Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the mouse galectin-4 N-terminal carbohydrate-recognition domain reveals the mechanism of oligosaccharide recognition

Veronika Krejčiříková,[a] Petr Pachl,[a] Milan Fábry,[a] Petr Malý,[b] Pavlína Řezáčová[a,c] and Jiří Brynda[a,c]*

[a]Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, v.v.i., Flemingovo nám. 2, 166 37 Prague 6, Czech Republic, [b]Institute of Biotechnology, Academy of Sciences of the Czech Republic, v.v.i., Vídeňská 1083, 142 20 Prague 4, Czech Republic, and [c]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, v.v.i., Flemingovo nám. 2, 166 10 Prague 6, Czech Republic

Correspondence e-mail: brynda@img.cas.cz

Received 30 November 2010
Accepted 1 February 2011

PDB Reference: mouse galectin-4 N-terminal domain CRD1, 3i8t.

Galectin-4, a member of the tandem-repeat subfamily of galectins, participates in cell-membrane interactions and plays an important role in cell adhesion and modulation of immunity and malignity. The oligosaccharide specificity of the mouse galectin-4 carbohydrate-recognition domains (CRDs) has been reported previously. In this work, the structure and binding properties of the N-terminal domain CRD1 were further investigated and the crystal structure of CRD1 in complex with lactose was determined at 2.1 Å resolution. The lactose-binding affinity was characterized by fluorescence measurements and two lactose-binding sites were identified: a high-affinity site with a K_d value in the micromolar range (K_{d1} = 600 ± 70 μ*M*) and a low-affinity site with K_{d2} = 28 ± 10 m*M*.

1. Introduction

Galectins are a family of evolutionarily conserved soluble animal lectins with binding affinity towards various galactose-containing oligosaccharides. Through interaction with cell-surface glycans, galectins regulate the immune response, participate in cell signalling and modulate cell–cell interactions (Rabinovich *et al.*, 2004; Hughes, 2001; Cooper, 2002). Galectins also play roles in regulating the cell cycle, in apoptosis and in cancer (Liu & Rabinovich, 2005; Liu *et al.*, 2002). The members of the galectin family share a conserved carbohydrate-recognition domain (CRD) of about 130 amino acids in length which has an affinity towards various β-galactosides.

15 structurally and functionally defined galectins have been identified to date and a new placenta-specific galectin transcript of the human gene cluster on chromosome 19 has recently been observed (Than *et al.*, 2009). Based on domain composition, the galectin family can be divided into three subfamilies: (i) monomeric (or prototype) galectins that frequently form noncovalent dimers, (ii) chimera-type galactins, the only known representative of which, galectin-3, possesses a C-teminal CRD linked to an N-terminal proline-, glycine- and tyrosine-rich domain, and (iii) the tandem-repeat subfamily characterized by the presence of two CRD domains.

Galectin-4, a member of the tandem-repeat subfamily, is composed of two distinct but homologous domains designated CRD1 (N-terminal) and CRD2 (C-terminal). These domains share 40% identity and are connected by a proline-rich and glycine-rich peptide linker (Huflejt & Leffler, 2004) which is highly sensitive to tissue proteases. In fact, galectin-4 was first discovered as a 17 kDa protein in rat intestinal extract (Leffler *et al.*, 1989); subsequent gene cloning revealed that this protein

was a proteolytic fragment of a larger 36 kDa protein (Oda *et al.*, 1993).

In mammals, galectin-4 is expressed in the epithelium of the small intestine and colon (Gitt *et al.*, 1998; Huflejt & Leffler, 2004). The lipid raft microdomains of intestinal epithelial cells contain galectin-4 as an organizer and stabilizer of the brush border membrane, preventing loss of digestive enzymes and protecting the glycolipid microdomains (Wrackmeyer *et al.*, 2006). The role of carbohydrate recognition by galectin-4 in intestinal inflammation during inflammatory bowel disease (Crohn's disease and ulcerative colitis) has recently been reported (Paclik *et al.*, 2008).

The function of galectin-4 at the molecular level still remains unclear; however, the presence of two non-identical CRD domains suggests a cross-linking function (Brewer, 2002). The binding specificity for separate CRD domains of galectin-4 confirmed differential specificity, selectivity and affinity for particular oligosaccharides (Sorme *et al.*, 2003; Wasano & Hirakawa, 1999; Wu *et al.*, 2002).

Studies of binding affinity have recently been performed for mouse galectin-4 (mGal-4), which shares 73% identity with its human homologue (Marková *et al.*, 2006). Oligosaccharide-binding profiles for recombinant variants of full-length mGal-4 and separate carbohydrate-recognition domains were established. mGal-4 binds to α-linked *N*-acetylgalactosamine (α-GalNAc) and α-linked galactose blood type A and B structures with or without fucose. While the CRD2 domain has a high specificity and affinity for A type-2 α-GalNAc structures, the CRD1 domain has a broader specificity compared with the total binding profile (Marková *et al.*, 2006).

The structures of mammalian galectins have been studied extensively; over 80 crystal structures are presently available. The structural basis of carbohydrate recognition has been addressed for a number of mammalian galectins and the crystal structures of carbohydrate complexes are available for human galectin-1, galectin-3, galectin-7 (López-Lucendo *et al.*, 2004; Seetharaman *et al.*, 1998; Leonidas *et al.*, 1998) and galectin-8 (PDB entry 2yxs; S. Kishishita, A. Nishino, K. Murayama, T. Terada, M. Shirouzu & S. Yokoyama, unpublished work), mouse galectin-9 (Nagae *et al.*, 2006) and fungal galectin-2 (Walser *et al.*, 2004). In this work, we report the crystal structure of the N-terminal carbohydrate-recognition domain of mouse galectin-4 in complex with lactose at 2.1 Å resolution.

