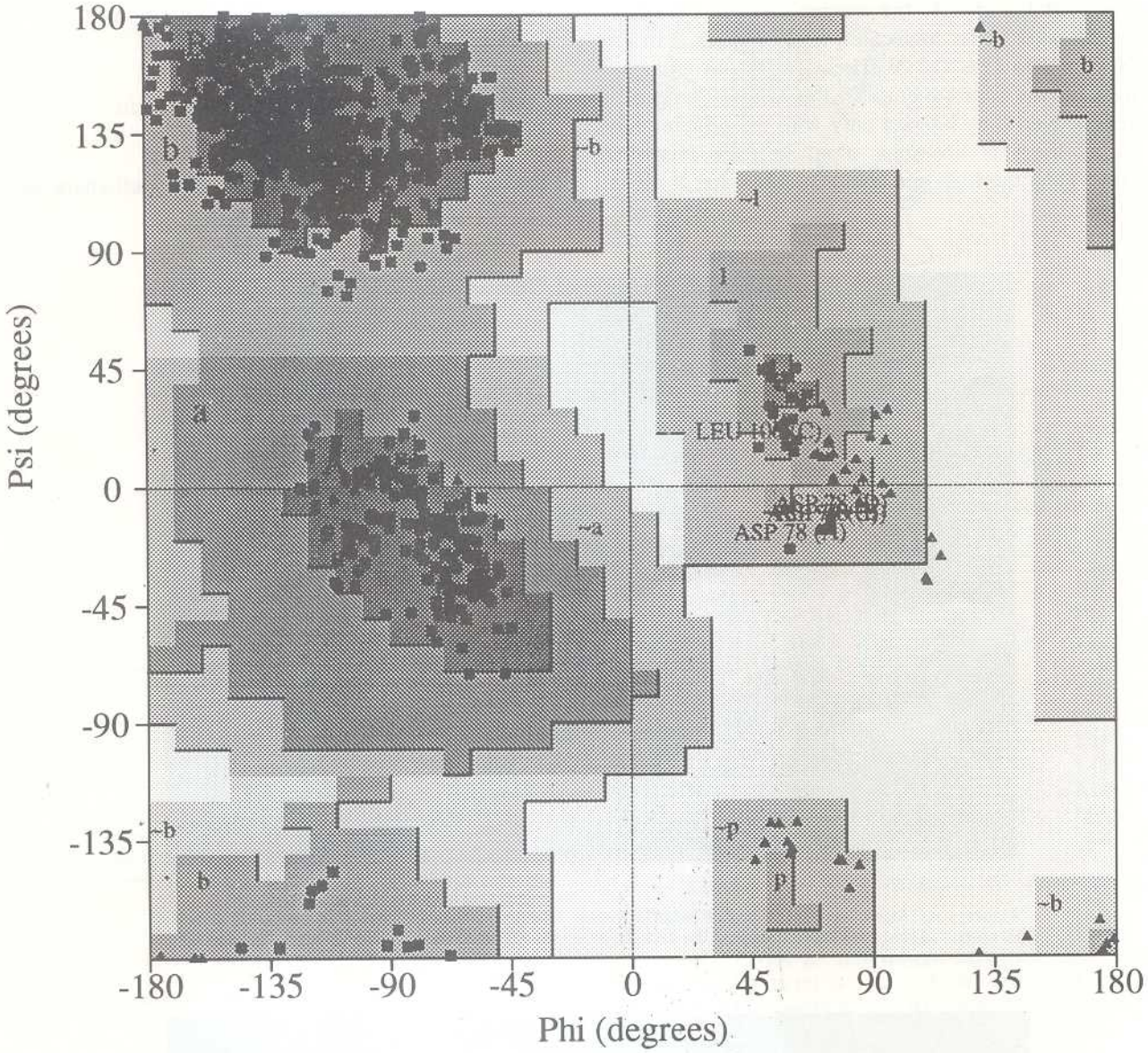

Figure 1. The Ramachandran map produced with PROCHECK. The squares and triangles represent non-glycine and glycine residues respectively.

