PIBB 生物化学与生物物理进展
Progress in Biochemistry and Biophysics
2011, 38(9): 810~815
www.pibb.ac.cn

A Novel Binding Pattern Unique in Two Ligands for One Carbohydrate Recognition Domain in Galectins*

BIAN Cheng-Feng[1,2], ZHANG Ying[1], LI De-Feng[1], WANG Da-Cheng[1]**
([1] *National Laboratory of Biomacromolecules, Institute of Biophysics, The Chinese Academy of Sciences, Beijing* 100101, China;
[2] *Graduate School of The Chinese Academy of Sciences, Beijing* 100039, China)

Abstract Galectins are a protein family with diverse biological functions, which are unique in specifically recognition and binding with β-galactosides as the primary structural basis for its functional performance. So far, all structurally characterized galectins display a conservative binding mode for the β-galactoside-containing carbohydrate ligands, in which one carbohydrate recognition domain (CRD) binds only one ligand. Here a novel binding pattern unique in two carbohydrate ligands for one CRD was reported, which is observed from the structure of Gal-3 CRD complexed with glycan TFN. In this doublet binding sites, Site 1 and Site 2, two TFN molecules interact with the CRD domain *via* two hydrogen-bond networks mediated by certain water molecules, respectively. The mutagenesis analysis shows that one of the binding sites, Site 1, is basic and essential for the carbohydrate ligands with a conservative binding mode, which should be commonly existed in galectins. While, the other binding site (Site 2) is easily discarded in a small structural interference from a single-site mutation, which illustrates that it should be conditionally appeared to play an additional and auxiliary role in ligand binding. The stereo-chemical analysis indicates that this doublet binding pattern may be suitable for some glycans with certain rather complicated constitutions, like branched structure. The possible functional role of this doublet binding pattern is also discussed.

Key words galectins, β-galactosides, carbohydrate recognition domain (CRD), glycan TFN, doublet binding pattern
DOI: 10.3724/SP.J.1206.2011.00162

The carbohydrate recognition domain (CRD) is widely found in a range of proteins, which can recognize some kinds of oligosaccharides to mediate interactions between proteins and glycoproteins or glycolipids. Galectins, a family of cell-surface lectins, are characterized by their affinity for recognizing β-galactoside containing oligosaccharides through the CRD. Their CRDs are evolutionarily conserved to be constituted of the shared consensus sequences of about 130 amino acids [1–5], which form two anti-parallel β-sheets showing a folding β-jelly-roll or β-sandwich topology. To date, sixteen mammalian galectins have been identified, which can be subdivided into three sub-families according to their three-dimension structures. Proto-galectins have one unique CRD, chimera-galectins contain at least one non-carbohydrate domain, and tandem-galectins possess two CRDs connected by a short linker peptide [6–7]. All these galectins, existing intracellularly including in the nucleus or extracellularly on the cell surface, fulfill diverse biological functions, such as impacting on the regulation of inflammation and allergies, cell adhesion, cell growth and proliferation, cell apoptosis, host-pathogen interactions and/or tumor progression. They mainly resort to the indispensable CRDs involving in distinct substantial glycoconjugate recognitions [8–11]. It is known that there are a lot of multifunctional counter receptors for CRDs including

*This work was supported by grants from National Basic Research Program of China (30800170) and The Special Project on Drug Discovery from The Ministry of Public Health of China (2009ZX09103-676).

**Corresponding author.
Tel: 86-10-64888547, E-mail: dcwang@ibp.ac.cn

Received: April 14, 2011 Accepted: April 29, 2011

mucins, fibronectin, neuropilin, tenascin, integrins, laminin, glycolipid GM1, as well as cell-type-specific glycoprotein CD3, CD4, CD7, CD43, CD45 and so on[12-13]. Investigating the binding properties of CRD for carbohydrates is essential for understanding the functional performance of galectins.