2. Experimental

2.1. Protein purification

The coding sequence of mGal-4 CRD1 was cloned into the pQE-31 expression vector (Qiagen) with a His tag located at the N-terminus of the recombinant product. To express CRD1, *Escherichia coli* M15 cells containing the expression vector were grown in LB medium at 310 K and recombinant expression was induced with 1 m*M* IPTG at 303 K. Purification of CRD1 involved two liquid-chromatography steps: affinity chromatography on a lactosyl-agarose column and gel filtration on Superdex 200 HR. The molecular cloning, protein expression and purification of CRD1 have been described in detail previously (Marková *et al.*, 2006).

Table 1
Crystal data and diffraction data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Data-collection statistics	
Space group	$P42_12$
Unit-cell parameters (Å)	a = 90.84, b = 90.84, c = 57.13
No. of molecules in asymmetric unit	1
Wavelength (Å)	0.979
Resolution (Å)	35.62–2.10 (2.18–2.10)
No. of unique reflections	14348 (1374)
Multiplicity	7.6 (5.4)
Completeness (%)	99.2 (97.5)
$R_{\rm merge}$†	0.07 (0.47)
Average $I/\sigma(I)$	9.2 (2.4)
Wilson B (Å^2)	20.5
Refinement statistics	
Resolution range (Å)	35.56–2.10 (2.16–2.10)
No. of reflections in working set	12878 (919)
No. of reflections in test set	722 (53)
R‡ (%)	18.7 (22.6)
$R_{\rm free}$§ (%)	22.4 (19.1)
R.m.s.d. bond lengths (Å)	0.02
R.m.s.d. angles (°)	1.82
No. of atoms in asymmetic unit	1302
No. of protein atoms in asymmetic unit	1170
No. of water molecules in asymmetic unit	90
Mean B value (Å^2)	40.3
Ramachandran plot statistics	
Residues in favoured regions (%)	85.3
Residues in allowed regions (%)	14.7

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where the $I_i(hkl)$ are individual intensities of the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the average intensity of reflection hkl with summation over all data. ‡ $R = \sum_{hkl}||F_{\rm obs}| - |F_{\rm calc}||/\sum_{hkl}|F_{\rm obs}|$, where $F_{\rm obs}$ and $F_{\rm calc}$ are observed and calculated structure factors, respectively. § $R_{\rm free}$ is equivalent to the R value but is calculated for 5% of the reflections chosen at random and omitted from the refinement process (Brünger, 1992).

2.2. Lactose-binding assay

Fluorescence measurements served as a method for the characterization of lactose binding. The emission spectra of CRD1 in the absence or presence of 5 m*M* lactose were measured on an Aminco Bowman DW2000 spectrometer at 298 K. The excitation wavelength was 280 nm (slit width of 4 nm) and emission spectra were recorded in the range 300–400 nm. The sample contained 5 µ*M* CRD1 in a buffer consisting of 10 m*M* sodium phosphate pH 7.4, 140 m*M* NaCl, 5 m*M* EDTA and 5 m*M* β-mercaptoethanol.

For the binding-affinity assay the fluorescence response to successive additions of lactose to a final concentration of 36 m*M* was recorded using a PerkinElmer LS 50B. Fluorescence values are data measured during 30 s (1 Hz reading) of steady state for every measured concentration of lactose. The excitation wavelength was 282 nm (slit width 10 nm) and the emission wavelength was 340 nm (slit width 6 nm). The sample composition was identical to that used for emission-spectra measurement. Experimental data were fitted using an equation describing the two ligand-binding site model,

$$I = I_0 + \frac{\Delta I_1[\mathrm{L}]}{([\mathrm{L}] + K_{\mathrm{d1}})} + \frac{\Delta I_2[\mathrm{L}]}{([\mathrm{L}] + K_{\mathrm{d2}})},$$

where I is the measured fluorescence intensity, I_0 is the initial fluorescence intensity, ΔI_1 and ΔI_2 are the fluorescence difference contributions of binding sites 1 or 2, respectively, k_{d1} and k_{d2} are the dissociation constants for each binding site and [L] is the ligand concentration.

The parameters of the equation were obtained by nonlinear regression using the least-squares method as implemented in the nonlinear toolbox of the *MATLAB* program (MathWorks, 2008).

2.3. Crystallization and data collection

Cocrystals of mGal-4 CRD1 with lactose were prepared at 291 K by the hanging-drop vapour-diffusion method. The reservoir contained 0.1 *M* sodium cacodylate pH 5.0, 80 m*M* lactose, 15%(*w*/*v*) PEG 4000, 0.2 *M* ammonium sulfate, 5 m*M* β-mercaptoethanol and 18%(*v*/*v*) glycerol. For crystal-growth optimization, the microseeding method was employed. The details of the crystallization procedure have been described previously (Krejčiříková *et al.*, 2008).

Diffraction data were collected at 100 K on beamline ID-19 of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory, Argonne, Illinois, USA. Crystal parameters and data-collection statistics are summarized in Table 1.

