X-ray Crystal Structure of the Human Galectin-3 Carbohydrate Recognition Domain at 2.1-Å Resolution

(Received for publication, December 9, 1997, and in revised form, February 10, 1998)

J. Seetharaman‡, Amit Kanigsberg‡, Rita Slaaby‡, Hakon Leffler§, Samuel H. Barondes¶, and James M. Rini**

From the ‡Departments of Molecular and Medical Genetics and Biochemistry, University of Toronto, Toronto, Ontario, MSS 1A5, Canada and the §Center for Neurobiology and Psychiatry, Department of Psychiatry, and the ¶Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94179

Galectins are a family of lectins which share similar carbohydrate recognition domains (CRDs) and affinity for small β-galactosides, but which show significant differences in binding specificity for more complex glycoconjugates. We report here the x-ray crystal structure of the human galectin-3 CRD, in complex with lactose and N-acetyllactosamine, at 2.1-Å resolution. This structure represents the first example of a CRD determined from a galectin which does not show the canonical 2-fold symmetric dimer organization. Comparison with the published structures of galectins-1 and -2 provides an explanation for the differences in carbohydrate-binding specificity shown by galectin-3, and for the fact that it fails to form dimers by analogous CRD-CRD interactions.

Galectin-3 is a member of the galectin family of lectins defined by a conserved 14-kDa carbohydrate recognition domain (CRD)1 showing affinity for β-galactosides (1, 2). Abundantly expressed in a few cell types, such as macrophages and polarized epithelial cells in adults (2, 3) and others during embryogenesis (4), it tends to be localized in the cytoplasm and the nucleus. Although functions for galectin-3 have been proposed in each of these subcellular locations (5–7), it is also secreted by a nonclassical pathway (8, 9) and is found on the cell surface and in the extracellular matrix. There it binds and cross-links selected carbohydrate-containing ligands (10, 11) and is thought to modulate cell adhesion (12–14) and cell signaling (15, 16). Many groups are currently studying the roles and uses of galectin-3 in cancer, inflammation, host-pathogen interaction, and nerve injury, among others (17, 18).

Galectin-3 is unique among the known galectins in that, in addition to the canonical CRD (located at the C terminus), it contains an unrelated, non-carbohydrate-binding N-terminal domain of between 120 (in human) and 166 (in dog) amino acids (3, 19). In contrast, galectins-1 and -2 are homodimers composed of the CRD alone, while galectins-4, -6, -8, and -9 possess an N- and C-terminal CRD linked in tandem by a short polypeptide segment (18). The galectin-3 CRD shows sequence identity ranging from 30–40% with galectins-4 through -10, to 20–25% with galectins-1 and -2. It has an affinity for lactose (Lac, $K_d = 1$ mM), and N-acetyllactosamine (LacNAc, $K_d = 0.2$ mM) similar to that of other galectins, but has a distinct profile for larger oligosaccharides (20, 21), including polyN-acetylglucosaminoglycan, a polymer of β(1,3)-linked LacNAc units found on many extracellular matrix and cell surface molecules.

Intact galectin-3, but not its CRD alone, shows avidity for multivalent glycoconjugates (10, 11), modulates cell adhesion (14), and induces intracellular signals (15). Thus it is thought that the N-terminal domain of galectin-3 promotes the formation of dimers or higher order oligomers, thereby permitting multivalent interactions essential for its biological activities.

We report here the x-ray crystal structure of the CRD of human galectin-3 in complex with Lac and LacNAc at 2.1-Å resolution. Previously we and others showed that galectin-1 (22, 23) and galectin-2 (24) are 2-fold symmetric homodimers of the canonical 14-kDa CRD. We now show that, although the galectin-3 CRD is similar to that found in galectin-1 and galectin-2, it displays structural features which provide an explanation for the known differences in galectin-3 carbohydrate-binding specificity and mode of self-association.

