Article No. jmbi.1999.2910 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 290, 867–879

JMB

The Structure of a Tunicate C-type Lectin from *Polyandrocarpa misakiensis* Complexed with D-Galactose

Sébastien F. Poget[1]*, Glen B. Legge[1], Mark R. Proctor[1], P. Jonathan G. Butler[2], Mark Bycroft[1] and Roger L. Williams[2]

[1]*Cambridge Centre for Protein Engineering, Department of Chemistry, University of Cambridge, Lensfield Road Cambridge, CB2 1EW, UK*

[2]*Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK*

**Corresponding author*

C-type lectins are calcium-dependent carbohydrate-recognising proteins. Isothermal titration calorimetry of the C-type *Polyandrocarpa* lectin (TC14) from the tunicate *Polyandrocarpa misakiensis* revealed the presence of a single calcium atom per monomer with a dissociation constant of 2.6 μM, and confirmed the specificity of TC14 for D-galactose and related monosaccharides. We have determined the 2.2 Å X-ray crystal structure of *Polyandrocarpa* lectin complexed with D-galactose. Analytical ultracentrifugation revealed that TC14 behaves as a dimer in solution. This is reflected by the presence of two molecules in the asymmetric unit with the dimeric interface formed by antiparallel pairing of the two N-terminal β-strands and hydrophobic interactions. TC14 adopts a typical C-type lectin fold with differences in structure from other C-type lectins mainly in the diverse loop regions and in the second α-helix, which is involved in the formation of the dimeric interface. The D-galactose is bound through coordination of the 3 and 4-hydroxyl oxygen atoms with a bound calcium atom. Additional hydrogen bonds are formed directly between serine, aspartate and glutamate side-chains of the protein and the sugar 3 and 4-hydroxyl groups. Comparison of the galactose binding by TC14 with the mannose binding by rat mannose-binding protein reveals how monosaccharide specificity is achieved in this lectin. A tryptophan side-chain close to the binding site and the distribution of hydrogen-bond acceptors and donors around the 3 and 4-hydroxyl groups of the sugar are essential determinants of specificity. These elements are, however, arranged in a very different way than in an engineered galactose-specific mutant of MBPA. Possible biological functions can more easily be understood from the fact that TC14 is a dimer under physiological conditions.

Keywords: invertebrate C-type lectin; carbohydrate recognition; galactose specificity; calcium-binding affinity; dimeric interface

Introduction

Molecular recognition in biological systems is often based on interactions between oligosaccharide structures and corresponding receptor proteins. The C-type lectins are a family of extracellular carbohydrate recognition proteins characterised by a common sequence motif of 115 to 130 amino acid residues. This domain, called the carbohydrate recognition domain (CRD), usually shows specific, but weak (with dissociation constants in the mM range) calcium-dependent binding to a variety of monosaccharides (Drickamer, 1993). C-type lectin CRDs are found as building blocks in a variety of

Present address: G. B. Legge, Department of Molecular Biology MB-2, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA.

Abbreviations used: CRD, carbohydrate recognition domain; ESEL, human E-selectin; HLIT, human lithostathine; HTNT, human tetranectin; ITC, isothermal titration microcalorimetry; MBPA, rat serum mannose-binding protein; MBPC, rat liver mannose-binding protein; TC14, *Polyandrocarpa* lectin (tunicate-derived C-type lectin of 14 kDa); 6-GPGP, 6-β-D-galactopyranosyl-D-galactopyranose; 4-GPGP, 4-α-D-galactopyranosyl-D-galactopyranose; DTT, dithiothreitol.

E-mail address of the corresponding author: sfp22@cam.ac.uk

0022-2836/99/290867–13 $30.00/0

multi-domain proteins involved in organising the extracellular matrix, in endocytosis, in the primary immune system and in interactions of blood cells (Gabius, 1997). In the current Pfam database of protein families (Sonnhammer *et al*., 1998), 389 C-type lectin sequences have been identified in a wide range of animals (in nematodes, molluscs, arthropods, echinoderms, tunicates and in a large number of vertebrates). The importance of the C-type lectins is also reflected by the fact that they represent the seventh most common protein domain identified in the *Caenorhabditis elegans* genome (The *C. elegans* Sequencing Consortium, 1998). C-type lectin domains adopt a typical fold: one half of the molecule consists of a long two-stranded β-sheet and two α-helices, while the second half contains the calcium and carbohydrate-binding site(s) and is mostly formed of non-repetitive loop structures. This fold is conserved in all the examples of known C-type lectin structures, including the human and rat serum mannose-binding proteins (MBPA; Sheriff *et al*., 1994; Weis *et al*., 1991), rat liver mannose-binding protein (MBPC; Ng *et al*., 1996), human E-selectin (ESEL; Graves *et al*., 1994) and human tetranectin (HTNT; Kastrup *et al*., 1998; Nielsen *et al*., 1997). The structure of human lithostathine (HLIT; Bertrand *et al*., 1996) also adopts the typical C-type lectin fold, but does not have any sugar binding activity. Another C-type lectin homologue without lectin activity is the coagulation factors-IX/X binding protein from snake venom, which shows a similar fold apart from a loop that projects into the other subunit of the heterodimer (Mizuno *et al*., 1997). The angiogenesis inhibitor endostatin has been found to have a fold distantly related to C-type lectins in spite of almost no sequence identity (Hohenester *et al*., 1998), and the NMR structure of the link module shows that this protein adopts the same basic fold containing all the elements of regular secondary structure, but the whole loop region is replaced by one short turn (Kohda *et al*., 1996).

Insight into the mechanism of carbohydrate binding in C-type lectins was first obtained from the crystal structure of MBPA complexed with an oligomannose asparaginyl-oligosaccharide (Weis *et al*., 1992). Later, the structures of MBPC complexed with different monosaccharides and a mutant of MBPA that bound galactose were solved by X-ray crystallography (Ng *et al*., 1996; Kolatkar & Weis, 1996). The binding site of the C-type lectins is quite exposed and is located on the surface of the loop region, with the 3 and 4-hydroxyl groups of the carbohydrate coordinating to a bound calcium ion. Additional hydrogen bonds are formed between the sugar and the protein side-chains involved in binding this calcium ion, with van der Waals contacts further stabilising the bound sugar. In all structures of galactose-specific lectins, including the galactose-binding mutant of MBPA, additional hydrophobic stacking between an aromatic protein side-chain and the apolar side of the galactose was observed (Rini, 1995).

Lectins often bind to natural polysaccharides with much higher affinity than to simple monosaccharides. This can be achieved by two methods: In the first case, increased binding affinity can be achieved by extended binding sites, where the terminal sugar is bound to the lectin, and further residues of the oligosaccharide make contact to the protein in either a direct or solvent-mediated way. An even more efficient way to increase the binding affinity to an extended oligosaccharide is the multiple binding by a cluster of several identical binding sites, often achieved through protein oligomerisation (Weis & Drickamer, 1996).

Here, we solved the X-ray crystal structure of *Polyandrocarpa* lectin (TC14, tunicate derived C-type lectin of 14 kDa) complexed with D-galactose. TC14 was isolated from the tunicate *Polyandrocarpa misakiensis* and its amino acid sequence was determined (Suzuki *et al*., 1990). Cloning of the cDNA encoding TC14 confirmed the original sequence (Shimada *et al*., 1995). It has D-galactose specificity and is suggested to play a role in generalised defence mechanisms of the organism because of its strong antibacterial activity (Suzuki *et al*., 1990). It was shown to be part of the extracellular matrix in developing buds, directing undifferentiated haemoblasts and pluripotent stem cells to the epithelium of the bud vesicle (Kawamura *et al*., 1991).

In spite of the fact that the structures of seven C-type lectins are known, the only natural protein-ligand complexes studied are those of the mannose-binding proteins with mannose carbohydrates. Some information about how galactose selectivity can be achieved has been obtained from structural studies on a galactose-specific mutant of MBPA (Kolatkar & Weis, 1996). The structure of TC14 complexed with D-galactose reveals a novel way in which galactose specificity can be achieved in the C-type lectins. TC14 is also the first non-vertebrate C-type lectin whose structure has been solved, and therefore an interesting model for analysing the evolution of carbohydrate recognition and the C-type lectin fold.

Results

Protein expression and purification

Recombinant TC14 was overexpressed in *Escherichia coli* and obtained as inclusion bodies. The protein was solubilised in 8 M urea and purified in the denatured state on a DEAE Sepharose ion exchange column (Pharmacia) using a NaCl gradient from 0 to 800 mM. Refolding was achieved by dialysing in three steps, removing the urea and increasing the calcium concentration to a final value of 20 mM. A second purification step was performed using an immobilised galactose resin (Pierce Chemical Co.). The protein was loaded on the column in the last refolding buffer and eluted with EDTA.