glucose in the case of lactose while it is a β1-3 linked *N*-acetyl galactose in T-antigen. PNA, however, binds T-antigen 20 times more strongly than it binds lactose[4]. The invariant interactions, outlined earlier, exist in both the complexes. The other hydrogen bonds involving the galactose moiety (Ser 211 OG–galactose O4, Ser 211 OG–galactose O5, Asp 80 OD2–galactose O6) are also the same in the two complexes. In both the cases the galactose moiety is additionally linked to the protein through two water bridges (Gly 104 N–W2–galactose O2, Glu 129 OE2–W1–galactose O2). Thus the hydrogen bonded interactions involving galactose are the same in both the complexes. The same is true about the non-polar contacts. In the lactose complex, the second ring is connected to the lectin through two hydrogen bonds, both involving glucose O3, with Ser 211 OG and Gly 213 N. The differences in the linkage and the conformation of the two disaccharides are such that O4 of the second ring in the T-antigen occupies the same position as O3 of the second ring of lactose does, with respect to the lectin (Figure 4 *a*). Therefore in the T-antigen complex, O4 of *N*-acetylgalactosamine is hydrogen bonded to Ser 211 OG and Gly 213 N. The number and the nature of non-polar protein–sugar contacts are also nearly the same in the two complexes. Therefore, the direct interactions of PNA with the two disaccharides are essentially the same and they do not explain substantially the higher affinity of the lectin for the T-antigen. The only additional interactions T-antigen has with PNA are two water bridges, involving the

carbonyl oxygen of the acetamido group. One water molecule (W3) connects the oxygen atom to Ile 101 O, while the other water molecule (W4) connects it to Asn 41 ND2 and to Leu 212 N. These two water molecules exist in all the four subunits in the lactose complex as well. However, they interact only with the protein atoms in it, as there is no sugar atom in their immediate vicinity. Thus the high specificity of PNA for T-antigen and its 20-fold higher affinity for this saccharide than that for lactose, are generated essentially by these water bridges.

Comparison with other relevant lectin–carbohydrate complexes

The only other known structurally well-characterized

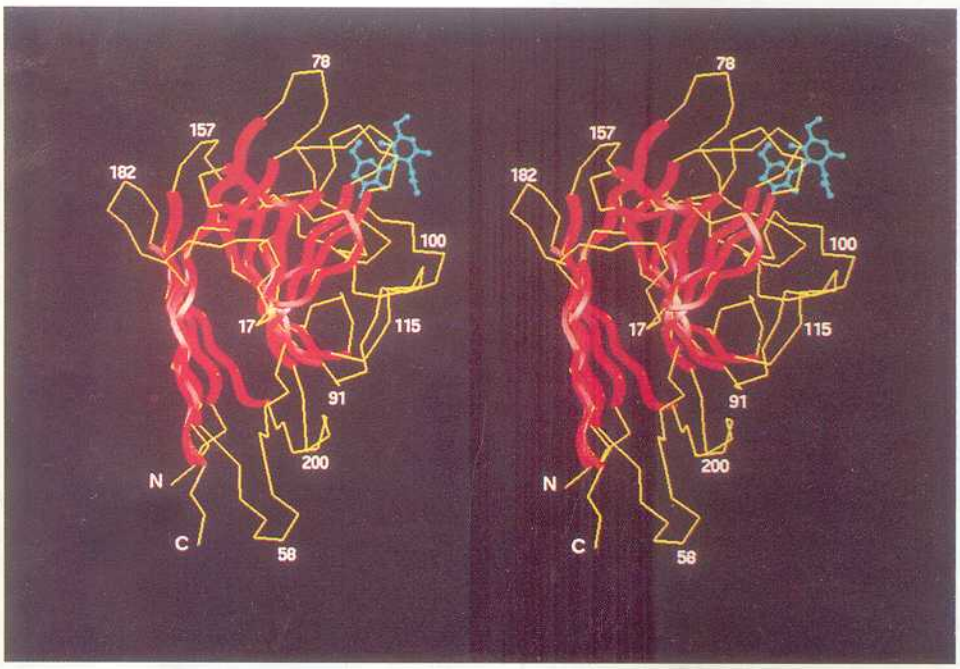

Figure 2 *a*. Stereo view of a peanut lectin monomer with the bound T-antigenic disaccharide shown in cyan, ball and stick representation. The beta sheets and loops are coloured red and yellow respectively (Insight, Biosym/MSI).

