doi:10.1016/j.jmb.2007.09.060 J. Mol. Biol. (2008) 375, 119–135

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Structural Analysis of the Human Galectin-9 N-terminal Carbohydrate Recognition Domain Reveals Unexpected Properties that Differ from the Mouse Orthologue

Masamichi Nagae[1], Nozomu Nishi[2], Sachiko Nakamura-Tsuruta[3] Jun Hirabayashi[3], Soichi Wakatsuki[1] and Ryuichi Kato[1]*

[1]*Structural Biology Research Center, Photon Factory, Institute of Materials Structure Science High Energy Accelerator Research Organization (KEK) 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan*

[2]*Departments of Endocrinology Faculty of Medicine, Kagawa University, 1750-1 Ikenobe Miki-cho, Kita-gun, Kagawa 761-0793, Japan*

[3]*Glycostructure Analysis Team, Research Center for Glycoscience, National Institute of Advanced Industrial Science and Technology, AIST Tsukuba Central 2, 1-1-1 Umezono Tsukuba, Ibaraki 305-8568 Japan*

Received 19 June 2007; received in revised form 7 September 2007; accepted 11 September 2007
Available online 26 September 2007

Edited by I. Wilson

Galectins are a family of β-galactoside-binding lectins that contain a conserved carbohydrate recognition domain (CRD). They exhibit high affinities for small β-galactosides as well as variable binding specificities for complex glycoconjugates. Structural and biochemical analyses of the mechanism governing specific carbohydrate recognition provide a useful template to elucidate the function of these proteins. Here we report the crystal structures of the human galectin-9 N-terminal CRD (NCRD) in the presence of lactose and Forssman pentasaccharide. Mouse galectin-9 NCRD, the structure of which was previously solved by our group, forms a non-canonical dimer in both the crystal state and in solution. Human galectin-9 NCRD, however, exists as a monomer in crystals, despite a high sequence identity to the mouse homologue. Comparative frontal affinity chromatography analysis of the mouse and human galectin-9 NCRDs revealed different carbohydrate binding specificities, with disparate affinities for complex glycoconjugates. Human galectin-9 NCRD exhibited a high affinity for Forssman pentasaccharide; the association constant for mouse galectin-9 NCRD was 100-fold less than that observed for the human protein. The combination of structural data with mutational studies demonstrated that non-conserved amino acid residues on the concave surface were important for determination of target specificities. The human galectin-9 NCRD exhibited greater inhibition of cell proliferation than the mouse NCRD. We discuss the biochemical and structural differences between highly homologous proteins from different species.

Keywords: glycosylation; carbohydrate recognition; lectin; crystal structure; ligand specificity

Introduction

Galectins, members of an animal lectin family, are defined by shared consensus amino acid sequences which confer specific binding to β-galactoside-containing glycoconjugates.[1,2] Galectin families are ubiquitously expressed from lower organisms, such as nematodes and sponges, to higher mammalian species, such as humans.[3] The presence of such protein across many species coupled with the highly conserved amino acid residues which are critical for ligand recognition in the carbohydrate recognition

*Corresponding author. E-mail address: ryuichi.kato@kek.jp.

Abbreviations used: CRD, carbohydrate recognition domain; NCRD, N-terminal CRD; CCRD, C-terminal CRD; Gal, β-D-galactose; Glc, β-D-glucose; GalNAc, *N*-acetyl-D-galactosamine; FAC, frontal affinity chromatography; SPR, surface plasmon resonance; PA, pyridylaminated; pNP, *p*-nitrophenyl; GST, glutathione *S*-transferase.

domain (CRD) suggests that galectins are involved in critical, conserved biological processes.[3,4] Galectin CRDs, which typically consist of approximately 130 amino acid residues tightly folded into a sandwich structure of five- and six-stranded β-sheets, recognizes the basic structure of *N*-acetyllactosamine (LacNAc).[3] Fifteen members of the mammalian galectin family have been identified to date.[5] Hirabayashi & Kasai proposed designating galectin subfamilies as proto-type, chimera-type, or tandem-repeat-type, based on their domain organization.[6] Proto-type galectins (galectins-1, 2, 5, 7, 10, 11, 13, 14, and 15) contain a single CRD with a short N-terminal sequence, while tandem-repeat-type galectins (galectins-4, 6, 8, 9, and 12) include two non-identical CRDs joined by a short linker peptide sequence. The single chimera-type galectin (galectin-3) has one CRD with an extended N terminus containing several repeats of a proline-tyrosine-glycine-rich motif.