Analytical ulracentrifugation

Analytical ultracentrifugation was employed to investigate unambiguously whether the protein is present as a monomer or dimer in solution. Sedimentation equilibrium runs were performed at different initial protein concentrations (21, 13 and 5 μM). The resulting plots of the weight average apparent molecular mass against concentration are shown in Figure 1(a), and show essentially a dimer, with no evidence of dissociation even at the lowest protein concentrations. There is some slight, non-equilibrating aggregation, but this is <5% of the total protein and this small amount is unlikely to have any biological relevance. An alternative method of analysis was also applied, where a model involving homodimerisation was used to fit the absorbance directly at the various radii. This model gave good fits to the sedimentation data in each cell, out to 7.0 cm radius (or 6.9 cm for the highest concentration cell), when a monomer mass ~15% higher than the true mass was used (i.e. the assumed partial specific volume is probably not accurate), and the dissociation constants (K_d) fitted for each cell were <0.1 μM. Plots of the residuals for the fit against radius are shown in Figure 1(b). This result is again compatible with a tightly aggregated dimer at physiological protein concentrations, with traces of higher-order aggregates disturbing the fit at the bottom of each cell.

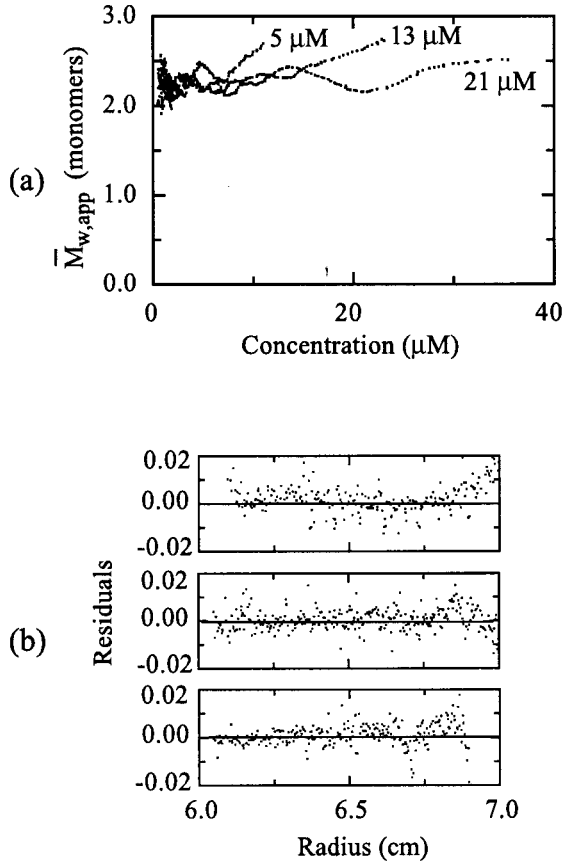

Figure 1. Sedimentation equilibrium analysis of aggregation state of TC14. (a) Plot of the mass apparent average molecular mass (as monomers of the sequence derived molecular mass) against concentration for samples of TC14 analysed at 27°C. Data from samples at initial concentrations of 5, 13 and 21 μM are plotted together, to show the extent of overlap. The three data sets are labelled with their initial concentration. (b) Residual errors between the measured optical density and that calculated from fitting to a model for the formation of homodimers, optimising the dissociation constant, baseline error and monomer molecular mass (to allow for error in the assumed partial specific volume), for each sample. The relatively even distributions around the zero lines show that the model gives a good fit to the data over the radii used for the fitting.

It can therefore be concluded that TC14 is a dimer under physiological conditions. Gel filtration analysis of native (Suzuki *et al*., 1990) and recombinant TC14 had suggested a molecular mass of 15 and 12 kDa, respectively. We would suggest that either the shape of the dimer caused it to behave anomalously in gel filtration or that the protein was retained somewhat by partial binding of the lectin to the polysaccharide-based resin.

Binding studies

Both the calcium binding and the affinity for different types of mono- and disaccharides were analysed by isothermal titration microcalorimetry (ITC). Typical titration curves for calcium and galactose binding are shown in Figure 2. The results for the calcium binding show that TC14 has a single binding site per CRD domain (the average estimate for the number of calcium sites from three experiments is 1.05(±0.11)) with a dissociation constant of 2.6(±0.2) μM for the protein-galactose complex. This is in the normal range observed for different calcium-binding proteins (e.g. 8.9 μM for calmodulin (Milos *et al*., 1986) or 0.6 μM for oncomodulin (Cox *et al*., 1990)). Binding studies with different mono- and disaccharides were also performed, and the results are summarised in Table 1.

The binding affinity of TC14 for monosaccharides is weak, which is consistent with that observed for most of the C-type lectin family (the complex of the galactose-specific mutant of MBPA with β-methyl galactoside has a K_d of 2.0 mM, and rat hepatic lectin 1 with the same ligand has a K_d of 1 mM as determined by NMR binding assay; Iobst & Drickamer, 1994). Recombinant TC14 shows the expected selectivity for a galactose-specific C-type lectin, and the results were also in agreement with the binding studies performed on the native protein (Suzuki *et al*., 1990). The highest affinities can be observed for D-galactose and D-fucose, with no detectable binding for D-mannose or D-glucose and only very weak binding for *N*-acetylgalactosamine. A binding affinity in the same range as for D-galactose was also detected for the two disaccharides 6-β-D-galactopyranosyl-D-galactopyranose (6-GPGP) and 4-α-D-galactopyranosyl-D-galactopyranose (4-GPGP). The slightly higher affinity for 6-GPGP is probably due to the fact that

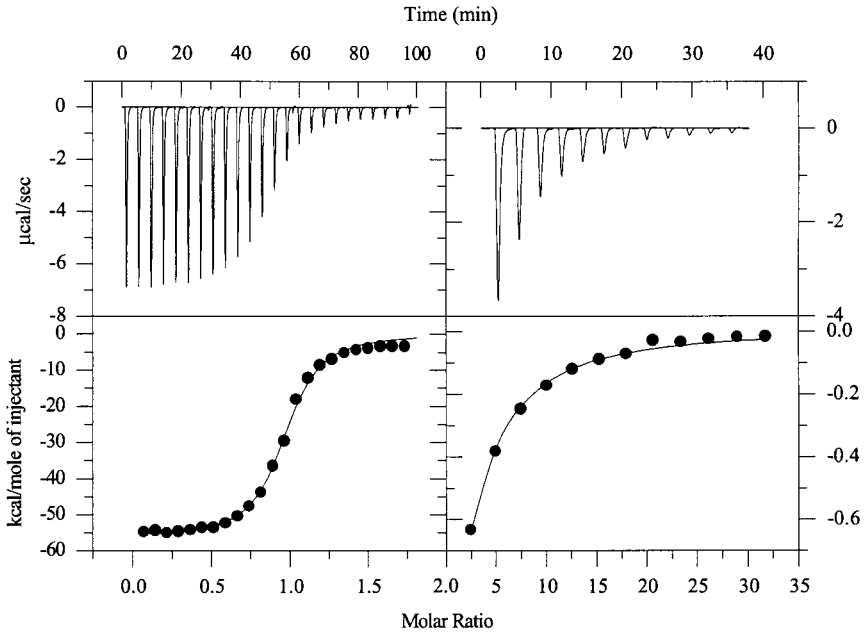

Figure 2. ITC scans of TC14 binding to calcium (left), and TC14 binding to D-galactose (right). The upper panel shows the incremental heat liberation upon each injection and the lower panel shows the integrated area of the above peaks following the subtraction of the heat of dilution plotted against the molar ratio of ligand to TC14 in the reaction cell.

the 3 and 4-hydroxyl groups of both galactose moieties have access to the TC14 binding site, whereas in 4-GPGP one of the galactose moieties becomes sterically hindered due to the α-1,4 glycosidic linkage.