2.4. Structure determination and refinement

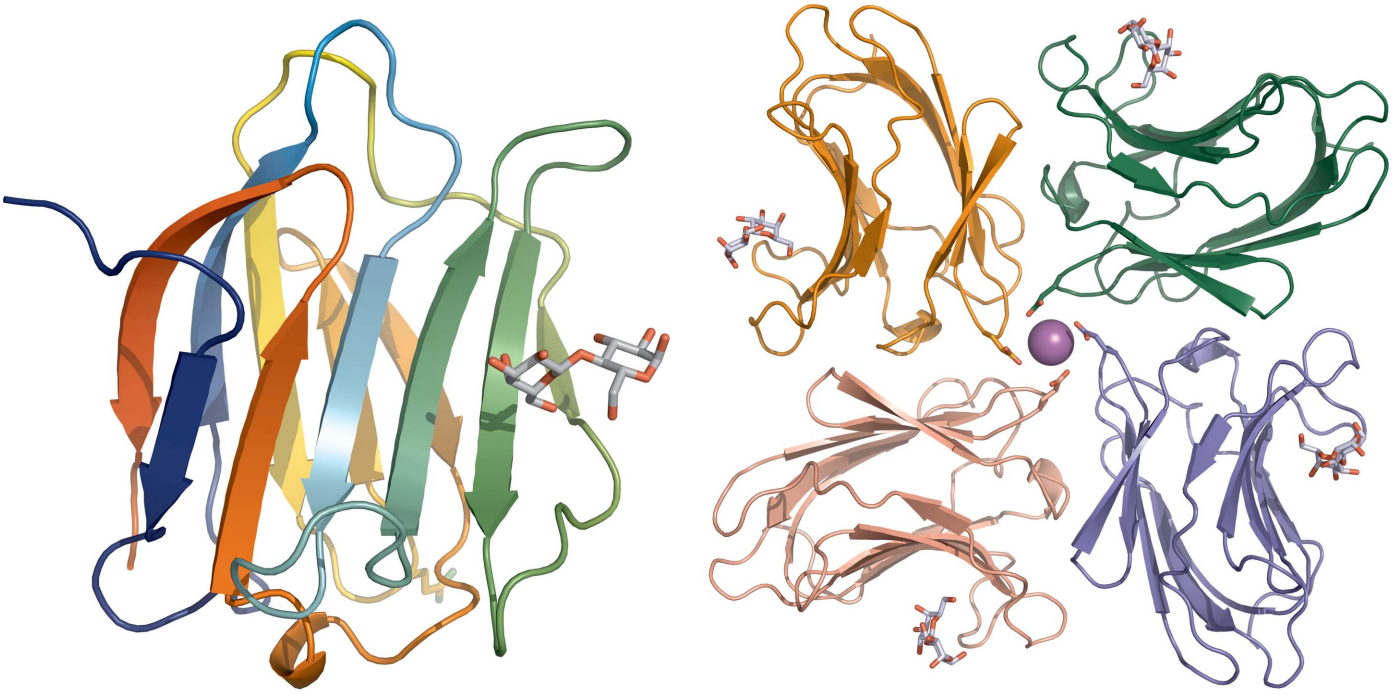
The crystal structure was solved by molecular replacement using the program *MOLREP* (Vagin & Teplyakov, 2010). The search model was derived from the structure of the homologous (40% sequence identity) human galectin-7 (PDB code 2gal; Leonidas *et al.*, 1998).

Model refinement was carried out using the program *REFMAC* v.5.3 (Murshudov *et al.*, 1997) from the *CCP*4 package (Collaborative Computational Project, Number 4, 1994). Manual protein and ligand building and adjustments were performed using the program *Coot* (Emsley & Cowtan, 2004). The quality of the final model was validated using the *MolProbity* service (Chen *et al.*, 2010). Final refinement statistics are summarized in Table 1. All figures showing structural representations were prepared with the program *PyMOL* (DeLano, 2002). The following services were used to analyze the structures: the *PISA* server (Krissinel & Henrick, 2005) and the protein–protein interaction server (Jones & Thornton, 1996).

Atomic coordinates and experimental structure factors have been deposited in the Protein Data Bank with accession code 3i8t.

2.5. Lactose docking

The mGal-4 CRD1 and lactose were prepared for docking in the *YASARA* modelling package (Krieger *et al.*, 2009). H atoms were added to the protein to mimic neutral pH and



(*a*) (*b*)

Figure 1
Crystal structure of the CRD1 domain of mouse galectin-4. (*a*) Overall structure and secondary-structure elements of mGal-4 complexed with lactose. The protein is represented as a rainbow ribbon diagram (blue to red: N-terminus to C-terminus) and the lactose molecule is shown as sticks (C and O atoms coloured grey and red, respectively). (*b*) Crystallographic tetramer of mGal-4 CRD1. Individual monomers are shown in cartoon representation, with lactose bound to the carbohydrate-binding site shown as sticks. The side chains of Glu122 coordinating the Na^+ ion located on the fourfold axis are represented as sticks and the ion is represented as a violet sphere.

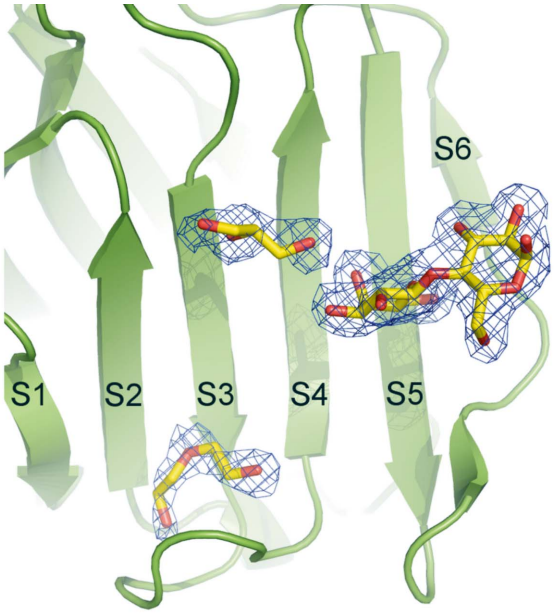
their positions were optimized. The side chains of His59 and His75 were implicated in interactions with a sugar ligand and a protein and thus a diprotonated form was assigned to these residues. The glycerol and water molecules were removed from the model. The parameter set used for the protein was AMBER ff03 (Duan *et al.*, 2003). The ligand was optimized in a vacuum and partial charges on its atoms were obtained by a restrained fit to the electrostatic potential (RESP) at the AM1BCC level (Jakalian *et al.*, 2002). The ligand was then docked to the protein using the *AutoDock* program (Morris *et al.*, 2009). 1000 poses were obtained using a local search protocol; these were subsequently clustered based on similarity (r.m.s.d. < 5 Å) and then scored.