So far, dozens of structures of galectins complexed with various glycans have been determined and shown a common regular pattern for CRD, in which one CRD can only bind to one carbohydrate ligand, which has already been a traditional view in the CRD investigation. Here we report a novel observation from the structure of Galectin-3 (Gal-3) CRD complexed with its glycan ligand TFN, which shows unexpectedly two TFN molecules in binding with one CRD of Gal-3. Galectin-3 with ~31 ku belongs to the type Ⅱ of galectins, which is composed of unusual tandem repeats of short amino-acid stretches in N-terminal, following by the CRD domain in C-terminal.

1 Materials and methods

1.1 Purification and crystallization

The protein used in the structural determination is the C-terminal part of Gal-3 containing a CRD domain (Gal-3 CRD). TFN (Galβ1-3GalNAcα1-*p*-nitrophenyl) is a glycan containing Thomsen-Friedenreich (TF) antigen, consisting of TF *p*-nitrophenyl.

Gal-3 CRD (117~250) and its single-site mutant E165A were recombined into the vector pET22b with no His-tag, and expressed in *Escherichia coli* strain BL21 (DE3), respectively. They were preliminarily purified on the lactosyl Sepharose column essentially as described previously [2]. The lactose used to elute target proteins, was removed *via* the size-exclusion chromatography with Hiload Superdex75 16/60 column (GE Healthcare). After two steps of chromatography, proteins were concentrated to ~30 g/L in the solution buffer, containing 10 mmol/L K, Na-phosphate and 100 mmol/L NaCl.

Co-crystals of Gal-3 CRD or E165A binding with TFNs were acquired using the hanging drop vapor diffusion method at room temperature. Before co-crystallization, TFN (from Merk) and proteins were mixed in a molecular ratio of 1∶1 and placed at 4℃ for 2 h. The crystals, in a hexagon-cylinder shape, appeared in the mother liquid condition of 2 mol/L $(NH_4)_2SO_4$ and 0.1 mol/L Bis-Tris pH 6.0 after 2 days.

1.2 Data collection, structure determination and refinement

Diffraction data were collected at Beamline 17U synchrotron, Shanghai Synchrotron Radiation Facility (SSRF, 17U). Before collections, crystals were transferred and soaked into the cryo-protectant including the mother liquid and 10% (*v*/*v*) glycerol for 15 s and then flash frozen to 83K.

Data were processed and merged using *MOSFLM*[14] and *CCP4* [15] programs. The phases were resolved by the molecule replacement (MR) method [16]. The TFN was added according to the shape of the associated electron density map, and further structural refinements were performed in a circle of *CNS*-*COOT*-regulating manually[17].

2 Results and discussions

2.1 A novel binding pattern containing two carbohydrate ligands in one CRD domain

The structure of the complex was solved at 1.9Å in $p3_221$ space group, and the final R and R_{free} factors were refined to 18% and 22%, respectively. Generally, Gal-3 CRD is folded into a β-sandwich formed by two anti-parallel β-strands, F1-F5 and S1-S6. The refined structure shows that there are two TFN molecules in binding with one CRD domain of Gal-3 in an asymmetric unit (Figure 1a). The 2Fo-Fc omit electron density maps show the definite and clear stereo-chemistry of two TFN molecules (Figure 1b).

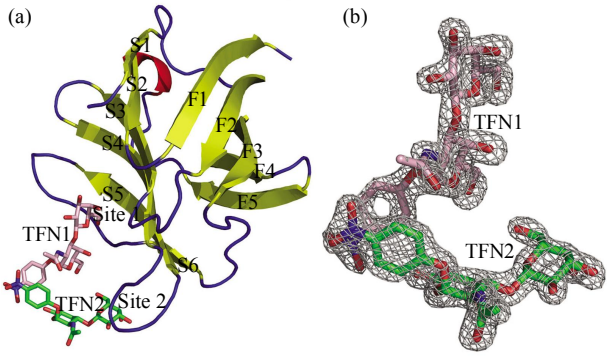

Fig. 1 Schematic ribbon representation of overall structure of Gal-3 CRD complexed with TFN as viewed toward the binding site

(a) The bound TFN molecules are shown in a stick representation. (b) The fit of two TFN molecules to the final 2Fo-Fc omit electron density map, contoured at 1.5 σ, is also excellent. The orientation of TFNs is the same as presented in left.