Structure solution and quality of the model

The protein was crystallised in the presence of calcium and D-galactose. Using a single samarium derivative, a 2.0 Å resolution model was built. The SIRAS map clearly showed that glycerol, present in large excess in the cryoprotectant solution used for diffraction data collection, was occupying the carbohydrate-binding site in the crystal. Therefore, a second dataset was collected to a resolution of 2.2 Å, using galactose as the cryoprotectant instead of glycerol. The galactose was clearly present in the $2mF_o - DF_c$ electron density map. Crystals frozen in galactose and glycerol solutions were isomorphous, and superposition of the coordinates of the glycerol-containing and the galactose-containing structures revealed that the rmsd over all C^α atoms is 0.18 Å with maximum deviations between protein atoms of less than 0.3 Å. Even the solvent spheres in both forms are very similar. The stereochemical parameters for both final models are summarised in Table 2. The asymmetric unit consists of two protein molecules labelled A and B. Molecules A and B can be superimposed with an rmsd over all C^α-atoms of 0.26 Å for the galactose and 0.23 Å for the glycerol form.

According to PROCHECK (Laskowski *et al.*, 1993), 93.0 % of the residues in the galactose form lie in the most favourable areas of the Ramachandran plot (90.2 % of the residues for the glycerol form), and no residues are found in disallowed areas in either of the two models.

The final model for the galactose form contains residues 2-124. No electron density could be found for Asp125, and the density for Asp124 is weak and non-continuous, indicating that the C terminus of the protein is partially disordered. Only partial electron density could be found for the surface-exposed side-chains of residues Lys40, Ala64 and Asp124 in molecule A and Gln22 and Asp124 in molecule B. Even though electron density for the first residue Met1 is evident, no model with acceptable geometry could be fitted to the map. Another interesting feature of the N-terminal region is a

Table 1. Thermodynamic parameters of the binding of mono- and disaccharides to TC14

Sugar ligand	K_d (mM)	ΔH (kJ mol^{-1})	ΔG (kJ mol^{-1})	ΔS (J mol^{-1}K^{-1})
D-Fucose	0.33	−25.5	−19.8	−19.9
D-Galactose	0.44	−77.8	−19.1	−197.0
GalNAc	2.0	−15.1	−15.5	1.4
6-GPGP	0.38	−46.8	−19.5	−91.6
4-GPGP	0.58	−41.1	−18.4	−76.1

GalNAc is *N*-acetylgalactosamine; 6-GPGP is 6-β-D-galactopyranosyl-D-galactopyranose; and 4-GPGP is 4-α-D-galactopyranosyl-D-galactopyranose. Measurement of the dissociation constants for D-mannose, D-glucose, L-fucose and *N*-acetylglucosamine was also attempted, but their binding affinities were too low to be detected by the instrument.

Table 2. Details and stereochemical parameters of the final models

	Glycerol form	Galactose form
Number of protein atoms	1927	1928
Number of solvent atoms	278	243
Metal atoms	4 Ca + 7 Zn	4 Ca + 5 Zn
Hetero-groups	2 glycerol + 4 acetate	2 galactose
rmsd for bond lengths (Å)	0.006	0.007
rmsd for bond angles (deg.)	1.6	1.7
Average *B*-factor over all protein atoms	20.0	25.5
G-factor [a]	0.15	0.10

[a] As determined with PROCHECK (Laskowski *et al.*, 1993).

chain of five adjacent spheres with high electron density. The middle sphere lies between the side-chains of Asp2 of one molecule in the asymmetric unit and a symmetry-related copy of the other molecule in the asymmetric unit. This region could not be clearly interpreted at 2.0 Å resolution but was modelled as three zinc atoms bridged by two water molecules. These, together with another zinc site next to Asp35 and a fully hydrated calcium site near Trp82, are believed to be due to crystal packing and the solution used for crystal growth. This assumption is supported by the ITC results, which only showed one calcium atom bound per monomer.

Structure description

As expected by sequence similarity, TC14 adopts a typical C-type lectin fold as shown in Figure 3. The elements of secondary structure are summarised in Figure 4. The structure can be roughly divided into two parts: a lower part of the molecule containing mostly elements of regular secondary structure (the two α-helices and the two long β-strands β1 and β5), and an upper part that consists of mostly non-repetitive secondary structure and includes the carbohydrate recognition site. The short β-strand β2 is at the boundary between the two parts of the molecule. This strand also acts as a connection between the two β-sheets formed by the N and C-terminal strands β1 and β5 in the lower part, and the β3 and β4 strands in the upper part of the molecule. The only other area of regular secondary structure in the upper part of the molecule is a very small β-sheet formed by β′ and β″. The extended loops 3 and 4 are involved in galactose binding together with strand β4. TC14 shows the typical two disulphide bridges for short-form C-type lectins, joining α1 to β5 and closing the β-sheet between strands 3 and 4.

The dimer detected in solution is manifested in the crystal by the two molecules in the asymmetric unit which form a symmetric dimer (Figure 5). It is stabilised by both intermolecular hydrogen bonds and hydrophobic interactions. Firstly, a four-stranded β-sheet extending through both molecules of the dimeric unit is formed through antiparallel pairing of the two N-terminal β-strands. The second element of the dimeric interface is a close contact between the second α-helices of each molecule, which forms a small hydrophobic core consisting of residues Phe45 and Phe7 from both molecules (Figure 6). The total buried surface area between the two molecules is 1727 Å^2 as calculated with the programme GRASP (Nicholls, 1992), using a probe radius of 1.4 Å and the program's default atom radii. This value is within the range found for proteins forming stable dimers in solutions.

Ligand binding

The sugar binding in TC14 occurs predominantly indirectly through a Ca^{2+} (see Figures 7 and 8(a)). ITC data indicate that there is only one calcium-binding site per domain, and in the TC14 structure, the sugar-binding calcium site is likely to be the only genuine metal site. This site corresponds to the samarium site in the samarium derivative. The protein ligands for this calcium ion are side-chain oxygen atoms of Glu86, Asn89, Asp107 and Asp108 as well as the main-chain carbonyl oxygen of Asp108. The last two sites to reach pentagonal bipyramidal coordination are occupied by the galactose 3 and 4-hydroxyl oxygen atoms. Hydrogen bonds between protein side-chains and the galactose hydroxyl groups are formed between hydroxyl group 3 and Glu86 and Ser88, as well as between hydroxyl group 4 and Asp107. Solvent molecules play a role in the structure of the complex by mediating indirect protein-carbohydrate interactions that are bridged by a single water atom. Such interactions can be found between the 2-hydroxyl group and Ser88 and between the 5-hydroxyl group and Asn89. The galactose 6-hydroxyl group is linked to the side-chains of the three residues Asp52, Gln98 and Arg115 through a single water molecule. Additional stabilisation of the complex is achieved through hydrophobic stacking of the side-chain of Trp100 to the apolar side of D-galactose. The importance of this hydrophobic interaction is also reflected in the slightly higher binding affinity of TC14 for D-fucose than for D-galactose. D-Fucose