3. Results

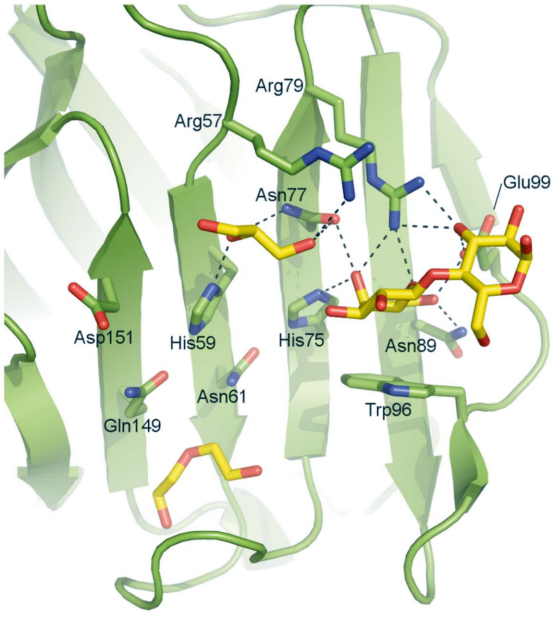
3.1. Overall structure and crystal packing

We determined the crystal structure of the N-terminal carbohydrate-recognition domain (CRD1) of mouse galectin-4 in complex with lactose at a resolution of 2.1 Å. The crystal structure was solved by the molecular-replacement method using the structure of human galectin-7 as a search model. The tetragonal crystal form contained one protein molecule in the asymmetric unit, with a solvent content of 60.9%. The majority of the protein residues could be modelled into the electron-density map, with the exception of 11 N-terminal mGal-4 residues and the N-terminal His tag. The electron density used for modelling the lactose molecule was of excellent quality. The final crystallographic model consisted of 140 protein residues (corresponding to residues 12–164 of the mGal sequence deposited in GenBank as AAC27245.1), one lactose molecule, two glycerol molecules, part of a polyethylene glycol molecule (comprising two ethylene glycol units), one sodium cation and 189 water molecules.

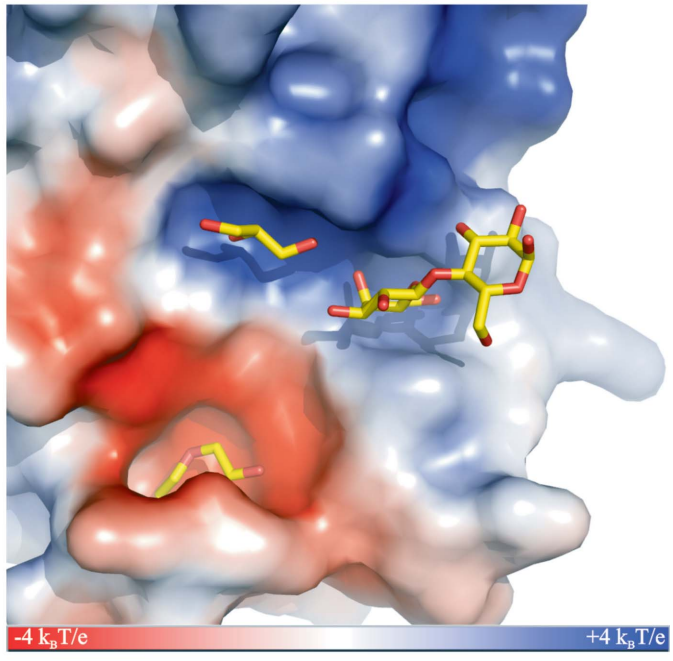
mGal-4 CRD1 exhibits a typical jelly-roll motif composed of two tightly associated β-sheets: a six-stranded β-sheet (S1–S6) and a five-stranded β-sheet (F1–F5) joined by connecting loops. This fold is common to all known galectin structures



(*a*)



(*b*)



(*c*)

Figure 2
Ligand binding to the mGal-4 carbohydrate-binding site. (*a*) Lactose, glycerol and PEG are shown as stick models (C and O atoms coloured yellow and red, respectively), with the $2F_o - F_c$ electron-density maps contoured at 1.0σ. The cartoon representation of the protein shows four strands of β-sheet (labelled S1–S6) forming the carbohydrate-binding site. (*b*) Side chains of the carbohydrate-binding site residues are shown as sticks. Hydrogen-bond interactions are depicted as dashed lines. (*c*) The shape and electrostatic potential of the carbohydrate-binding site are depicted. The protein is represented by its solvent-accessible surface coloured by electrostatic potential (red for negative, blue for positive); ligands are represented as sticks.

(Barondes *et al.*, 1994), with the carbohydrate-binding site being formed by the S4, S5 and S6 strands (Fig. 1*a*).

Analysis of the crystal packing in the mGal-4 crystal structure revealed an interesting quaternary assembly. The asymmetric unit of the mGal-4–lactose complex contains one monomer, which forms a tetramer around the crystallographic fourfold axis (see Fig. 1*b*). A distinct positive difference density (peak at 4σ) on the fourfold axis indicated the presence of an ion. Based on the geometric arrangement of the glutamic acid side chains (Glu122) and the composition of the crystallization buffer, this electron density was assigned as an Na^+ ion (Fig. 1*b*). The interface area buried in the tetrameric assembly is 6830 Å^2, representing about 23.5% of the overall accessible surface area of the four interacting monomers. Each protein monomer loses 48.3% of its surface upon tetramer formation. The interactions at the tetrameric interface have 40.6% polar and 59.4% nonpolar nature. Analysis of the free-energy barrier of tetramer dissociation ($\Delta G^{\rm diss}$) using the *PISA* server (Krissinel & Henrick, 2005) yielded a $\Delta G^{\rm diss}$ value of 348 kJ mol^{-1}. This positive value suggests the stability of the tetrameric assembly in the solution. Interestingly, when the sodium ion was omitted from the analysis the protein tetrameric assembly was evaluated to be thermodynamically unstable.