One of the TFN molecules, TFN1, is plunged into the conservative carbohydrate binding site, which is formed by β-strands S4～S6. While the other TFN molecule, TFN2, is lodged in a novel binding site *via* recognizing its β-galactose moiety, which is not found before in other galectins (Figure 2a). All of the carbohydrate moieties are bound in a boat conformation. The deeply buried β-galactose moiety of TFN1 occupies the same position as that in other galectins, such as Galectin-1 and AAL, a mushroom galectin from *Agrocybe aegerita* [18-19]. It is anchored through the van der waals interactions between Trp181 and the carbon atoms C3, C4, C5 and C6 on one face, and *via* a hydrogen-bond network, involved residues Arg148, His158, Asn160, Arg162 and Asn174, to recognize the corresponding atoms just like that was reported before [20]. The additional TFN2 is contact with TFN1 molecule and located in an interesting position at the bottom of β-strands S5-6 and the loop L4 connecting the two strands (Figure 2c). The β-galactose (Gal) moiety is grasped by resorting to 4 waters, which are indispensable and coordinately connected to 4 residues, Glu165, Asn166, Glu184 and Arg186, respectively, in the same asymmetric unit to form a water-mediated hydrogen-bond network. In details, Glu165 mediating W1 recognizes the hydroxyl group (—OH) on C2′, besides involving in the recognition of N-acetyl group of TFN1 *via* another water molecule, W6. Asn166 and Arg186 together interact with —OH on C3′ *via* W2 and W3, respectively, and Glu184 recognizes —OH on C6′ with the help of W4. The two TFNs make contacts between their hydroxyl groups on C6′ and C6 of the Gal moieties through a water molecule, W5. Furthermore, both of the *p*-nitrophenyl groups participate in stacking contacts with each other to enhance their interactions.