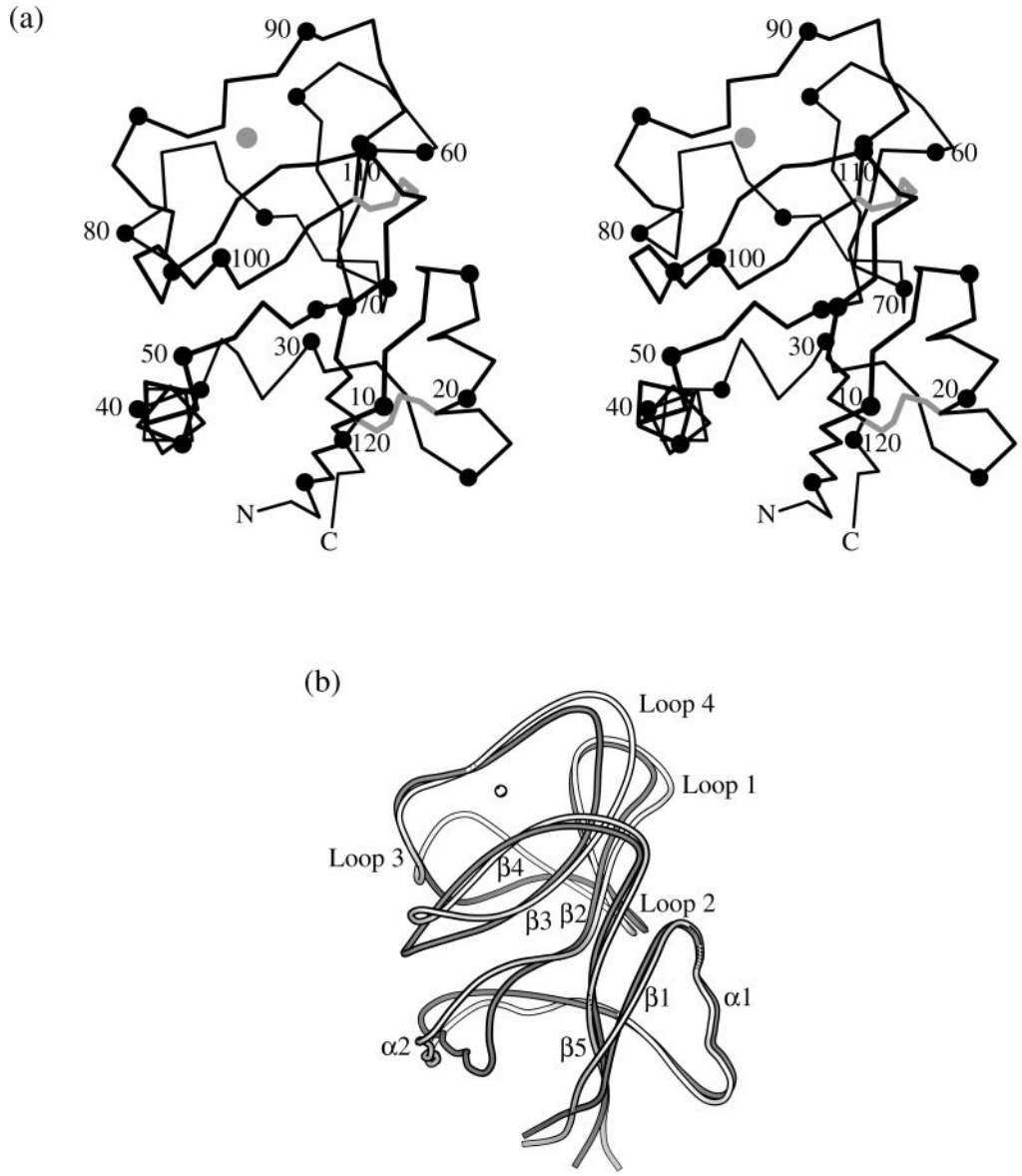

Figure 3. (a) Stereoview of the backbone of TC14, starting from the first residue visible in the map, Asp2. Every fifth residue is represented with a black sphere, every tenth residue is numbered. The galactose-binding calcium ion is represented as a grey sphere and the disulphide bridges are shown as grey bonds. (b) Worm representation of a superimposition of TC14 (in white) on MBPA (in grey). The major elements of secondary structure are labelled as in TC14. Both Figures were produced with MOLSCRIPT (Kraulis, 1991).

only differs from D-galactose by the absence of a hydroxyl group in the C6 position. The less polar sugar can therefore form a stronger hydrophobic interaction with the side-chain of Trp100.

Discussion

Homology to other C-type lectins

Superposition of the structure of TC14 on the structures of the C-type lectins shown in the structural sequence alignment (Figure 4) shows a C^{α} rmsd of between 1.36 (MBPA, see Figure 3(b)) and 1.78 Å (ESEL). This demonstrates that in spite of the sequence divergence (sequence identity between 29% for ESEL and 17% for HTNT) of TC14 relative to the mammalian lectins, the overall structure is highly conserved. Generally, the structural diversity between the different C-type lectins is higher in the loop regions, mainly caused by insertions or deletions of amino acids, but the structure of the first part of loop 3 in TC14 shows a significant divergence (deviations of more than 5 Å) from all other C-type lectins. The second element of structure where TC14 greatly differs from other C-type lectins is helix α2. Its axis is turned by about 20° relative to all other structures. This helix is important in the formation of the dimeric interface, and as TC14 is the only C-type lectin among the analysed structures that forms a dimer of physiological importance, we can assume

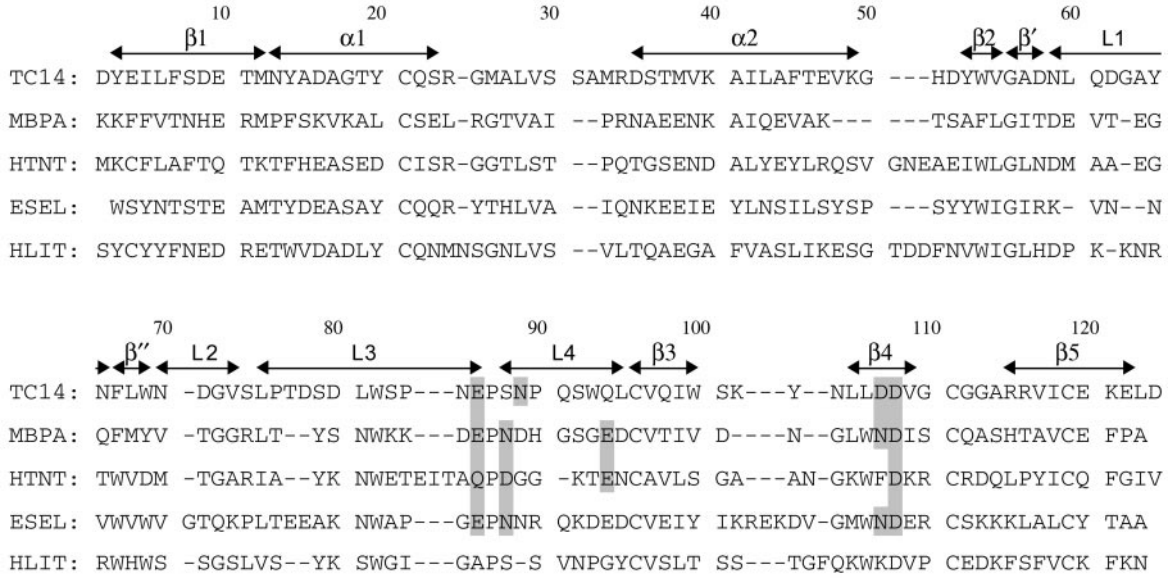

Figure 4. Structural sequence alignment of TC14 with rat mannose-binding protein A (MBPA), human E-selectin (ESEL), human tetranectin (HTNT) and human lithostathine (HLIT). The alignment has been made manually considering structural similarities. The elements of secondary structure shown refer to TC14. Residues of the carbohydrate binding site (or putative carbohydrate binding site for ESEL and HTNT) are shaded in grey.

that this conformation of the helix is required for stabilisation of the dimer.

The structure of TC14 is the first example of a naturally dimeric C-type lectin. It is rather unlikely that this type of dimerisation occurs in C-type lectin domains that are part of multi-domain proteins, as both N and C termini are located close to the dimeric interface and therefore neighbouring domains would probably disturb the dimer formation. This organisation might be typical for invertebrate isolated C-type lectin domains. One biological function of TC14 is the direction of differentiating cells in the developing bud of *P. misakiensis*. The ability of TC14 to dimerise might allow it to perform this function by binding to both the extracellular matrix and to the surface of the migrating cells.

Galactose binding

The D-galactose selectivity of TC14 observed in the binding studies can be explained mainly by two aspects of the structure: the distribution of hydrogen-bond acceptors and donors around the sugar hydroxyl groups 3 and 4 and the presence of a tryptophan side-chain close to the binding site.

In TC14, the sugar hydroxyl groups 3 and 4 are hydrogen-bonded to two acidic side-chains on one side of the sugar ring and to one serine residue and one water molecule on the other side of the ring (see Figure 8(a)). As it is very unlikely that acidic side-chains in the calcium ligand sphere are protonated at a pH over 6.5, it is assumed that they act as hydrogen acceptors in TC14, and therefore Ser88 and the water molecule have to act as proton donors. The importance of the hydrogen donor/acceptor distribution for sugar specificity has been demonstrated earlier in MBPA, where the two mutations Glu185 → Gln and Asn187 → Asp are enough to induce weak preferential galactose binding (Drickamer, 1992). These mutations result in the hydrogen donors (amide side-chains) being on one side of the sugar ring and the hydrogen acceptors (acidic side-chains) being on the other side, whereas in wild-type MBPA one donor and one acceptor are found on each side (see Figure 8(b) and (c)). Comparison of the binding site of TC14 and the galactose specific QPDWG mutant of MBPA (Kolatkar & Weis, 1996) shows that the hydrogen donor/acceptor distribution is inverted, and as a consequence the orientation of the galactose ring in the binding site is flipped by 180 degrees in the two proteins (compare Figure 8(a) and (b)). This observed relation between hydrogen donor/acceptor distribution and sugar specificity/orientation allows the lowest energy non-covalent bonding partner distribution around the sugar hydroxyl groups to be achieved. Newman projections along the oxygen-carbon bonds of the 3 and 4-hydroxyl groups of all sugar-C-type lectin complexes studied (see Figure 9) reveal that the hydroxyl-proton (and therefore the hydrogen-bond acceptor) is never *gauche* to both the ring carbon atoms and therefore always points away from the ring. This behaviour is consistent with an observation made by Elgavish & Shaanan (1997) for the geometry around the 4-hydroxyl group for all lectin-sugar complexes known so far. Hydrogen donors cluster in the region between *gauche*$^-$ and *gauche*$^+$ to C3, whereas hydrogen acceptors are found in the region between *trans* to C3 and *trans* to C5 (Elgavish & Shaanan, 1997). In the C-type lectins this rule also seems to apply for the 3-