3.2. Ligand binding in the crystal structure

The lactose molecule was modelled into well defined electron density observed in the canonical sugar-binding site (Fig. 2*a*) of mGal-4 CRD1 formed by the S4, S5 and S6 β-strands. Residues His75, Asn77, Arg79, Asn89 and Glu99 are involved in forming nine hydrogen bonds to the lactose molecule. In addition to direct hydrogen bonds, the side chain of His59 interacts with lactose through a water-mediated hydrogen bond (Fig. 2*b*). Residue Trp96 also contributes to lactose binding through the interaction of the lactose H atoms with π-electrons from the aromatic system of the Trp96 side chain.

The two carbohydrate units of lactose (galactose and glucose) contribute differently to the interaction with the sugar-binding site. The galactose moiety is deeply buried in a deep pocket, while the less stabilized glucose residue is more exposed to the solvent. Six of the nine direct hydrogen bonds are formed between the protein and the galactose moiety. The average atomic displacement parameters (ADPs) for the galactose moiety are also much lower than those for the glucose moiety (38.8 *versus* 47.0 Å^2), indicating higher stabilization of galactose upon binding.

The galactose is bound in a surface pocket lined by positively charged residues (Fig. 2*c*). The binding pocket for galactose in mGal-4 can be described as a groove which extends from the bound lactose towards the centre of the S side of the β-sheet and forms a deep cavity lined by negatively charged residues at the base of the S2 strand (Fig. 2*c*). A distinct negatively charged pocket is also notable in the vicinity of the lactose-binding site. This pocket, which is located within a distance of 4–5 Å of the bound lactose, is formed by a loop connecting β-strands S3 and S4.

Both pockets adjacent to the lactose-binding site allow the binding of two components of the crystallization solution:

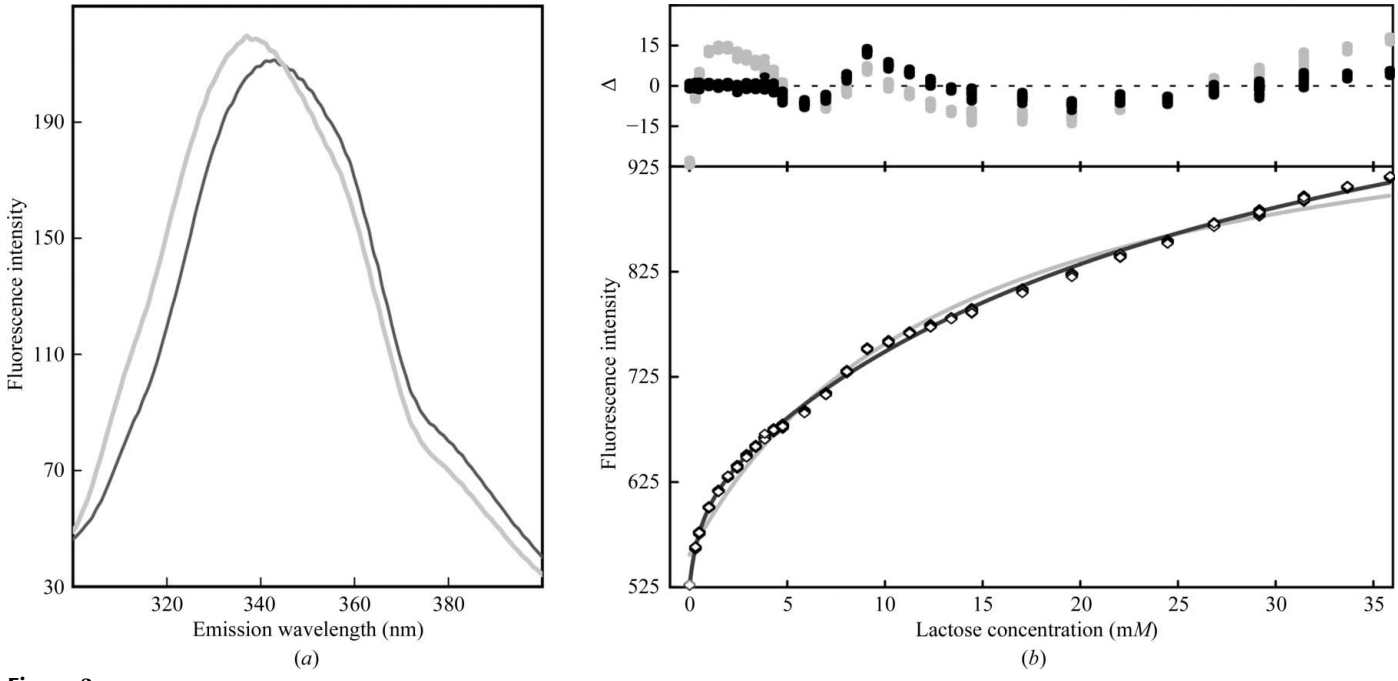


Figure 3
Lactose-binding analysis. (*a*) Fluorescence emission spectra of free mGal-4 CRD1 (black line) and upon the addition of lactose (grey line). The excitation wavelength was 280 nm. (*b*) Response of the fluorescence intensity to the concentration of lactose added to a solution of 5.2 µ*M* mGal-4 CRD1. Diamonds represent measured intensities. Least-squares best fits for the experimental data were calculated for a one-site binding model (grey line) and for a two-site binding model (black lines). Residuals expressed as differences (Δ) between the theoretical curves and the experimental data in the top plot show a much better fit for the two-site model (black circles) than for the one-site model (grey circles).

glycerol and polyethylene glycol (Fig. 2). Polyethylene glycol 4000 and glycerol were present in the crystallization mixture at concentrations of 15 and 18%(*v*/*v*), respectively. All atoms of glycerol and two ethylene glycol units of PEG 4000 could be modelled into the electron-density maps (Fig. 2*a*).