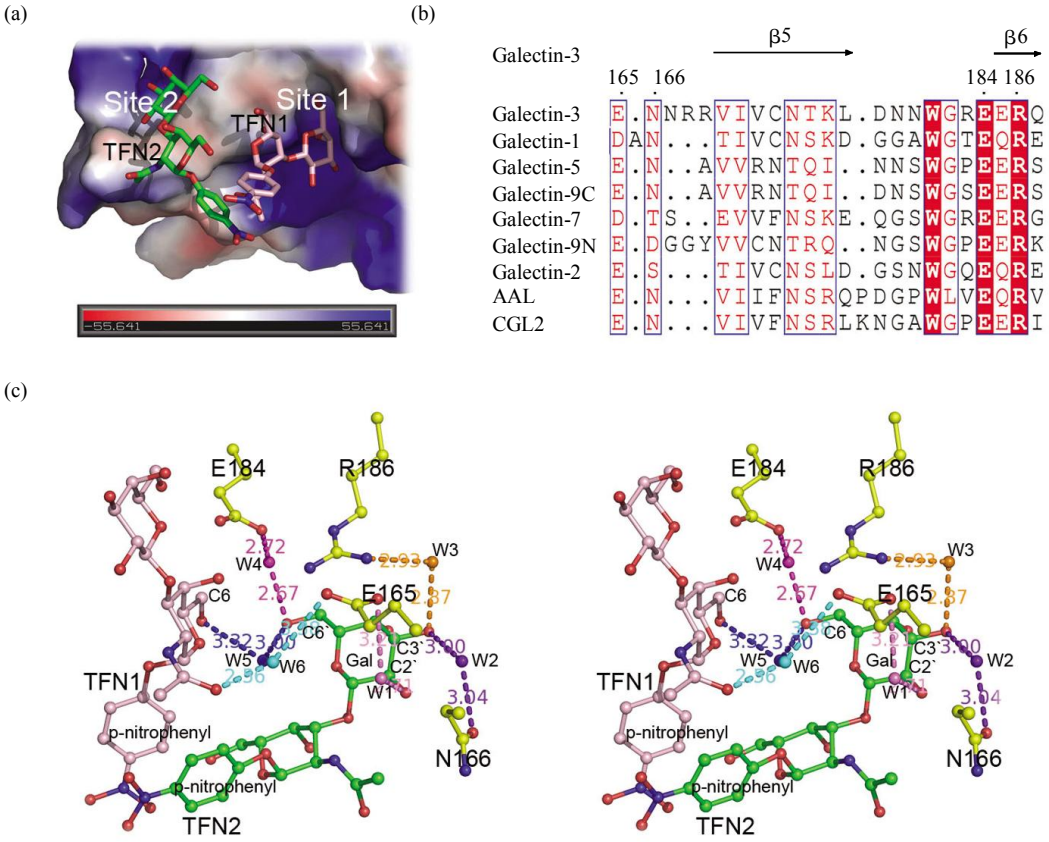

Fig. 2　A view of the molecular surface structure and the stereo diagram illustrating TFN-doublet binding interaction

(a) Surface drawing of the doublet carbohydrate binding sites in Gal-3 CRD. (b) The conservative residues (in red) involved in ligand binding of the different glycans. (c) Stereo drawing of the two hydrogen-bond networks connecting the TFN molecules and CRD domain. TFN1, TFN2 and the residues from Gal-3 CRD are respectively colored in pink, green and yellow. Hydrogen bonds are shown as dashed lines.

2.2 Significance of two binding sites for glycan ligands

To distinguish the ligand binding abilities for two binding sites found in Gal-3 CRD, a mutant E165A was constructed and structurally characterized. E165 plays roles simultaneously for two binding sites of Gal-3 CRD, which interacts with both TFN1 and TFN2 *via* two water molecules. The structure of mutant E165A complexed with TFN ligand was determined at 2.0Å in the same space group, $P3_221$. The final refined R and R_{free} factors are 24% and 28%, respectively. In the complex structure, TFN1 is entirely kept at the binding site commonly found in galectins, which will be called binding Site 1 hereafter, while the TFN2 molecule is completely disappeared in the binding site, which will be called Site 2 (Figure 3). Evidently, the Site 1 is basic and essential for carbohydrate ligand binding, and the Site 2 is additional and auxiliary for the ligand binding.

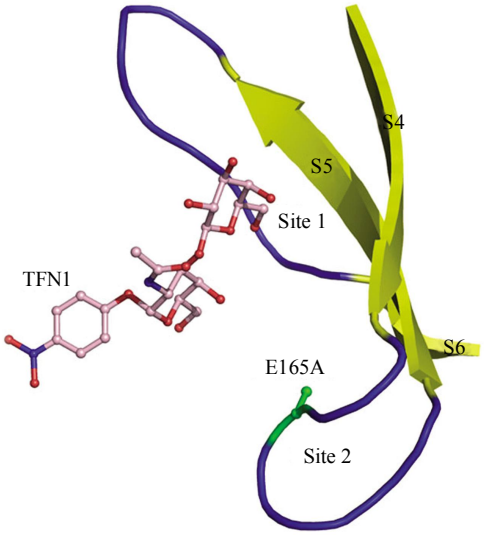

Fig. 3 The CRD domain in the structure of mutant E165 showing the TFN2 molecule is discarded while TFN1 is still appeared at the original binding site 1

Sequence alignments show that the residues critical for ligand binding at Site 2, such as E165, N166, E184 and R186 are in fact conserved in the most galectins known to date (Figure 2b). It implies that the structural elements for the additional binding Site 2 are intrinsically involved in the most galectins, but they could function conditionally. It may be one of the cases that the two-binding-site pattern would be suitable for binding with certain branched saccharides. It is possible that one antennary β-galactoside inserts into the basic Site 1, whereas another may be anchored into the auxiliary Site 2. It has been reported that the binding affinities of galectins to some branched glycans are generally increased compared to monovalent analogues [21]. And from the CFG website(http://www.functionalglycomics.org/glycomics/publicdata/selectedScreens.jsp), we can also find that galectins show relative higher affinities to some branched saccharides.