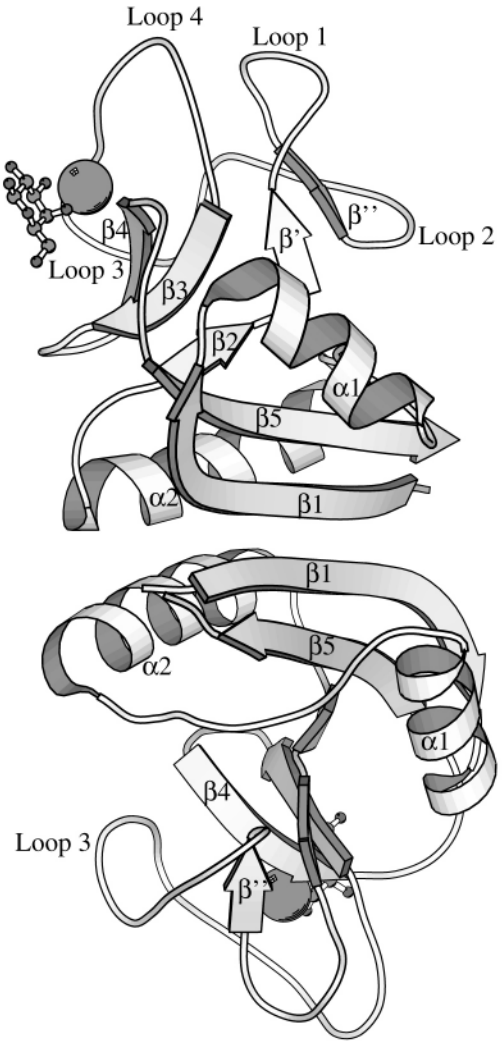

Figure 5. Ribbon diagram of the TC14 dimer. All elements of secondary structure are labelled in the upper molecule, and labelling is repeated for some elements in the lower molecule to improve clarity. The sugar-binding calcium ions are shown as grey spheres and the galactose molecules in a ball-and-stick representation.

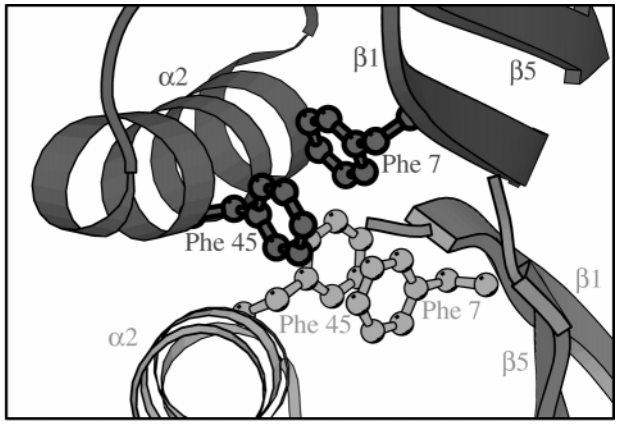

Figure 6. Representation of the hydrophobic interactions at the dimeric interface. The phenylalanine residues involved in hydrophobic contacts are shown in ball-and-stick representations. Molecule A is shown in light grey and molecule B in dark grey.

hydroxyl group, whereas no such pattern can be found when observing all lectin families.

The second element ensuring galactose specificity in TC14 is the tryptophan residue Trp100, which packs against the apolar side of the galactose (see Figure 8(a)). Because of the different stereochemistry of D-mannose, achieving an optimal distribution of binding partners around the 3 and 4-hydroxyl groups would mean a more upright pyranose ring position which would sterically interfere with Trp100. The importance of the packing interaction between the apolar face of D-galactose and an aromatic side-chain has been demonstrated earlier (Rini, 1995), and was a crucial factor in obtaining the galactose-specific QPDWG mutant of MBPA (Iobst & Drickamer, 1994; see Figure 8(b)). The specificity-inducing tryptophan residue in the QPDWG mutant was obtained through a mutation in loop 4 that was held close to the sugar binding site through insertion into loop 4 of five glycine residues. In contrast, the tryptophan (Trp100) in TC14 is located on the opposite side of the binding site in the strand β3, as required by the inverted orientation of the galactose. The mutations in QPDWG had been introduced using mammalian galactose-specific proteins (mostly the asialoglycoprotein receptors) as models, and this difference in the tryptophan position might reflect a general difference between mammalian/vertebrate and invertebrate galactose-specific C-type lectins. Among the known and characterised C-type lectins, only two have a tryptophan residue in the position equivalent to Trp100. One of them, the *Sarcophaga* lectin from flesh fly, is indeed an invertebrate galactose-specific protein (Takahashi *et al*., 1985), whereas in the *N*-glucosamine-specific chicken hepatic lectin (Drickamer, 1981) this tryptophan residue is probably not involved in sugar binding. Another invertebrate D-galactose-specific C-type lectin, the *Drosophila* lectin, has a tyrosine residue in the equivalent position of its sequence, which could play the same role as Trp100 in TC14 (Haq *et al*., 1996). In addition, three C-type lectin domains found in the *C. elegans* genome (genes F47C12.4., F26A1.11. and F26A1.12.) also show a tryptophan in the equivalent position (Wilson *et al*., 1994).

The fact that D-galactose can be bound in two completely different orientations involving a very different tryptophan position by two different C-type lectins shows the remarkable versatility of the C-type lectin binding site. The flexibility of the loop regions allows the introduction of various elements of binding specificity in different spatial arrangements around the primary binding site, therefore creating a wide range of possible functions on the same structural scaffold. The wide dis-

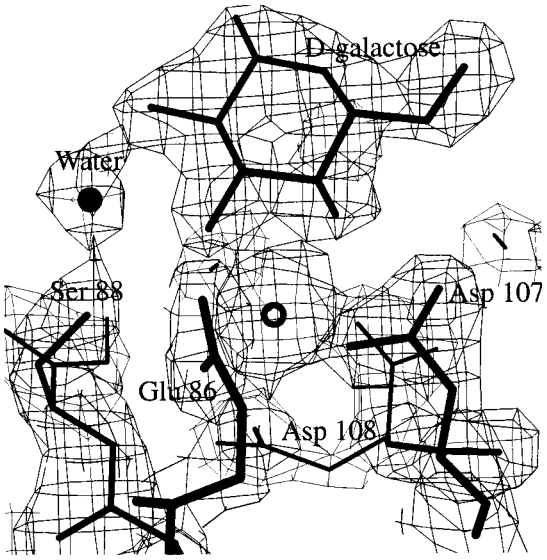

Figure 7. Extract of the final $2|F_o| - |F_c|$ electron density map in the sugar binding site, contoured at 1.7 σ (Figure produced with BOBSCRIPT (Esnouf, 1997)).

tribution of C-type lectins in animals reflects this ability of the fold to evolve very different binding specificities.

Materials and Methods

Materials

Unless otherwise stated, all chemicals used in all experiments were purchased from standard laboratory suppliers and were of the highest available quality. All water used was distilled and purified on an Elgastat system.

Cloning of the gene encoding TC14

A synthetic gene, based on the primary sequence of TC14 as determined by Suzuki *et al.* (1989), was designed with optimal codon usage for expression in *E. coli*. The gene was constructed according to the method by Moore (1989), using four partially overlapping oligonucleotides which were annealed, extended with phage T7 DNA polymerase (Sequenase, U.S. Biochemical) and then cleaved with *Nde*I and *Eco*R1. The cleaved gene was ligated into the protein expression vector pRSETA (Invitrogen Corporation) cut with *Nde*I and *Eco*R1. Sequencing of the DNA product showed that the finally obtained pRSETA-lec vector encoded the correct amino acid sequence.