The molecule of glycerol interacts by polar interactions with the side chains of Arg57, Asn77 and His59. These polar interactions as well as the conformation of glycerol are highly similar to those observed for the adjoining galactose moiety of the bound lactose (Fig. 2).

The PEG molecule interacts with residues in a loop connecting the antiparallel β-sheets S3 and S4 and with residues from β-sheets S2, S3 and S4. A deeply buried terminal O atom of the PEG molecule forms two polar interactions with Asp67 and Gly65. An O atom of the PEG molecule at the opening of the binding pocket is located within a hydrogen-bonding distance of 2.9 Å of the carbonyl O atom of Ala70; however, the hydrogen bond is not achievable in the context of the whole PEG polymer. In addition to polar interactions, the PEG molecule makes numerous van der Waals interactions with residues Lys32, Asn61, Ala63, Gly65, Asp67, Asp68, Gly69, Ala70, Val72 and Gln149.

Comparison of our structure with the crystal structure of unliganded mGal-4 CRD1 (PDB entry 2dyc; M. Kato-Murayama, K. Murayama, T. Terada, M. Shirouzu & S. Yokoyama, unpublished work) revealed that ligand binding does not cause any structural changes in the protein. Apart from the different N- and C-termini (which contain cloning artifacts), these two structures are highly similar, with an r.m.s.d. of 0.31 Å for 130 C^{α} atoms (residues 27–163 and 15–151 of our structure and 2dyc, respectively). Superposition of all identical atoms yielded an r.m.s.d. of 0.90 Å for 1065 atoms. Major differences can be observed in the structures of the surface-exposed loop connecting β-sheets S4 and S5, probably owing to the inherent flexibility of this region.

3.3. Lactose-binding studies

Since a tryptophan residue is involved in lactose binding, the change in its intrinsic fluorescence could be followed and used in mGal-4 ligand-binding assays (Iglesias *et al.*, 1998). Using an excitation wavelength of 280 nm, the fluorescence emission spectra of mGal-4 CRD1 had a maximum at 343 nm. Upon the addition of lactose, the fluorescence maxima shifted to a shorter wavelength of 336 nm and the intensity of fluorescence emission increased (Fig. 3*a*). This specific effect on fluorescence was used to determine the dissociation constant. The values of the dissociation constant were determined on the basis of the response of the fluorescence signal to incremental additions of lactose (Fig. 3*b*). The fluorescence intensity is proportional to the concentration of the bound ligand and thus it could be used as a measure of the lactose bound to the CRD1 of the m-Gal4. In the bound to free ligand ratio, the depletion of lactose upon binding to the protein could be neglected for the range of protein and ligand concentrations used.

The experimental data were fitted using a one-site binding model; however, a better fit was obtained using a two non-equivalent binding sites model (Fig. 4*b*). The best fit to a two-site model provided K_d values of $K_{d1} = 600 \pm 70$ μM and $K_{d2} = 28 \pm 10$ mM.

In order to obtain insight into the location and molecular interactions of the lactose molecule in the second binding site, a docking study was performed. Possible lactose binding in the vicinity of the lactose-binding site that was observed in the crystal structure was probed *in silico*. The majority of the resulting poses docked the lactose molecule into the standard binding site; however, some poses were docked into a secondary binding site (Fig. 4). This second binding site was formed by residues Pro30, Arg57, Phe58, His59, Asn61, His75, Asn77, Gln149, Asp151 and Gly152. The polar interaction of the glucose moiety with Arg57, His59 and Asn77 resembled the binding of the glycerol molecule that was observed in the crystal structure. This second binding pose placed some galactose atoms in proximity to residue Trp96 (the distance between the galactose C6 and O6 atoms and the Trp96 $C^{\varepsilon 2}$ and $C^{\delta 1}$ atoms is <6 Å).

4. Discussion

Galectin-4 is a member of the galectin family with two carbohydrate-recognition domains. The amino-acid sequence of mouse galectin-4 shares 91% similarity (76% identity) to human galectin-4, for which the crystal structure is not available. We have solved the crystal structure of the N-terminal carbohydrate-binding domain CRD1 of mouse galectin-4 in complex with lactose. The structure of CRD1 has the all-β fold typical of other members of the galectin family formed by two

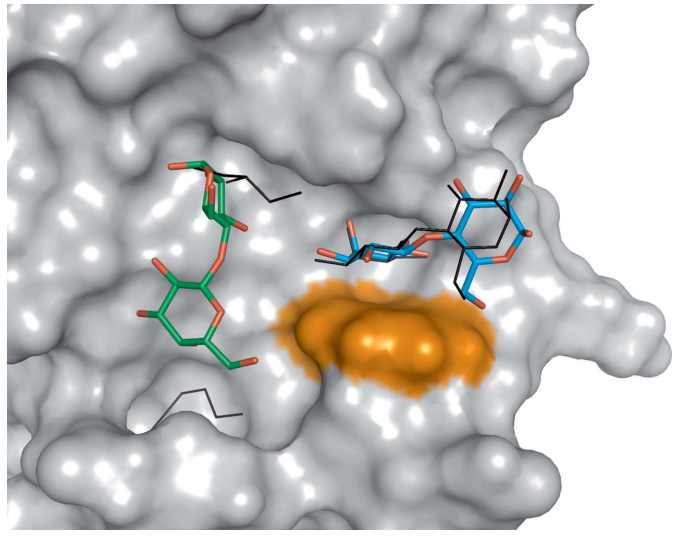


Figure 4
Docking of lactose into the mGal-4 CRD1 carbohydrate-binding site. The molecular surface of mGal-4 CRD1 is coloured grey, with the Trp96 residue highlighted in orange. Small-molecule ligands (lactose, glycerol and PEG) determined in the crystal structure are depicted as black wires. Two poses of the lactose molecule resulting from docking are depicted as stick models (O atoms are coloured red and C atoms light blue and green).