EXPERIMENTAL PROCEDURES

Protein Purification and Crystallization—Intact human galectin-3 was expressed in Escherichia coli strain BL21(DE3), followed by purification on a lactosyl-Sepharose column essentially as described previously (11). A C-terminal fragment of human galectin-3 (galectin-3-C; residues 107–250) containing the CRD was then produced by collagenase digestion of the intact molecule followed by purification on a lactosyl-Sepharose column. Galectin-3-C was dialyzed against 50 mM Tris-HCl, pH 7.5, 100 mM MgCl$_2$, and 0.02% NaN$_3$ to remove lactose, and concentrated to 10–15 mg/ml using a Centricon-10 (Amicon, Beverly, MA) ultrafiltration unit. Crystals were grown using the hanging drop vapor diffusion method from drops containing equal volumes of the protein (10–15 mg/ml) in crystallization buffer containing 32–35% PEG 6000, 100 mM Tris-HCl, pH 8.5, and 100–150 mM MgCl$_2$, and 2-mercaptoethanol. The crystals attain dimensions of $0.5 \times 0.3 \times 0.2$ mm$^3$ within 1 to 2 weeks and diffract to approximately 1.9 Å resolution. They grow in space group P2$_1$2$_1$2$_1$ with unit cell dimensions of $a = 86.0$ Å, $b = 58.4$ Å, and $c = 64.0$ Å. There is one molecule in the asymmetric unit resulting in a solvent content of 48%.

Data Collection, Structure Determination, and Refinement—Five heavy atom derivatives were used to calculate the initial phases (see Table 1). With the exception of dimethylmercury, the heavy atom compounds were dissolved in artificial mother liquor containing 32–35% PEG 6000, 100 mM Tris-HCl, pH 8.5, and 100–150 mM MgCl$_2$, at concentrations of 1–3 mM, and soaked into crystals for 2–3 days. The dimethylmercury derivative was obtained through vapor diffusion. After mounting a crystal in a standard glass x-ray capillary, a small
volume (~2 μl) of dimethylmercury was pipetted into the end of the capillary before sealing it with wax and epoxy. The crystal was then allowed to equilibrate for 48 h before data collection. Native and derivative data sets were collected at room temperature on a Siemens multiwire area detector with a conventional rotating anode x-ray source, and reduced using XDS (25). Heavy atom parameter refinement, phase calculations, and solvent flattening were performed using PHASES (26). An initial phase set was calculated using the isomorphous and anomalous components of the derivative data sets (Table I) in conjunction with solvent flattening. The resulting electron density map at 2.5-Å resolution was of very high quality permitting an unambiguous chain tracing. The initial model was built using O (27). Model refinement was performed with X-PLOR (28) using the simulated annealing and conventional energy minimization protocols with $F_{obs} > 1.0(F_{obs})$ between 8.0- and 2.1-Å resolution. Restrained atomic temperature factors were refined using $F_{obs} > 2σ(F_{oobs})$ between 5.0- and 2.1-Å resolution. Water molecules were selected based on difference electron density and hydrogen bond geometry and assigned an occupancy of 0.6 (29). All figures were made with SETOR (30) with the exception of Figs. 4 and 7, which were created with GRASP (31).

**RESULTS AND DISCUSSION**

**Structure Description**—The structure of galectin-3-C has been determined in the presence of both Lac and LacNAc, and in both cases refined at 2.1-Å resolution with good geometry (Table I). Analysis using PROCHECK (32) shows that all non-proline and non-glycine residues are found in the most favored or additionally allowed regions of the Ramachandran plot. In both complexes, Leu-114 is the first residue for which electron density is observed, and hence, the first six residues are presumed to be disordered. Well defined electron density is observed for all other residues, including the C-terminal residue Ile-250. Both complexes are very similar to each other, and all further reference to the structure will pertain to the LacNAc complex unless otherwise indicated.