Protein expression and purification

The pRSETA-lec plasmid was transformed into *E. coli* strain TG2 by the method described by Chung *et al.* (1989). The cells were grown in supplemented M9 minimal medium (Sambrook *et al.*, 1989) at 37 °C until the A_{600nm} reached 0.5 cm^{-1} and then induced by addition of IPTG (to 0.5 mM) and M13 helper phage carrying the gene for phage T7 RNA polymerase. The cells were grown for 16 hours prior to harvesting by centrifugation.

The cells were resuspended in 50 mM Tris-HCl (pH 8.4), 5 mM EDTA, 0.5 mM PMSF and 1% (v/v) Triton X-100 and lysed by sonication. The protein was obtained as inclusion bodies and solubilised in 50 ml of 8 M urea, 50 mM Tris-HCl (pH 8.4), 100 mM NaCl, 0.1 M dithiothreitol (DTT) and 5 mM EDTA at 4 °C for six hours. Following centrifugation and filtration, the supernatant was applied to a DEAE Sepharose fast flow column (Pharmacia) equilibrated in 4 M urea, 1 mM EDTA, 50 mM Tris-HCl (pH 8.4), 5 mM DTT and 100 mM NaCl at a flow rate of 2 ml/minute. The protein was eluted from the column with a linear NaCl gradient to a final concentration of 800 mM. The lectin-containing fractions were refolded by dialysing for 24 hours against four litres of 10 mM Tris-HCl (pH 8.4), 50 mM NaCl, 1 mM $CaCl_2$ and 1 mM *trans*-4,5-dihydroxy-1,2-dithiane, for 24 hours against four litres of 10 mM Tris-HCl (pH 7.2), 50 mM NaCl and 2 mM $CaCl_2$, and finally overnight against four litres of 20 mM imidazole-HCl (pH 7.8), 100 mM NaCl and 20 mM $CaCl_2$ at 4 °C. The refolded protein was applied to an affinity column of immobilised D-galactose agarose resin (Pierce Chemical Co.) pre-equilibrated in the last refolding buffer and washed with approximately 50 ml buffer before the protein was eluted with 5 mM EDTA and 100 mM NaCl. The purity and mass of the affinity-purified protein was confirmed by laser ionisation/time of flight mass spectrometry using a Kratos Kompact 4 spectrometer. Analytical gel filtration was performed on a Pharmacia Superdex 75 column with 150 mM NaCl, 1 mM $CaCl_2$ and 50 mM Tris-HCl (pH 8.0) at a flow rate of 0.75 ml/minute.

Analytical ultracentrifugation

Sedimentation equilibrium analysis was performed in a Beckman Optima XLA analytical ultracentrifuge, using an An-60Ti rotor, at 27 °C and at 14,000 rev/minute, scanning at 280 nm. Runs were overspeeded at 27,000 rev/minute for six hours, to reduce the time to reach equilibrium (Van Holde & Baldwin, 1958) and then scanned every 24 hours until two successive scans superimposed exactly. The later scan was taken as operationally at equilibrium. Mass average apparent molecular weights were calculated with the equation:

$$\bar{M}_{w,app} = \frac{d\ \ln(c)}{dr^2} \quad \frac{2RT}{\omega^2(1-\bar{\nu}\rho)}$$

where c is the concentration at radius r, ω is the angular velocity in radians/second, $\bar{\nu}$ is the partial specific volume (taken as 0.72 ml/g) and ρ is the solvent density (taken as 1.01 g/ml); for running sets of 41 data points and plotted against the concentration for the central point of each set.

As an alternative, the sedimentation data were analysed by direct fitting, using ProFit 5.1, to the equation:

$$A_r = \varepsilon_1\left(c_0 \exp(\sigma(r^2 - r_0^2)) + \frac{2(c_0 \exp(\sigma(r^2 - r_0^2)))^2}{K_d}\right) + A_{err}$$

with:

$$\sigma = \frac{M_1(1-\bar{\nu}\rho)\omega^2}{2RT};$$

where A_r is the absorbance at radius r, ε_1 is the molar extinction coefficient for the monomer (and dimer is assumed to have twice the extinction), c_0 is the monomer

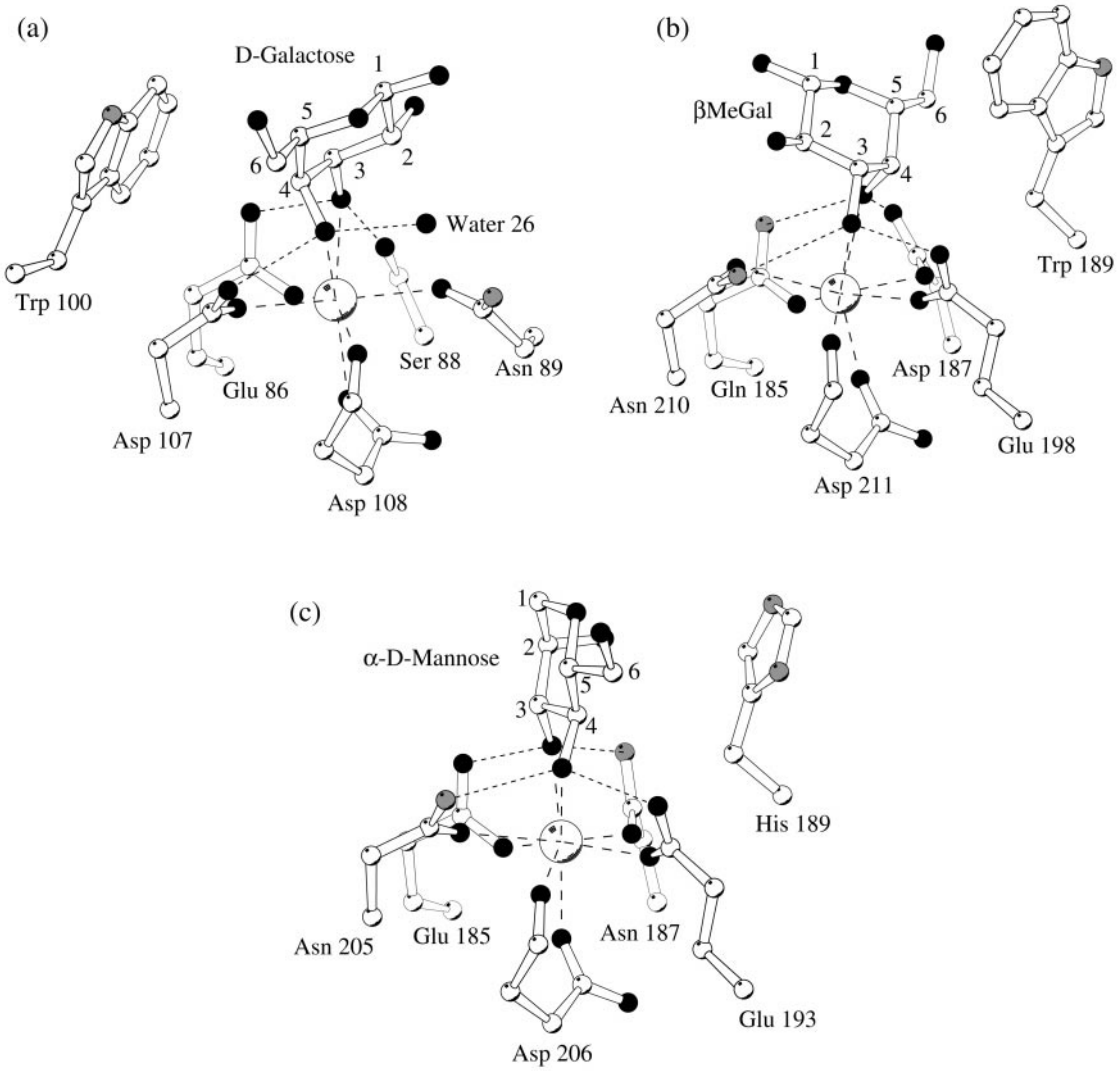

Figure 8. (a) The galactose binding in TC14, with carbon atoms represented in white, oxygen atoms in black and nitrogen atoms in grey. Hydrogen bonds are shown as broken lines. (b) The carbohydrate-binding site of the galactose binding mutant QPDWG of MBPA and (c) the mannose binding site of MBPA, both in the same representation as TC14. The three panels are in the same orientation as obtained by C^{α} superimposition of the three proteins.