β-sheets containing six and five strands (Leffler *et al.*, 2004). The closest known structural homologue of the galectin-4 CRD1 is the galectin-9 N-terminal CRD (Nagae *et al.*, 2006), with an r.m.s.d. for the main-chain atoms of below 0.9 Å, yet the structural similarity to other galectin structures is also high.

Although mGal-4 CRD1 was only detected as a monomer in solution, it forms a crystallographic tetramer with an extensive interaction surface area (Fig. 1*b*) stabilized by a sodium cation located on the tetramer fourfold axis. A crystallographic tetrameric assembly was also identified in the crystal structure of free mGal-4 CRD1 solved by others (PDB code 2dyc). Both the free and ligand-bound mGal-4 crystallized in tetragonal crystal forms; however, the space groups differed ($I422$ and $P42_12$ for the free and liganded structures, respectively). The crystal packing differs in these two crystal structures, but despite this the mGal-4 CRD1 monomers are arranged into a highly similar tetrameric assembly which superposes with an r.m.s.d. of 0.41 Å for 544 C^{α} atoms of the four mGal-4 molecules. Although the tetrameric assembly in both crystal structures is the result of the presence of a crystallographic fourfold axis, its repeated occurrence in the structure of free and liganded CRD1 points to some biological relevance. This tetrameric assembly would also be achievable for the whole mGal-4, since the C-termini connecting the N-terminal CRD1 domain to CRD2 are free in this tetrameric arrangement.

The physiological relevance of the tetrameric quaternary assembly observed in the crystal remains an open question. Analysis of the oligomeric state of mGal-4 CRD1 in solution by gel filtration did not reveal any tetramerization. However, the formation of an mGal-4 tetramer might be plausible in the specific environment at the outer membrane of intestinal cell walls, where a high content of galectin-4 has been observed (Danielsen & van Deurs, 1997). Further studies are required in order to evaluate the physiological relevance of mGal-4 tetrameric assembly.

The lactose molecule binds in the binding site formed by the S4, S5 and S6 β-strands and makes an extensive hydrogen-bonding network with the side chains of amino-acid residues His75, Asn77, Arg79, Asn89 and Glu99 (Fig. 2*b*). Moreover, the galactose residue of the lactose forms a stacking interaction with the indole group of Trp96. This interaction represents a feature that is typical of other galectins and some other members of the lectin family (Elgavish & Shaanan, 1997; Sujatha *et al.*, 2004). We used this specific contribution of the Trp96 residue to carbohydrate binding to characterize lactose binding to mGal-4 CRD1 using intrinsic tryptophan fluorescence measurements (Lakowicz, 2006).

This method has previously been used to determine the binding affinity of ovine galectin-1 to lactose (Iglesias *et al.*, 1998): the quenching of fluorescence observed upon lactose binding allowed determination of a $K_{\rm d}$ value of 157 ± 20 μM. In binding assays performed for the CRD1 of mGal-4, lactose binding resulted in an increase in fluorescence intensity and we identified two binding sites for lactose on mGal-4 CRD1. The first binding site is of high affinity, with a $K_{\rm d}$ value in the micromolar range ($K_{\rm d1} = 600 \pm 70$ μM), while the second binding site exhibits significantly lower affinity ($K_{\rm d2} = 28 \pm 10$ mM).

In the crystal structure only the high-affinity binding site corresponding to the canonical carbohydrate-binding pocket is occupied by lactose; no second lactose was found. Instead, a glycerol molecule and polyethylene glycol were bound in the adjacent pocket extending from Trp96 across the S β-sheet (Fig. 2). During crystallization, glycerol and PEG were present at high concentrations and thus the second molecule of lactose could be competed out from the low-affinity binding site.

Glycerol and polyethylene glycol binding represents a crystallization artifact, but nevertheless these simple molecules probably mark additional binding pockets for complex carbohydrates, which are the physiological ligands of galectin-4. In the oligosaccharide-binding profile analysis mGal-4 CRD1 exhibited a rather broad affinity for complex glycans, especially lactose derivatives modified at C2 and C3 (Marková *et al.*, 2006). The groove spanning from the position of the glycerol molecule to the PEG molecule (see Fig. 2) is well defined in shape and is lined with residues that are available for polar interactions: His59, Asn61, Asp67, His75, Asn89, Glu99, Gln149 and Asp151 (Fig. 2). It is large enough to accommodate a second lactose molecule and well positioned with respect to the high-affinity lactose-binding site to allow the binding of other carbohydrate derivatives in complex branched glycans.

In conclusion, the fine details of the complex saccharide-binding site in the mGal-4 CRD1 domain obtained in this study could provide leads for the design of specific reagents for probing and/or inhibiting its function.

This study was supported by grants from the Grant Agency of the Czech Republic (Nos. 203/09/0820 and 304/03/0090), by grant 1M0505 from the Czech Ministry of Education and by the institutional research projects AV0Z40550506, AV0Z50520514 and AV0Z50520701 awarded by the Academy of Sciences of the Czech Republic. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-06CH11357. The authors would like thank Dr Martin Lepšík for help with the docking, Devon Maloy for critical reading of the manuscript and the EMBL for access to the X12 beamline at DESY in Hamburg.