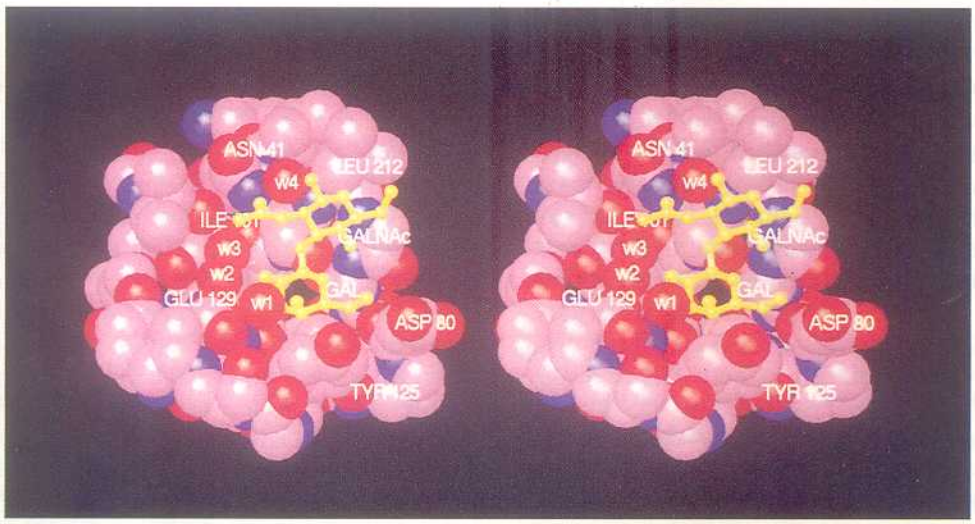

Figure 2 *b*. Solid van der Waals surface (oxygen in red, nitrogen in blue and carbon in pink) of the carbohydrate-binding region with the disaccharide shown in ball and stick (Insight, Biosym/MSI).

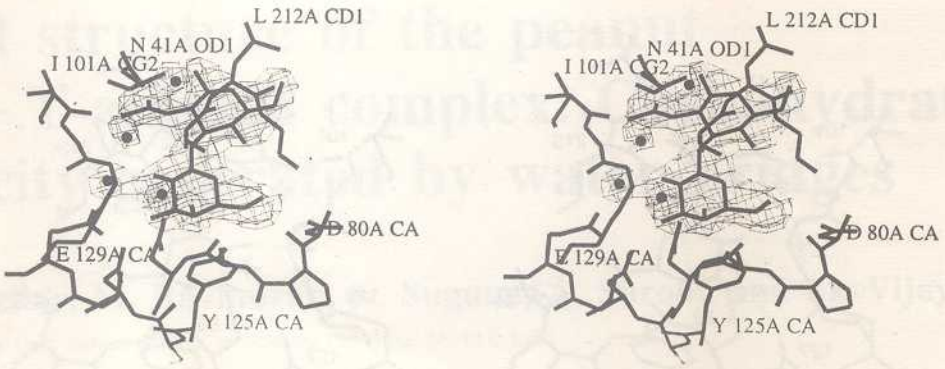

Figure 3 *a*. Omit map corresponding to the sugar molecule and the water molecules involved in protein–carbohydrate water bridges. The surrounding peptide stretches are also shown. The map is contoured at 3.2σ level.

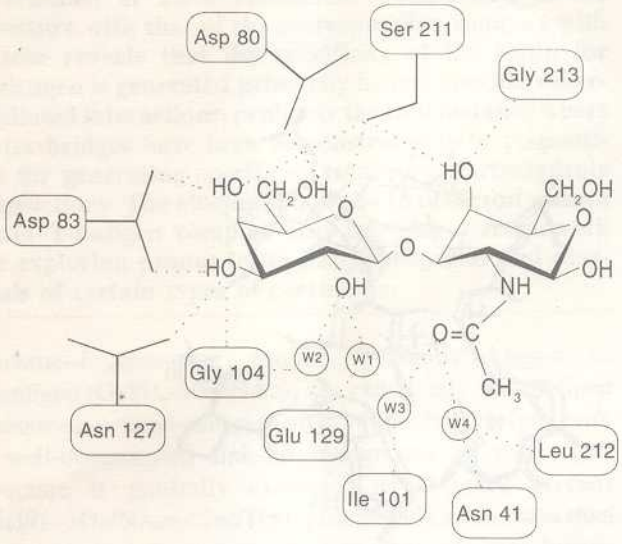

Figure 3 *b*. Schematic representation of protein–carbohydrate interactions in the complex.