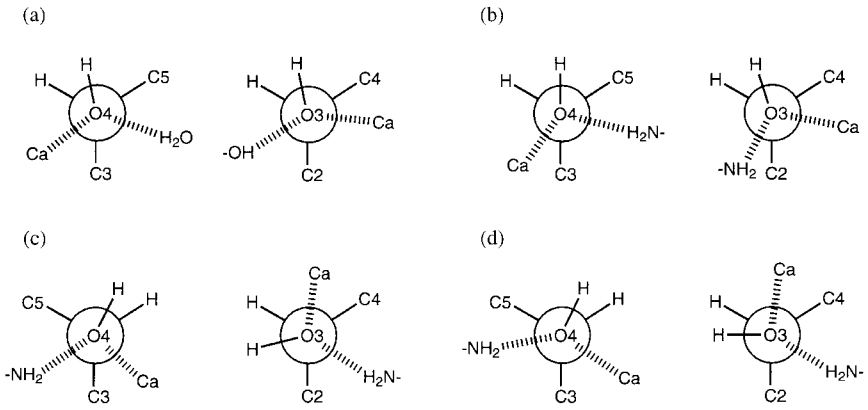

Figure 9. Newman projections along the oxygen-carbon bond of the sugar 3 and 4-hydroxyl groups of the complexes of (a) TC14 with D-galactose, (b) QPDWG with D-galactose, (c) MBPA with D-mannose, and (d) MBPC with D-mannose.

concentration at reference radius r_0, K_d is the dissociation constant for the dimerisation, A_{err} is a correction for error in the baseline determined by overspeeding the rotor to obtain essentially zero protein concentration at the meniscus, M_1 is the monomer mass, and $\bar{\nu}$, ρ and ω are as defined above. ProFit uses a Levenberg-Marquardt algorithm to give a least-squares fit. Estimated errors in the total absorbance used during fitting were taken as the standard deviation of ten readings at each radius during the data collection. The fitted parameters were K_d, A_{err} and M_1, with this latter correcting for any error in the assumed values for $\bar{\nu}$ or ρ. The residuals between the fitted absorbance and the experimental values were plotted, to test for the validity of the model used.

Titration calorimetry

A MicroCal OMEGA calorimeter and ORIGIN software supplied with the instrument were used for the ITC experiments. All solutions were filtered using a Millipore 0.22 μm filter prior to use. Protein concentration was determined by measuring the A_{280} and using the calculated extinction coefficient (Gill & von Hippel, 1989). All experiments were performed at a temperature of 24.7(±0.2)°C. For measuring the calcium affinity, the protein was first dialysed against 5 mM EDTA, 5 mM Mops (pH 7.8) to remove any traces of bound calcium. Subsequently all the experiments were performed in Chelex 100 (BioRad) treated 5 mM Mops (pH 7.8) buffer. The cell was loaded with protein solution at a concentration of 42(±2) μM and titrated with a 1.02 or 2.04 mM $CaCl_2$ solution (concentration determined by complexometric titration of the stock solution). A total of 24 automatic injections of 4 μl each were carried out with 250 seconds between each injection. A 5 mM Mops (pH 7.8) and 5 mM $CaCl_2$ buffer was used for the determination of the sugar-binding affinities. For these experiments, the cell was loaded with 30(±2) μM protein and titrated with 12 injections of 20 μl each of 5 mM sugar solution with 120 seconds between each injection. For all experiments, a blank sample was run in the absence of protein to determine the heat of dilution of the ligand, which was subtracted from the apparent heat of binding prior to data analysis. Integrated heats derived from the injection series were fitted by the software to the one set of sites model, yielding the molar enthalpy of association, the dissociation constant and an estimate for the stochiometry (set to one in the sugar binding experiments).

Protein crystallisation

Crystals were grown at 17 °C using the hanging drop technique. The affinity-purified protein (0.4 mM in 12.5 mM $CaCl_2$, 2.5 mM D-galactose, 10 mM Tris (pH 7.2), 25 mM NaCl) was mixed with an equal amount of a precipitant solution containing 7 % (w/v) PEG 8000, 0.2 M $Zn(ac)_2 \dot{v} 2H_2O$, 0.1 M sodium cacodylate (pH 6), and then equilibrated over a reservoir of the same precipitant solution. Crystals appeared after one day and grew to a size of about 0.1 mm × 0.1 mm × 0.2 mm within a week. They had $P2_12_12_1$ symmetry with $a = 54.18$ Å, $b = 66.45$ Å and $c = 85.90$ Å.

Data collection

All data were collected using an Enraf-Nönius GX13 CuK_α X-ray source and an 18 cm MAR-research imaging plate system (MAR-research Corp.). Crystals were transferred into a cryoprotectant solution consisting of 25 % (v/v) glycerol, 25 mM NaCl, 10 mM Tris (pH 7.2), 2.5 mM D-galactose, 0.1 M sodium cacodylate (pH 6.5), 0.2 M $Zn(ac)_2 \cdot 2H_2O$, 7 % PEG 8000, 12.5 mM $CaCl_2$, and frozen in a stream of nitrogen gas at 100 K using an Oxford Cryostream cooler prior to data collection. The Sm derivative was made by soaking a crystal for three hours in a cryoprotectant solution containing 20 mM $SmCl_3$ instead of the $CaCl_2$. A complex with bound galactose was obtained by soaking a crystal for five minutes in a cryoprotectant containing 30 % (w/v) D-galactose instead of the glycerol. Each dataset was collected from a single crystal. Diffraction intensities were integrated using the programme MOSFLM (Leslie, 1992) and scaled and merged with the CCP4 (1994) programme SCALA (Evans, 1997). The data collection statistics are summarised in Table 3.

Structure solution

Most crystallographic calculations were carried out using the CCP4 programme suite. The Sm sites were located using the programme SOLVE (Terwilliger *et al.*, 1987). The programme SHARP (de la Fortelle & Bricogne, 1997) was used to refine all heavy-atom parameters using both anomalous and isomorphous differences (the phase determination statistics are summarised in Table 4). The initial map showed two molecules and two Sm positions in the asymmetric unit. Solvent flattening was carried out with the programme SOLOMON (Abrahams & Leslie, 1996), using a solvent content of 48 %.

A model containing all the residues from 2 to 123 of both molecules in the asymmetric unit was built into the easily interpretable SIRAS electron density map with the programme O (Jones *et al.*, 1991) and refined with the programme REFMAC (Murshudov *et al.*, 1997), using 19,550 reflections between 35 Å and 2 Å resolution to a final conventional *R*-factor of 20.6 % and a free *R*-factor of 25.0 % based on 1039 random reflections in thin shells of resolution. All the protein atoms of this structure were used as an initial model for the structure solution of the galactose crystal form. After two rounds of rigid body refinement with REFMAC using reflections from 26 Å to 5 Å and 4 Å, refinement was completed by alternative cycles of refinement (with REFMAC using 15,259 reflec-

Table 3. Statistics for the crystallographic data collection

Data set	d_{min} (Å)	Measurements	$\langle I/\sigma \rangle$	Unique reflections	Completeness (at highest res.) (%)	R^a_{merge}
Native	2.0	115,502	8.3	20,632	95.9 (87.9)	6.8 (18.5)
Sm derivative	2.5	29,962	7.2	9707	87.7 (68.2)	8.2 (15.7)
Galactose form	2.2	108,587	11.6	16,111	99.1 (83.8)	5.6 (17.1)

[a] $R_{merge} = \Sigma_{hkl}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I_i(hkl)$.

Table 4. Phase refinement statistics

R_{iso} (%)[a]	18.3
R_{cullis}[b]	0.72
Centric phasing power	1.5
Isomorphous acentric phasing power	1.9
Anomalous acentric phasing power	1.7
Figure of merit before solvent flattening (2.8-2.5 Å resolution)	0.36
Figure of merit after solvent flattening (overall)	0.89
Figure of merit after solvent flattening (2.1-2.0 Å resolution)	0.80

[a] $R_{iso} = \Sigma\ ||F_{deriv}| - |F_{native}|| / \Sigma |F_{native}|$.
[b] $R_{cullis} = \Sigma\ ||F_{PH} \pm F_P| - F_{H(calc)}| / \Sigma |F_{PH} \pm F_P|$, shown for isomorphous, acentric differences.

tions between 26 Å and 2.2 Å) and model building with O. The final model had a conventional *R*-factor of 21.1 % (R_{free} 26.9 %).

Protein Data Bank accession numbers

The coordinates of both final models have been submitted to the Brookhaven Protein Data Bank (accession codes 1BYF for the glycerol complex and 1TLG for the galactose complex).