References

Barondes, S. H., Cooper, D. N., Gitt, M. A. & Leffler, H. (1994). *J. Biol. Chem.* **269**, 20807–20810.
Brewer, C. F. (2002). *Biochim. Biophys. Acta*, **1572**, 255–262.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cooper, D. N. (2002). *Biochim. Biophys. Acta*, **1572**, 209–231.
Danielsen, E. M. & van Deurs, B. (1997). *Mol. Biol. Cell*, **8**, 2241–2251.

DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J. & Kollman, P. (2003). *J. Comput. Chem.* **24**, 1999–2012.
Elgavish, S. & Shaanan, B. (1997). *Trends Biochem. Sci.* **22**, 462–467.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gitt, M. A., Colnot, C., Poirier, F., Nani, K. J., Barondes, S. H. & Leffler, H. (1998). *J. Biol. Chem.* **273**, 2954–2960.
Huflejt, M. E. & Leffler, H. (2004). *Glycoconj. J.* **20**, 247–255.
Hughes, R. C. (2001). *Biochimie*, **83**, 667–676.
Iglesias, M. M., Rabinovich, G. A., Ivanovic, V., Sotomayor, C. & Wolfenstein-Todel, C. (1998). *Eur. J. Biochem.* **252**, 400–407.
Jakalian, A., Jack, D. B. & Bayly, C. I. (2002). *J. Comput. Chem.* **23**, 1623–1641.
Jones, S. & Thornton, J. M. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 13–20.
Krejčiříková, V., Fábry, M., Marková, V., Malý, P., Řezáčová, P. & Brynda, J. (2008). *Acta Cryst.* F**64**, 665–667.
Krieger, E., Joo, K., Lee, J., Raman, S., Thompson, J., Tyka, M., Baker, D. & Karplus, K. (2009). *Proteins*, **77**, Suppl. 9, 114–122.
Krissinel, E., Henrick, K. (2005). *CompLife* 2005, edited by M. R. Berthold, R. Glen, K. Diederichs, O. Kohlbacher & I. Fischer, pp. 163–174. Berlin, Heidelberg: Springer-Verlag.
Lakowicz, J. R. (2006). *Principles of Fluorescence Spectroscopy*, 3rd ed. New York: Kluwer/Plenum.
Leffler, H., Carlsson, S., Hedlund, M., Qian, Y. & Poirier, F. (2004). *Glycoconj. J.* **19**, 433–440.
Leffler, H., Masiarz, F. R. & Barondes, S. H. (1989). *Biochemistry*, **28**, 9222–9229.
Leonidas, D. D., Vatzaki, E. H., Vorum, H., Celis, J. E., Madsen, P. & Acharya, K. R. (1998). *Biochemistry*, **37**, 13930–13940.
Liu, F.-T., Patterson, R. J. & Wang, J. L. (2002). *Biochim. Biophys. Acta*, **1572**, 263–273.
Liu, F.-T. & Rabinovich, G. A. (2005). *Nature Rev. Cancer*, **5**, 29–41.
López-Lucendo, M. F., Solís, D., André, S., Hirabayashi, J., Kasai, K., Kaltner, H., Gabius, H.-J. & Romero, A. (2004). *J. Mol. Biol.* **343**, 957–970.
Marková, V., Smetana, K. Jr, Jeníková, G., Láchová, J., Krejciríková, V., Poplstein, M., Fábry, M., Brynda, J., Alvarez, R. A., Cummings, R. D. & Maly, P. (2006). *Int. J. Mol. Med.* **18**, 65–76.
MathWorks (2008). *MATLAB – The Language Of Technical Computing*,. http://www.mathworks.com/products/matlab/.
Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. & Olson, A. J. (2009). *J. Comput. Chem.* **30**, 2785–2791.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Nagae, M., Nishi, N., Murata, T., Usui, T., Nakamura, T., Wakatsuki, S. & Kato, R. (2006). *J. Biol. Chem.* **281**, 35884–35893.
Oda, Y., Herrmann, J., Gitt, M. A., Turck, C. W., Burlingame, A. L., Barondes, S. H. & Leffler, H. (1993). *J. Biol. Chem.* **268**, 5929–5939.
Paclik, D., Danese, S., Berndt, U., Wiedenmann, B., Dignass, A. & Sturm, A. (2008). *PLoS One*, **3**, e2629.
Rabinovich, G. A., Toscano, M. A., Ilarregui, J. M. & Rubinstein, N. (2004). *Glycoconj. J.* **19**, 565–573.
Seetharaman, J., Kanigsberg, A., Slaaby, R., Leffler, H., Barondes, S. H. & Rini, J. M. (1998). *J. Biol. Chem.* **273**, 13047–13052.
Sorme, P., Kahl-Knutson, B., Wellmar, U., Nilsson, U. J. & Leffler, H. (2003). *Methods Enzymol.* **362**, 504–512.
Sujatha, M. S., Sasidhar, Y. U. & Balaji, P. V. (2004). *Protein Sci.* **13**, 2502–2514.
Than, N. G. *et al.* (2009). *Proc. Natl Acad. Sci. USA*, **106**, 9731–9736.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Walser, P. J., Haebel, P. W., Kunzler, M., Sargent, D., Kues, U., Aebi, M. & Ban, N. (2004). *Structure*, **12**, 689–702.
Wasano, K. & Hirakawa, Y. (1999). *J. Histochem. Cytochem.* **47**, 75–82.
Wrackmeyer, U., Hansen, G. H., Seya, T. & Danielsen, E. M. (2006). *Biochemistry*, **45**, 9188–9197.
Wu, A. M., Wu, J. H., Tsai, M.-S., Liu, J.-H., André, S., Wasano, K., Kaltner, H. & Gabius, H.-J. (2002). *Biochem. J.* **367**, 653–664.