protein complex with the T-antigenic disaccharide is that involving the heat-labile enterotoxin (LT) from *E. coli*[32]. In this complex, however, only the galactose moiety interacts with the protein. The *N*-acetyl-galactosamine moiety has virtually no interaction with LT. This is not surprising as LT has evolved to recognize the terminal sugars (galactose and sialic acid) of the GM1 pentasaccharide, and not T-antigen. Thus, the complex of LT with the T-antigenic disaccharide does not reflect the protein's natural binding propensity unlike in the case of the PNA complex of the disaccharide.

The crystal structure of jacalin[33], the second lectin to be demonstrated to have T-antigen specificity, has been determined recently in our laboratory, in its complex with methyl-α-galactose. This tetrameric protein with a novel lectin fold has an unusual carbohydrate binding site in which the specificity is generated by a post-

Table 2. Protein–sugar interactions

Protein atom	Sugar atom	Distance (Å) in Mon1	Mon2	Mon3	Mon4
Hydrogen Bonds					
GalO3	Asp83 OD1	2.80	2.61	2.67	2.66
	Gly104 N	3.22	2.98	2.88	2.91
	Asn127 ND2	2.76	2.97	2.78	2.96
GalO4	Asp83 OD2	2.71	2.75	2.70	2.87
	Ser211 OG	2.97	3.18	2.81	2.54
GalO5	Ser211 OG	3.02	3.00	2.83	2.95
GalO6	Asp80 OD2	2.83	2.94	3.43	3.19
GalNAcO3	Ser211 OG	3.15	3.17	3.12	3.20
GalNAcO4	Ser211 OG	3.33	2.98	3.08	2.89
	Gly213 N	3.29	3.06	3.14	2.81

Water mediated interactions (Distances (Å) averaged over four subunits)

GalO2- -W1- -Glu129 OE1	(O2- -W1 = 2.94; W1- -OE1 = 3.36)
GalO2- -W2- -Gly104 N	(O2- -W2 = 2.95; W2- -N = 2.97)
GalNAcO7- -W3- -Ile101 O	(O7- -W3 = 3.04; W3- -O = 2.94)
GalNAcO7- -W4- -Leu212 N	(O7- -W4 = 2.81; W4- -N = 2.79)
- -Asn41 ND2	(W4- -ND2 = 2.61)

Residues less than 4 Å from any sugar atom

Asp80, Ala82, Asp83, Gly103, Gly104, Tyr125, Asn127, Ser211, Leu212 and Gly213

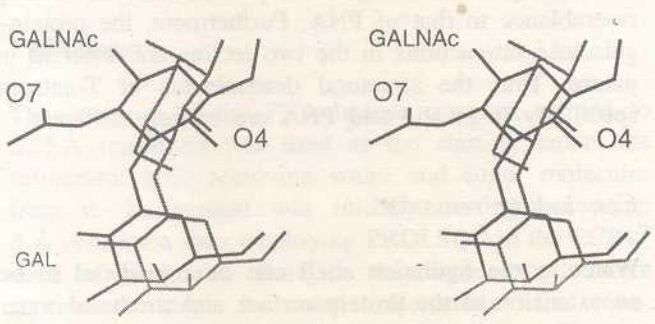

Figure 4 *a*. Superposition of the T-antigenic disaccharide and lactose (light lines) as they occur in the respective complexes. The lactose has been shifted slightly for clarity as the galactose rings superpose almost exactly.

