Version: 17 September 2013

For: *Biomol. NMR Assign.*

¹H, ¹³C, and ¹⁵N backbone and side-chain chemical shift assignments for the 29 kDa human chimera-type galectin-3, an adhesion/growth-regulatory lectin

Hans Ippel^{1,2}, Michelle C. Miller¹, Manuel Alvaro Berbís³, Dennis Suylen²,
Sabine André⁴, Tilman Hackeng², F. Javier Cañada³, Christian Weber^{2,5},
Hans-Joachim Gabius⁴, Jesús Jiménez-Barbero³, Kevin H. Mayo¹

¹Departments of Biochemistry, Molecular Biology & Biophysics,
University of Minnesota, Minneapolis, Minnesota USA; ²Department of Biochemistry
and CARIM, University of Maastricht, Maastricht, The Netherlands,
³Chemical and Physical Biology Department, Centro de Investigaciones Biológicas,
CSIC, Madrid, Spain, and ⁴Institute of Physiological Chemistry of the Faculty of
Veterinary Medicine and ⁵Institute for Cardiovascular Prevention,
Ludwig-Maximilians-University, Munich, Germany

Running head: ¹H, ¹³C, and ¹⁵N backbone and side-chain chemical shift for Gal-3

Keywords: apoptosis, galectin, glycan, proliferation, signaling

Abbreviations used: Gal-3, human full-length galectin-3

Address correspondence to: Kevin H. Mayo, Dept. of Biochemistry, 6-155 Jackson Hall,
University of Minnesota, 321 Church Street, Minneapolis, Minnesota 55455. Tel: 612-
625-9968; Fax: 612-624-5121; Email: mayox001@umn.edu

Abstract

Galectin-3 has a unique trimodular design consisting of the canonical lectin domain, a collagen-like tandem-repeat section and an N-terminal peptide with two sites for Ser phosphorylation. Structural characterization of the full-length protein with its non-lectin part (115 of 250 residues) will help understand the multifunctionality of this potent cellular effector. Here, we report ^1H , ^{13}C , and ^{15}N chemical shift assignments as determined by heteronuclear NMR spectroscopy.

Biological Context

Complex carbohydrates on cell surfaces are increasingly viewed as biochemical messages read and translated into effects by endogenous lectins (Gabius et al. 2011). Among them, the members of the galectin family mediate adhesion and regulate cell growth (as mitogen or as anti/pro-apoptotic effector), attracting special interest due to their context-dependent multifunctionality and involvement in disease processes (Kaltner and Gabius 2012). Galectin-3 (Gal-3, also known as MAC-2 antigen) is unique in this family with its chimera-type design, meaning that the canonical carbohydrate recognition domain (CRD) is linked to a collagen-like section (with nine Pro/Gly-rich repeats in the human protein) and an N-terminal peptide with two sites for Ser phosphorylation (Hughes 1994). Its structural analysis in solution has been initiated by assignment of ^1H , ^{13}C and ^{15}N resonances of the truncated Gal-3 CRD (Umemoto and Leffler 2001). However, emerging evidence on the broad physiological significance of the full-length protein (with the other two structural parts), well beyond glycan binding, has emphasized the need for NMR assignment of the full-length protein.

Furthermore, the application of the full-length protein as a histochemical tool has revealed robust intracellular staining, supporting the steadily growing evidence of binding partners for Gal-3 other than glycans, for example in growth regulation such as bcl-2 or in pre-mRNA splicing (Haudek et al. 2010; Dawson et al. 2013). Structural analysis of such complexes will only be possible with complete resonance assignment information on Gal-3. The N-terminal tail is also a key factor for the markedly high affinity of Gal-3 to glycan counter receptors clustered in microdomains via aggregation and as a contact site for bacterial lipopolysaccharide (Kopitz et al. 2010; Quattroni et al. 2012). Following phosphorylation of Ser5 (and Ser11), a signal for nuclear export (Haudek et al. 2010), the lectin strongly promotes axon branching, suggesting the presence of intra-protein interactions between the N-terminal tail and CRD in counterreceptor selection, complex stability and routing (Díez-Revuelta et al. 2010). In addition, Tyr