The CRD (117–250) of galectin-3-C displays an identical topology and very similar three-dimensional structure to that reported for the CRDs of the homodimeric galectins-1 and -2 (22, 24), with which it shares 20–25% sequence identity. Like galectins-1 and -2, it is composed of 5-stranded (F1–F5) and 6-stranded (S1–S6a/S6b) β-sheets which associate in a β-sandwich arrangement (Fig. 1). The Ca atoms of 43 homologous residues in the β-sheet core of galectin-3-C align with root mean square deviations of 0.52 and 0.61 Å with those of human galectin-2 (24) and bovine galectin-1 (22), respectively (Fig. 1). In the canonical dimeric galectins-1 and -2, β-strands F1 and S1 from each monomer extend the antiparallel β-strand interactions across the 2-fold symmetric dimer interface (S1–S1’ and F1–F1’ in Fig. 1), whereas the S1 and F1 β-strands of galectin-3-C form a solvent-exposed surface as discussed below in detail.

**Primary Carbohydrate Binding Site**—As shown in Fig. 1 the Lac/LacNAc binding site is formed by β-strands S4–S6a/S6b. With S3 these β-strands define a carbohydrate-binding cleft, encoded for by a single DNA exon (2, 24), which is evolutionarily conserved among members of the galectin family. The amino acids making direct interaction with the bound carbohydrate are highly conserved among all galectins sequenced to date and are contained on these β-strands. The galactose moiety of Lac/LacNAc is most deeply buried in the binding site (Fig. 2); 166 Å² of its total 230-Å² surface area is buried by the protein. Its C-4 hydroxyl group plays a central role in binding, likely accepting hydrogen bonds from the highly conserved residues His-158 and Arg-162, while donating hydrogen bonds to Asn-160 and W1 (Table II). The galactose C-6 hydroxyl group also displays this cooperative hydrogen bonding pattern (33), interacting with Glu-184, Asn-174, and W3. The planar C-3, C-4, C-5, and C-6 carbon atoms of the
galactose moiety are in van der Waals contact with the aromatic side chain of Trp-181 in a fashion similar to that seen in a number of other galactose and lactose binding lectins (34).

The N-acetylglucosamine (GlcNAc) moiety is more solvent-exposed, with only 91 Å² buried by the protein. Only its C-3 hydroxyl group, which hydrogen bonds to Glu-184 and Arg-162, makes direct hydrogen bonds with the protein (Fig. 2 and Table II). The only other contacts involving the GlcNAc moiety are mediated through its N-acetyl group; the amide proton is hydrogen bonded through water (W2) to Glu-165, and the methyl group makes a van der Waals contact with the guanidino head group of Arg-186 (−20 Å² buried surface). Although there is a water molecule in an analogous position in the lactose complex, which hydrogen bonds to the lactose O-2 hydroxyl group, the hydrogen bond distance is much greater than that found in the LacNAc complex (Table II). The van der Waals interactions involving the 2-position of the glucose/N-acetylgalactosamine moiety represent the only significant differences between the Lac and LacNAc complexes, and presumably account for the approximately 5-fold higher binding affinity, shown by human galectin-3, for N-acetyllactosamine over lactose (20). Interestingly, galectin-1 also shows a water-mediated hydrogen bond involving the NH of the GlcNAc moiety (22), even though there is a difference in the way in which the water molecule is hydrogen bonded to the protein.

The bound galactose and N-acetylglucosamine moieties are very well defined, with average temperature factors of 21 and 32 Å², respectively. The ϕψ values for the β(1,4)-glycosidic linkage of Lac and LacNAc, respectively, are −70°/−103° and