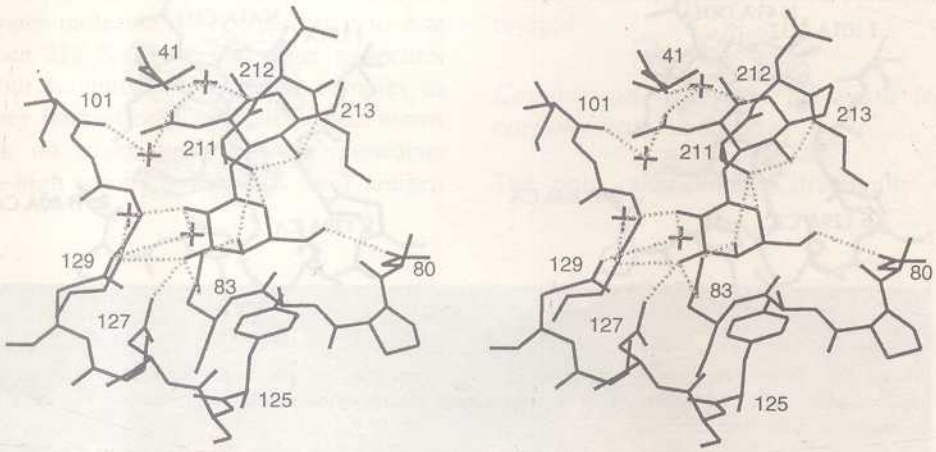

Figure 4 *b*. Stereo view of the protein–T-antigen interactions in PNA. The crosses represent water molecules and the broken lines hydrogen bonds.

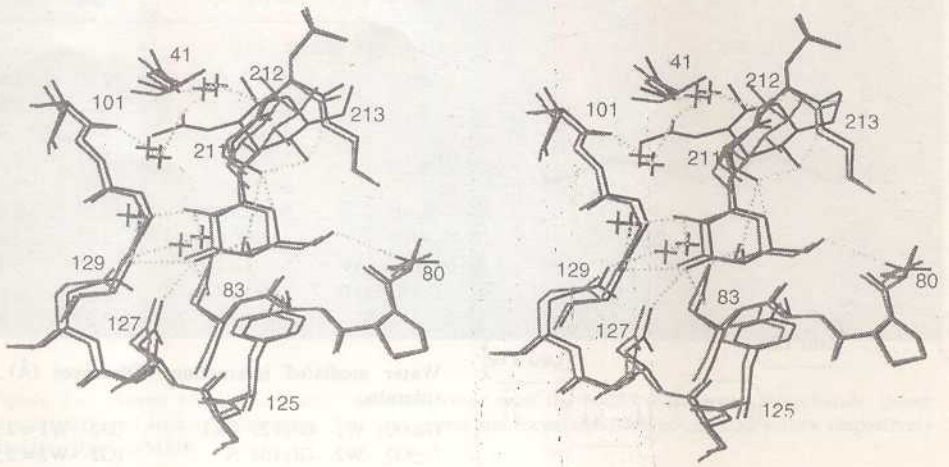

Figure 4 *c*. Stereo view of the superposition of the protein–sugar interactions in the T-antigen (dark lines) and the lactose (light lines) complexes. Only hydrogen bonds in the T-antigen complex are indicated.

translational modification. The structure of its complex with T-antigen has not yet been determined. However the carbohydrate-binding site of jacalin bears no striking resemblance to that of PNA. Furthermore, the protein–galactose interactions in the two lectins are different in nature. Thus the structural determinants of T-antigen specificity in jacalin and PNA are entirely different.

Concluding remarks

Water in the hydration shell can be considered to be an extension of the protein surface and structural water can effectively mediate carbohydrate–protein interactions by modifying the contours of the binding site. In addition to providing a detailed characterization of the PNA–T-antigen interactions, the crystal structure of the complex provides probably the first instance of carbohydrate specificity generated by water bridges. In addition to its intrinsic importance in relation to the growing body of detailed knowledge on protein–carbohydrate interactions, the information provided by the structure could be invaluable for the design of a PNA-based carcinoma marker.

1. Springer, G. F., *Science*, 1984, **224**, 1198–1206.
2. Irimura, T., Kawaguchi, T., Terao, T. and Osawa, T., *Carbohydr. Res.*, 1975, **39**, 317–327.
3. Lotan, R., Skutelsky, E., Danon, D. and Sharon, N., *J. Biol. Chem.*, 1975, **250**, 8518–8523.
4. Pereira, M. E. A., Kabat, E. A., Lotan, R. and Sharon, N., *Carbohydr. Res.*, 1976, **51**, 107–118.
5. Salunke, D. M., Swamy, M. J., Khan, M. I., Mande, S. C., Surolia, A. and Vijayan, M., *J. Biol. Chem.*, 1985, **250**, 13576–13579.