It is known that the interactions between galectins and glycans are cross-reactions *in vivo* to expand signals through the cascades. In 1994, Bourne *et al.* [22] found two galectin-1 proteins conjugated by the biantennary saccharides of N-acetyllactosamine, which is one kind of cross-linking. CRDs binding more than one saccharide may also result in their cross-linking metastatic cascades, in which the two binding sites maybe behave as amplifiers. For Gal-3, it may speculate that the intrinsically N-terminal domain composed of tandem repeats increases the aggregation to forming oligomers. But the following cascade reactions with ligands may rely on the additional binding site under some conditions, besides the common essential recognition.

Acknowledgements We thank the staff at Beamline 3W1A, BSRF, and Beamline 17U, SSRF, for their assistance during data collections. We also thank HAN Yi, LIU Sheng-Quan of the Institute of Biophysics, The Chinese Academy of Sciences, for their technical supports in the X-ray data collection in house. BIAN Cheng-Feng is really grateful to Dr. HU Yong-Lin, Dr. ZHU De-Yu and Dr. DING Jing-Jin for their valuable helps in this work.

References

[1] Massa S M, Cooper D N, Leffler H, *et al*. L-29, an endogenous lectin, binds to glycoconjugate ligands with positive cooperativity. Biochemistry, 1993, **32**(1): 260−267

[2] Barondes S H, Castronovo V, Cooper D N, *et al*. Galectins: a family of animal beta-galactosidebinding lectins. Cell, 1994, **76**(4): 597−598

[3] Leffler H, Arlsson S, Edlund M, *et al*. Introduction to galectins.

Glycoconj J, 2004, **19**(7−9): 433−440

[4] Lahm H, Andre S, Hoeflich A, *et al*. Tumor galectinology: insights into the complex network of a family of endogenous lectins. Glycoconj J, 2004, **20**(4): 227−238

[5] Houzelstein D, Goncalves I R, Fadden A J, *et al*. Phylogenetic analysis of the vertebrate galectin family. Mol Biol Evol, 2004, **21** (7): 1177−1187

[6] Hirabayashi J, Ayaki H, Soma G, *et al*. Production and purification of a recombinant human 14 ku beta-galactoside-binding lectin. FEBS Letters, 1989, **250**(2): 161−165

[7] Camby I, Mercier M L, Lefranc F, *et al*. Galectin-1: a small protein with major functions. Glycobiology, 2006, **16**(11): 137−157

[8] Liu F T, Patterson R J, Wang J L. Intracellular functions of galectins. Biochim. Biophys Acta, 2002, **1572**(2−3): 263−273

[9] Rabinovich G A, Baum L G, Tinari N, *et al*. Galectins and their ligands: amplifiers, silencers or tuners of the inflammatory response?. Trends Immunol, 2002, **23**(6): 313−320

[10] Perillo N L, Marcus M E, Baum L G. Galectins: versatile modulators of cell adhesion, cell proliferation, and cell death. J Mol Med, 1998, **76**(6): 402−412

[11] Yang R Y, Liu F T. Galectins in cell growth and apoptosis. Cell Mol Life Sci, 2003, **60**(2): 267−276

[12] Kopitz J, von Reitzenstein C, Burchert M, *et al*. Galectin-1 is a major receptor for ganglioside GM1, a product of the growthcontrolling activity of a cell surface ganglioside sialidase, on human neuroblastoma cells in culture. J Biol Chem, 1998, **273**(18): 11205−11211

[13] Sundblad V, Croci D O, Rabinovich G A. Regulated expression of galectin-3, a multifunctional glycan-binding protein, in haematopoietic and non-haematopoietic tissues. Histol Histopathol, 2011, **26**(2): 247−265

[14] Rossmann M G, van Beek C G. Data processing. Acta Crystallogr D Biol Crystallogr, 1999, **55**(10): 1631−1640