Galectin-9, a tandem-repeat-type galectin, was originally isolated from mouse embryonic kidney cells and later found to be widely distributed throughout rat and mouse tissues.[7] In contrast, expression of human galectin-9 is restricted to peripheral blood leukocytes and lymphatic tissues.[8] It has been reported that the potent eosinophil chemoattractant ecalectin, which was originally cloned from a human T cell line, is identical with human galectin-9.[9,10] Several isoforms of mammalian galectin-9 exist, each of which has a linker of various length.[11,12] Galectin-9 exhibits a variety of biological functions, including cell aggregation, eosinophils chemoattraction, and apoptosis of murine thymocytes and T cells and human melanoma cells.[13–16] Mouse galectin-9 induces thymocyte apoptosis in a lactose-inhibitable manner.[14] The chemoattractant activity of galectin-9 depends on its carbohydrate-binding activity and requires both CRDs.[10] Thus, the physiological function of galectin-9 likely depends on its carbohydrate recognition ability.

Two physiological targets for mouse galectin-9 have been reported, Tim-3[17] and GLUT-2.[18] Tim-3 is specifically expressed on the surface of T helper type 1 (T_H1) cells. Galectin-9 recognized the carbohydrate(s) covalently bound to Tim-3; the galectin-9-Tim3 pathway induces cell death in T_H1 cells, suggesting that galectin-9 plays a crucial role in down-regulating T_H1 responses.[17] Galectin-9, however, also interacts on the extracellular surface with GLUT-2, a glucose transporter expressed on the surface of pancreatic β cells that is essential for glucose-stimulated insulin secretion. The recognition of GLUT-2 by galectin-9 through recognition of the carbohydrate moiety is required for the residency of GLUT-2 on the cell surface. Loss of glycosylation or the addition of glycans attenuates the half-life of GLUT-2 on the cell surface and elicits receptor endocytosis with redistribution into endosomes and lysosomes.[18]

Although the exact target carbohydrate structures recognized by galectin-9 *in vivo* have not been identified, *in vitro* analyses have demonstrated that human galectin-9 has a high affinity for multiple oligosaccharides containing β-galactosides.[11,19,20] *In vitro* analyses revealed that the N-terminal CRD (NCRD) and C-terminal CRD (CCRD) of human galectin-9 have different oligosaccharide binding affinities. In comparison to the CCRD of human galectin-9, the NCRD exhibits a striking affinity for complex glycoconjugates, such as Forssman pentasaccharide and polymerized *N*-acetyllactosamine.[11,19] As the biological activity of human galectin-9 will depend on the ligand specificity of each CRD and the subsequent multivalent binding conferred by two CRDs, structural analysis of the specific carbohydrate recognition properties of the human galectin-9 NCRD should provide insight into the physiologic function of galectin-9.

Previously we determined the crystal structures of the mouse galectin-9 NCRD in the absence and presence of carbohydrates.[21] Our results demonstrated that the mouse galectin-9 NCRD forms a unique dimer that differs significantly from the canonical 2-fold symmetric dimer seen for galectins-1 and 2. We report here the crystal structures of the human galectin-9 NCRD complexed to lactose and Forssman pentasaccharide. The human galectin-9 NCRD is related to the mouse galectin-9 NCRD, with 67% overall amino acid sequence identity. Significant local conformational differences, however, are observed between the two crystal structures. Our structural and biochemical results suggest that these differences are integral in both oligomerization and target specificity. We discuss the differences in biochemical properties between the two mammalian orthologues.

Results

Overall structure of human galectin-9 NCRD

We determined the crystal structures of the human galectin-9 NCRD in the presence of two different carbohydrates, lactose and Forssman pentasaccharide. The crystal structure of the human galectin-9 NCRD–lactose complex was determined by molecular replacement using the mouse galectin-9 NCRD structure as a search model. Full-length human galecitn-9 shares 70% amino acid identity with mouse galectin-9, with 67% amino acid sequence identity of the NCRDs (Figure 1). The asymmetric unit of the human galectin-9 NCRD–lactose complex contains three pairs of complex molecules (Figure 2(a)), two of which have clear electron densities for the lactose molecules (molecules A and C), while the other (molecule B) has an ambiguous density map at the carbohydrate binding site. As the average temperature factor of the main-chain atoms in molecule B is higher (30.9 $Å^2$) than those of either molecule A (14.8 $Å^2$) or molecule C (19.1 $Å^2$), the high mobility would affect the quality of the molecule B electron density. The protein molecules of these three complexes are highly similar; the root-mean-square deviation (rmsd) values for all