6. Rinderle, S. J., Goldstein, I. J., Matta, K. L. and Ratcliffe, R. M., *J. Biol. Chem.*, 1989, **264**, 16123–16131.
7. Swamy, M. J., Gupta, D., Mahanta, S. K. and Surolia, A., *Carbohydr. Res.*, 1991, **213**, 59–67.
8. Wu, A. M., *Mol. Cell. Biochem.*, 1984, **61**, 131–141.
9. Clansen, H., Stroud, M., Parker, J., Springer, C. and Hakomon, S., *Mol. Immunol.*, 1988, **25**, 199–204.
10. O'Keefe, D. and Ashman, L., *Clin. Exp. Immunol.*, 1982, **48**, 329–338.
11. Cling, C. K. and Rhodes, J. K., *Br. J. Cancer*, 1989, **59**, 949–953.
12. Zabel, P. L., Noujaim, A. A., Shysh, A. and Bray, J., *Eur. J. Nucl. Med.*, 1983, **8**, 250–254.
13. Zebda, N., Bailly, M., Brown, S., Dore, J. F. and Berthier-Vergnes, O., *J. Cell. Biochem.*, 1994, **54**, 161–173.
14. Reisner, Y., Biniaminov, M., Rosenthal, E., Sharon, N. and Ramot, B., *Proc. Natl. Acad. Sci. USA*, 1979, **76**, 447–451.
15. Banerjee, R. *et al.*, *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 227–231.
16. Ravishankar, R.; Ravindran, M., Suguna,K., Surolia, A. and Vijayan, M., *Prog. Biophys. Mol. Biol.*, 1996, **65**, Suppl 1, 33.
17. Majumdar, T. and Surolia, A., *Prep. Biochem.*, 1978, **8**, 119–131.
18. Howard, A. J. *et al.*, *J. Appl. Crystallogr.*, 1987, **20**, 383–387.
19. Hendrickson, W. A. and Konnert, J. H., *Computing in Crystallography* (eds Diamond, R., Ramaseshan, S. and Venkatesan, K.), Indian Academy of Sciences, Bangalore, 1980, pp. 13.01–13.23.
20. Collaborative Computational Project, Num. 4 *Acta Crystallogr.*, 1994, **D50**, 760–763.
21. Jones, T. A., *J. Appl. Crystallogr.*, 1978, **11**, 268–272.
22. Brunger, A. T., *X-PLOR Version 3.1 Manual*, Yale University, 1992.
23. Vijayan, M., *Computing in Crystallography* (eds Diamond, R., Ramaseshan, S. and Venkatesan, K.), Indian Academy of Sciences, Bangalore, 1980, pp. 19.01–19.26.
24. Bhat, T. N. and Cohen, G. H., *J. Appl. Crystallogr.*, 1984, **17**, 244–248.
25. Laskowski, R. A., Mac Arthur, M. W., Moss, D. S. and Thornton, J. M., *J. Appl. Crystallogr.*, 1993, **26**, 283–291.
26. Ramachandran, G. N. and Sasisekharan, V., *Adv. Protein Chem.*, 1968, **23**, 283–438.
27. Luzzatti, V., *Acta Crystallogr.*, 1952, **5**, 802–810.
28. Banerjee, R. *et al.*, *J. Mol. Biol.*, 1996, **259**, 281–296.
29. Young, N. M. and Oomen, R. P., *J. Mol. Biol.*, 1992, **228**, 924–934.
30. Sharon, N., *Trends Biochem. Sci.*, 1993, **18**, 221–226.
31. Weis, W. I. and Drickamer, K., *Annu. Rev. Biochem.*, 1996, **65**, 441–475.
32. Akker, F. V. D., Steensma, E. and Hol, W. G., *J. Protein Sci.*, 1996, **5**, 1184–1188.
33. Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A. and Vijayan, M., *Nature. Struct. Biol.*, 1996, **3**, 596–603.

ACKNOWLEDGEMENTS. The diffraction data were collected on the Area Detector Facility supported by the Department of Science and Technology (DST) and the Department of Biotechnology (DBT). Facilities at the Supercomputer Education and Research Centre and the Interactive Graphics Based Molecular Modelling Facility and Distributed Information Centre (both supported by DBT) were used in the work. The work was funded by the DST.

Received 5 May 1997; accepted 14 May 1997