| Atom 1    | Atom 2      | Distance (Å) |
|-----------|-------------|--------------|
| Gal O-3   | W1          | 2.7          |
| W1        | Gal O-4     | 3.0          |
| W1        | Arg144 NH-1 | 2.5          |
| W1        | Asn160 OD-1 | 3.0          |
| Gal O-4   | His158 NE-2 | 2.6          |
| Gal O-4   | Arg162 NH-2 | 3.1          |
| Gal O-4   | Asn160 OD-1 | 3.2          |
| Gal O-5   | Arg162 NH-2 | 3.0          |
| Gal O-6   | Glu184 OE-2 | 2.8          |
| Gal O-6   | Asn174 ND-2 | 2.9          |
| Gal O-6   | W3          | 3.2          |
| W3        | Glu184 OE-2 | 3.1          |
| W3        | W3          | 2.9          |
| Nag/Glc O-6 | W3   | 2.7          |
| Nag/Glc O-3 | Arg162 NH-1 | 3.0          |
| Nag/Glc O-3 | Arg162 NH-2 | 3.0          |
| Nag/Glc O-3 | Glu184 OE-2 | 2.6          |
| Nag N-2   | W2          | 2.8          |
| Glc O-2   | W2          | 3.3          |
| W2        | Glu165 OE-2 | 2.8          |

* W, water molecule.
–68°/–103°, close to the calculated minima for the saccharides in solution (36).

**Extended Ligand Binding Site**—Examination of the solvent-exposed surfaces of galectin-3-C, shows that the Lac/LacNAc binding site, formed by β-strands S4–S6, is a cleft, open at both ends (see Fig. 3). At the nonreducing end (galactose) of the bound carbohydrate the cleft is extended by residues on β-strands S1–S3, whereas at the reducing (GlcNAc) end it is open to the surrounding solution. The carbohydrate binding site is similar in galectin-1 and galectin-2, consistent with the demonstrated ability of galectins-1 and -3 to bind longer oligosaccharides such as polylactosaminoglycans (21, 37–39). These lectins, however, do show differences in affinity for longer oligosaccharides, particularly those substituted on the O-3 of the nonreducing galactose moiety (21, 38). Galectin-3, for example, binds GalNAc1–3(Fuc1–2)Galβ1–4Glc with almost 100-fold higher affinity than does galectin-1 (20, 21). The α-linked GalNAc moiety would be expected to interact with residues in the extended cleft formed by β-strands S1–S3. As shown in Fig. 4, although the identity and conformation of residues involved in binding the Lac/LacNAc moiety in the primary binding site of galectin-1 and galectin-3 are very similar, they differ in the vicinity of the galactose O-3. Galectin-3 has an arginine residue at position 144 which is well positioned to interact with the GalNAc moiety or other saccharide residues linked to the galactose O-3 (Fig. 4). The serine found in galectin-1 is presumably unable to make similar interactions. In addition, the bulky leucine residue at position 31 in galectin-1 is reduced to an alanine (Ala-146) in galectin-3, creating more space for O-3 substituents (Fig. 4).

**Possible Site of Interaction with RNA**—Mouse galectin-3 (formerly known as CBP35) has been shown to be part of the heterogeneous nuclear ribonucleoprotein complex (40, 41) and may be involved in pre-mRNA splicing in vitro (5). In addition, human galectin-3 has recently been shown to directly bind RNA fragments in a gel-shift assay (42). Electrostatic potential calculations (see Fig. 3) show that the carbohydrate binding cleft of galectin-3 is flanked by a linear array of three positively charged arginine residues (Arg-186, Arg-162, and Arg-144). These residues are spaced approximately 5.3 and 7.9 Å apart, close to the phosphate repeat distance in RNA, leading to the possibility that the carbohydrate binding cleft may also be the RNA binding site. Consistent with this suggestion is the fact that the RNA splicing assay is inhibited by soluble oligosaccharides in a rank order reflecting their affinity for galectin-3 (5), even though the interaction of galectin-3 with heterogeneous nuclear ribonucleoprotein particles (41) and RNA fragments (42) appears not to be inhibited by lactose.