Acknowledgements

Financial support was obtained from the Medical Research Council of the UK. S.F.P. gratefully acknowledges financial support from an FCO British Chevening Scholarship, the Roche Research Foundation and the Freiwillige Akademische Gesellschaft Basel. G.B.L. is grateful for financial support from the Cambridge Commonwealth Trust, an ORS award, the Herchel Smith Endowment, a Raymond Beverly Sackler Studentship and Gonville and Caius College.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). Methods used in the structure determination of bovine mitochondrial F-1 ATPase. *Acta Crystallog. sect. D,* **52**, 30-42.

Bertrand, J. A., Pignol, D., Bernard, J.-P., Verdier, J.-M., Dagorn, J.-C. & Fontecilla-Camps, J. C. (1996). Crystal-structure of human lithostathine, the pancreatic inhibitor of stone formation. *EMBO J.* **15**, 2678-2684.

Chung, C. T., Niemela, S. L. & Miller, R. H. (1989). One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc. Natl Acad. Sci. USA,* **86**, 2172-2175.

Collaborative Computational Project No. 4 (1994). The CCP4 suite-programs for protein crystallography. *Acta Crystallog. sect. D,* **50**, 760-763.

Cox, J. A., Milos, M. & MacManus, J. P. (1990). Calcium- and magnesium-binding properties of oncomodulin: direct binding studies and microcalorimetry. *J. Biol. Chem.* **265**, 6633-6637.

de la Fortelle, E. & Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472-494.

Drickamer, K. (1981). Complete amino acid sequence of a membrane receptor for glycoproteins. Sequence of the chicken hepatic lectin. *J. Biol. Chem.* **256**, 5827-5839.

Drickamer, K. (1992). Engineering galactose-binding activity into a C-type mannose-binding protein. *Nature,* **360**, 183-186.

Drickamer, K. (1993). Ca^{2+}-dependent carbohydrate-recognition domains in animal proteins. *Curr. Opin. Struct. Biol.* **3**, 393-400.

Elgavish, S. & Shaanan, B. (1997). Lectin-carbohydrate interactions: different folds, common recognition-principles. *Trends Biochem. Sci.* **22**, 462-467.

Esnouf, R. M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**, 132-136.

Evans, P. R. (1997). ''SCALA''. *Joint CCP4 and ESF-EACBM Newsletter for Protein Crystallography,* **33**, 22-24.

Gabius, H. J. (1997). Animal lectins. *Eur. J. Biochem.* **243**, 543-576.

Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319-326.

Graves, B. J., Crowther, R. L., Chandran, C., Rumberger, J. M., Li, S., Huang, K.-S., Presky, D. H., Familletti, P. C., Wolitzky, B. A. & Burns, D. K. (1994). Insight into E-selectin/ligand interaction from the crystal structure and mutagenesis of the lec/EGF domains. *Nature,* **377**, 532-538.

Haq, S., Kubo, T., Kurata, S., Kobayashi, A. & Natori, S. (1996). Purification, characterization, and cDNA cloning of a galactose-specific c-type lectin from *Drosophila melanogaster*. *J. Biol. Chem.* **271**, 20213-20218.

Hohenester, E., Sasaki, T., Olsen, B. R. & Timpl, R. (1998). Crystal structure of the angiogenesis inhibitor endostatin at 1.5 angstrom resolution. *EMBO J.* **17**, 1656-1664.

Iobst, S. T. & Drickamer, K. (1994). Binding of sugar ligands to Ca^{2+}-dependent animal lectins. *J. Biol. Chem.* **269**, 15512-15519.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Crystallog. sect. A,* **47**, 110-119.

Kastrup, J. S., Nielsen, B. B., Rasmussen, H., Holtet, T. L., Graversen, J. H., Etzerodt, M., Thogersen, H. C. & Larsen, I. K. (1998). Structure of the C-type lectin carbohydrate recognition domain of human tetranectin. *Acta Crystallog. sect. D,* **54**, 757-766.

Kawamura, K., Fujiwara, S. & Sugino, Y. M. (1991). Budding-specific lectin induced in epothelial cells is

an extracellular matrix component for stem cell aggregation in tunicates. *Development,* **113**, 995-1005.
Kohda, D., Morton, C. J., Parkar, A. A., Hatanaka, H., Inagaki, F. M., Campbell, I. D. & Day, A. J. (1996). Solution structure of the link module: a hyaluronan-binding domain involved in extracellular matrix stability and cell migration. *Cell,* **86**, 767-775.
Kolatkar, A. R. & Weis, W. I. (1996). Structural basis of galactose recognition by C-type animal lectins. *J. Biol. Chem.* **271**, 6679-6685.
Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein. *J. Appl. Crystallog.* **24**, 946-950.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK-a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.
Leslie, A. G. W. (1992). Recent changes to the to the MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EACMB Newsletter for Protein Crystallography,* **26**.
Milos, M., Schaer, J.-J., Comte, M. & Cox, J. A. (1986). Calcium proton and calcium magnesium antagonisms in calmodulin - microcalorimetric and potentiometric analyses. *Biochemistry,* **25**, 6279-6287.
Mizuno, H., Fujimoto, Z., Koizumi, M., Kano, H., Atoda, H. & Morita, T. (1997). Structure of coagulation factors IX/X-binding protein, a heterodimer of C-type lectin domains. *Nature Struct. Biol.* **4**, 438-441.
Moore, D. D. (1989). Gene synthesis: assembly of target sequences using mutually priming long oligonucleotides. In *Current Protocols in Molecular Biology*, vol. Suppl. 6, pp. 8.2.8-8.2.13, John Wiley & Sons, Inc., New York.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D,* **53**, 240-255.
Ng, K. K. S., Drickamer, K. & Weis, W. I. (1996). Structural analysis of monosaccharide recognition by rat liver mannose-binding protein. *J. Biol. Chem.* **271**, 663-674.
Nicholls, A. (1992). *GRASP: Graphic Representation and Analysis of Surface Properties*, Columbia University, New York, USA.
Nielsen, B. B., Kastrup, J. S., Rasmussen, H., Holtet, T. L., Graversen, J. H., Etzerodt, M., Thogersen, H. C. & Larsen, I. K. (1997). Crystal structure of tetranectin, a trimeric plasminogen-binding protein with an alpha-helical coiled coil. *FEBS Letters,* **412**, 388-396.
Rini, J. M. (1995). Lectin structure. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 551-577.
Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). Bacterial media, antibiotics, and bacterial strains. In *Molecular Cloning: A Laboratory Mannual* (Ford, N. & Nolan, C., eds), pp. A.1-A.13, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sheriff, S., Chang, C. Y. & Ezekowitz, R. A. B. (1994). Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple α-helical coiled-coil. *Nature Struct. Biol.* **1**, 789-794.
Shimada, M., Fujiwara, S. & Kawamura, K. (1995). Expression of genes for two C-type lectins during budding of the ascidian *Polyandrocarpa misakiensis*. *Roux's Arch. Dev. Biol.* **204**, 406-411.
Sonnhammer, E. L., Eddy, S. R., Birney, E., Bateman, A. & Durbin, R. (1998). Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucl. Acids Res.* **26**, 320-322.
Suzuki, T., Takagi, T., Furukohri, T., Kawamura, K. & Nakauchi, M. (1990). A calcium-dependent galactose-binding lectin from the tunicate *Polyandrocarpa misakiensis*. *J. Biol. Chem.* **265**, 1274-1281.
Takahashi, H., Komano, H., Kawaguchi, N., Kitamura, N., Nakanishi, S. & Natori, S. (1985). Cloning and sequencing of cDNA of *Sarcophaga peregrina* humoral lectin induced on injury of the body wall. *J. Biol. Chem.* **260**, 12228-12233.
Terwilliger, T. C., Kim, S.-H. & Eisenberg, D. (1987). Generalized method of determining heavy-atom positions using the difference Patterson function. *Acta Crystallog. sect. A,* **43**, 1-5.
The *C. elegans* Sequencing Consortium (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science,* **282**, 2012-2018.
Van Holde, K. E. & Baldwin, R. L. (1958). Rapid attainment of sedimentation equilibrium. *J. Phys. Chem.* **62**, 734-743.
Weis, W. I. & Drickamer, K. (1996). Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* **65**, 441-473.
Weis, W. I., Kahn, R., Fourme, R., Drickamer, K. & Hendrickson, W. A. (1991). Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. *Science,* **254**, 1608-1615.
Weis, W. I., Drickamer, K. & Hendrickson, W. A. (1992). Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature,* **360**, 127-134.
Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulsen, A., Craxton, M., Dear, S., Du, Z., Durbin, R. *et al*. (1994). 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature,* **368**, 32-38.

Edited by R. Huber

(Received 26 February 1999; received in revised form 24 May 1999; accepted 25 May 1999)