[15] Dodson E J, Winn M, Ralph A. Collaborative Computational Project, number 4: providing programs for protein crystallography. Methods Enzymol, 1997, **277**: 620−633

[16] McCoy A J, Grosse-Kunstleve R W, Adams P D, *et al*. Phaser crystallographic software. J Appl Crystallogr, 2007, **40**(4): 658−674

[17] Jones T A, Zou J Y, Cowan S W, *et al*. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A, 1991, **47**(2): 110−119

[18] López-Lucendo M F, Solís D, André S, *et al*. Growth-regulatory human galectin-1: crystallographic characterisation of the structural changes induced by single-site mutations and their impact on the thermodynamics of ligand binding. J Mol Biol, 2004, **343**(4): 957−970

[19] Feng L, Sun H, Zhang Y, *et al*. Structural insights into the recognition mechanism between an antitumor galectin AAL and the thomsen-friedenreich antigen, Faseb J, 2010, **24**(10): 3861−3868

[20] Seetharaman J, Kanigsberg A, Slaaby R, *et al*. X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-Å resolution. J Biol Chem, 1998, **273**(21): 13047−13052

[21] Hirabayashi J, Hashidate T, Arata Y, *et al*. Oligosaccharide specificity of galectins: a search by frontal affinity chromatography. Biochim Biophys Acta, 2002, **1572**(2−3): 232−254

[22] Bourne Y, Bolgiano B, Liao D I, *et al*. Crosslinking of mammalian lectin (galectin-1) by complex biantennary saccharides. Nat Struct Biol, 1994, **1**(12): 863−870

半乳糖凝集素糖结合的新模式：一个糖结合域的双重配体结合 *

卞乘凤[1, 2)]　张　英[1)]　李德峰[1)]　王大成[1)**]

([1)]中国科学院生物物理研究所，生物大分子国家重点实验室，北京 100101；[2)]中国科学院研究生院，北京 100039)

摘要　半乳糖凝集素家族通过糖识别结构域(CRD)可以专一性识别和结合含 β- 半乳糖的多糖配体来发挥其生物学功能．到目前发现的 CRD 对 β- 半乳糖的识别模式是非常保守的，在结构已知的半乳糖凝集素结构中，一个 CRD 只能结合一个多糖配体分子．最近，通过对人源半乳糖凝聚素 -3 CRD 与对硝基 TF 二糖(TFN)复合物的晶体结构解析首次发现，一个 CRD 可以同时结合 2 个 TFN 分子．与这 2 个 TFN 分子有双向结合的残基突变体 E165A 结构分析显示，一个残基的突变引起的结构上的微小变化会使结合位点 2 丧失结合糖底物的能力，而位点 1 的配体结合却不受影响．这表明，结合位点 1 对糖底物保守的识别和结合是基本的、主要的，而结合位点 2 对于糖有条件的结合，是额外的、次要的．序列比对和立体化学分析显示，参与新位点 2 结合的关键残基在其他半乳糖凝集素分子中都是保守的，而它们参与糖配体结合并不常见，表明它们作用的发挥是有条件的．可能在复杂寡聚结构的情况下，如有多重分支结构，双重结合位点将有利于对这类配体分子的辨识和结合，已有一系列研究报道，具有分支结构的寡糖与半乳糖分子的亲和势明显高于单价糖配体，与上述分析相一致．对这类双重位点糖结合的可能生物学意义进行了讨论．

关键词　半乳糖凝集素家族，糖识别结构域(CRD)，β- 半乳糖，对硝基 TF 二糖(TFN)，双重结合模式

学科分类号　Q5　　**DOI**: 10.3724/SP.J.1206.2011.00162

---

* 国家重点基础研究发展计划(973)(30800170)和卫生部重大新药创制专项(2009ZX09103-676)资助项目.

** 通讯联系人.

Tel: 010-64888547, E-mail: dcwang@ibp.ac.cn

收稿日期：2011-04-14，接受日期：2011-04-29