**The Homologous Galectin-1/Galectin-2 Dimer Interface**—A striking feature of galectin-3-C is that, in contrast with galectins-1 or -2, it is found to be monomeric in solution at protein concentrations of up to approximately 0.1 mM (11). For this reason it was of particular interest to examine the region in the galectin-3 CRD corresponding to the canonical galectin-1 and

![Fig. 3. Electrostatic potential of the galectin-3-C surface viewed into the carbohydrate binding site.](image1)

![Fig. 4. Stereo view of the carbohydrate binding site of galectin-1 superimposed on that of galectin-3.](image2)
-2 dimer interface. As shown in Figs. 5 and 6 the galectin-1 dimer interface is a very apolar surface composed of residues Leu-4, Ala-6, Leu-9, Phe-133, Val-131, and Ile-128. In the galectin-3 CRD, the apolar nature of this interface has largely been eliminated by a reduction of the F1-S1 β-strand separation and the introduction of Tyr-118 and Tyr-247. In both cases the ring hydroxyl group of the Tyr residues point into solution creating a much more polar exposed surface (Fig. 6). Furthermore, the canonical dimer interface is partially obstructed by residues Leu-114, Ile-115, and Val-116, the end of the galectin-3 N-terminal domain (see Fig. 5). In fact, Val-116 is the start of a conserved tripeptide sequence (V(L)PY) found in galectin-3, galectin-4, and many other galectins, but not found in galectins-1 and -2. It makes a cis-peptide linkage with the highly conserved Pro-117 which is in turn followed typically by a tyrosine residue, in this case Tyr-118, one of the two tyrosine residues responsible for the increased polarity of the canonical dimer interface. Taken together it would appear that the region corresponding to the dimer interface in galectin-1 and galectin-2 does not serve a similar role in galectin-3.

Although intact galectin-3 also migrates as a monomeric (30 kDa) protein by gel filtration at concentrations of up to approximately 0.1 mM, several lines of evidence suggest that it can form dimers or oligomers at higher concentrations, where it binds multivalent ligands better than monovalent ones. The latter properties appear to be largely dependent on the proline/glycine rich N-terminal domain, although interactions with the CRD may also be important (10, 11, 43). In fact, analysis of the galectin-3 CRD (not shown) reveals an apolar patch in the face of the 5-stranded β-sheet which may provide a site for monomer-monomer interactions. Associations involving both the N- and C-terminal domains could easily be achieved by a parallel orientation of monomers, an arrangement seen in the collectins, soluble C-type lectins with an N-terminal collagenous triple helical oligomerization domain (44).

The CRDs of many lectin types have been found in different structural arrangements among members of their respective families. In fact, the precise structural arrangement of CRDs in multimeric lectins appears to be an important means of conferring both specificity and affinity on their interactions with multivalent carbohydrates (34). The fact that galectin-3 does not possess the canonical 2-fold symmetric dimer interface characteristic of galectins-1 and -2 suggests that the galectins are no exception. Moreover, the fact that monomeric galectin-3 is in equilibrium with higher order oligomers under appropri-
conditions, further suggests that in this case the valency is dynamic.

Acknowledgment—We thank Robert Atchison for technical assistance in producing galectin-3.

REFERENCES

1. Barondes, S. H., Castronovo, V., Cooper, D. N., Cummings, R. D., Drickamer, K., Feizi, T., Gitt, M. A., Hirabayashi, J., Hughes, C., Kasai, K., Leffler, H., Liu, F. T., Lotan, R., Mercurio, A. M., Monsigny, M., Pillai, S., Pouyet, F., Raz, A., Rigby, P. W. J., Rini, J. M., and Wang, J. L. (1994) Cell 76, 597–598.

2. Barondes, S. H., Cooper, D. N., Gitt, M. A., and Leffler, H. (1994) J. Biol. Chem. 269, 20807–20810.

3. Barondes, S. H. (1994) Glycobiology 4, 5–12.

4. Fowlis, D., Colnot, C., Ripoche, M. A., and Pouyet, F. (1995) Dev. Dyn. 203, 241–251.

5. Bagher, S. F., Wang, J. L., and Patterson, R. J. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 1213–1217.

6. Goletz, S., Hanisch, F. G., and Karsten, U. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 6737–6742.

7. Yang, R.-Y., Hsu, D. K., and Liu, F.-T. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 1585–1586.

8. Yang, R.-Y., Hsu, D. K., and Liu, F.-T. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 1585–1586.

9. Yang, R.-Y., Hsu, D. K., and Liu, F.-T. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 1585–1586.

10. Hsu, D. K., Zuberi, R. I., and Liu, F. T. (1992) J. Biol. Chem. 267, 15250–15252.

11. Massa, S. M., Cooper, D. N., Leffler, H., and Barondes, S. H. (1993) Biochemistry 32, 260–267.

12. Inohara, H., and Raz, A. (1995) Cancer Res. 55, 3267–3271.

13. Bao, Q., and Hughes, R. C. (1995) J. Cell Sci. 108, 2791–2800.

14. Kuwabara, I., and Liu, F. T. (1996) J. Immunol. 156, 3839–3844.

15. Yamazaki, A., Kuwabara, I., Frigeri, L. G., and Liu, F. T. (1995) J. Immunol. 154, 3479–3487.

16. Zuberi, R. I., Frigeri, L. G., and Liu, F. T. (1994) Cell. Immunol. 156, 1–12.

17. Akahani, S., Inohara, H., Nangia-Makker, P., and Raz, A. (1997) Trends Glycosci. Glycotechnol. 9, 69–75.

18. Leffler, H. (1997) Trends Glycosci. Glycotechnol. 9, 9–19.

19. Herrmann, J., Turek, C. W., Atchison, B. E., Hufjej, M. E., Poulter, L., Gitt, M. A., Burlingame, A. L., Barondes, S. H., and Leffler, H. (1993) J. Biol. Chem. 268, 26704–26711.

20. Sparrow, C. P., Leffler, H., and Barondes, S. H. (1987) J. Biol. Chem. 262, 7833–7839.

21. Leffler, H., and Barondes, S. H. (1986) J. Biol. Chem. 261, 10119–10126.

22. Liao, D. I., Kapadia, G., Ahmed, H., Vasta, G. R., and Herzberg, O. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 1428–1432.

23. Bourne, Y., Bolgiano, B., Liao, D. I., Strecker, G., Cantau, P., Herzberg, O., Feizi, T., and Cambillau, C. (1994) Nat. Struct. Biol. 1, 863–870.

24. Lobsanov, Y. D., Gitt, M. A., Leffler, H., Barondes, S. H., and Rini, J. M. (1995) J. Biol. Chem. 269, 27034–27038.

25. Kabach, W. (1988) J. Appl. Crystallogr. 21, 916–924.

26. Furey, W., and Swamianathan, S. (1997) Methods Enzymol. 277, 590–620.

27. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Acta Crystallogr. A 47, 110–119.

28. Brunger, A. (1992) X-PLOR, v3.1 A system for x-ray crystallography and NMR, Yale University Press, New Haven, CT.

29. Kundrot, C. E., and Richards, F. M. (1987) Acta Crystallogr. Sec. B 43, 544–547.

30. Evans, S. V. (1993) J. Mol. Graph. 11, 134–138, 127–138.

31. Nicholls, A., Sharp, K. A., and Honig, B. (1991) Proteins 11, 281–296.

32. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) J. Appl. Crystallogr. 26, 283–291.

33. Quiocho, F. A. (1988) Curr. Top. Microbiol. Immunol. 139, 135–148.

34. Rini, J. M. (1995) Annu. Rev. Biophys. Biomol. Struct. 24, 551–577.

35. Abbott, W. M., Hounsell, E. F., and Feizi, T. (1988) Biochem. J. 252, 283–287.

36. Imberty, A., Delage, M. M., Bourne, Y., Cambillau, C., and Perez, S. (1991) Glycoconjug. J. 8, 456–483.