phosphorylation by non-receptor kinases cAbl/Arg (at Tyr79, 107 and 118) further underscores the potential of the N-terminal tail, without and with phosphorylation, to contribute to Gal-3 functionality (Balan et al. 2010; Li et al. 2010). To engage in thorough NMR studies regarding the impact of phosphorylation on Gal-3 structure and function, as well as various counter-receptor interactions, we present here ^1H , ^{13}C and ^{15}N resonance assignments for full length Gal-3. This information also enables detailed comparisons of this chimera-type galectin to proto-type galectins-1 and -7, for which we have previously reported heteronuclear resonance assignments (Nesmelova et al. 2008, 2012).

Experimental Procedures

[^{15}N , ^{13}C]-labeled Gal-3 and peptide synthesis

Recombinant expression of uniformly [^{13}C , ^{15}N]-labeled Gal-3 in commercial, isotopically enriched medium led to a yield of about 40 mg protein per 100 mL stock solution. Purity of the lectin was rigorously ascertained by one- and two-dimensional electrophoresis, gel filtration, and nano-electrospray ionization mass spectrometry (ESI MS), as well as by activity measurements for hemagglutination, cell surface binding, and cell proliferation (Kopitz et al. 2003; Amano et al. 2012). Two non-isotopically-enriched peptides derived from the N-terminal tail sequence of Gal-3 (residues 1-50 and residues 51-107) were synthesized using solution-phase tBOC peptide chemistry as described previously (Hackeng et al. 2001). HPLC and mass spectrometry indicated that purity was > 95%.

NMR spectroscopy

NMR samples of Gal-3 were prepared in 20 mM potassium phosphate buffer at pH 6.8. NMR spectra were acquired at 30 °C, using Bruker AVANCE 700 or 900 MHz spectrometers equipped either with triple resonance CryoProbe or TCI cryoprobe. Assignments of backbone and side-chain resonances were done based on HNCA, HNcoCA, HNCACB, CBCAcoNH, HNCO, HNcaCO, HCCH-TOCSY, ^{15}N -separated NOESY-HSQC, and ^{13}C -separated NOESY-HSQC spectra (Muhandiram and Kay 1994), essentially as described and performed previously for galectins-1 and -7 (Nesmelova et al. 2008; 2012). In addition, ^{13}C -detected 2D C_CON, C_CACO and C_CAN spectra were used to aid in the sequential assignment of proline residues in the N-terminal region (Bermel et al. 2008).

The two peptides (residues 1-50 and residues 51-107) were prepared in 20 mM potassium phosphate solution adjusted to pH 3 in order to maximize NH resonance intensity. NMR spectra were acquired at 30 °C using a Bruker AVANCE 700 MHz spectrometer equipped with a triple resonance

CryoProbe. Assignments of backbone and side-chain resonances were made based on ^1H DIPSII and NOESY spectra as well as on natural abundance ^{13}C , ^1H -HSQC, ^{15}N , ^1H -HSQC and ^{13}C -HSQC-DIPSII spectra. NMR data were processed with the program NMRPipe (Delaglio et al. 1995) and analyzed by using the program CARA (Keller 2004).

Assignments and Data Deposition

Gal-3 can be essentially separated into two distinct parts: CRD and N-terminal tail. These structurally distinct elements are reflected in the ^{15}N , ^1H -HSQC spectrum shown in Figure 1. Resonances from the CRD are relatively well dispersed for the canonical β -sheet structure as previously reported for truncated Gal-3 CRD (Umemoto and Leffler 2001), whereas those arising from the N-terminal tail exhibit rather narrow line widths, reflecting comparatively increased internal mobility. In addition, resonances arising from the N-terminal tail are minimally dispersed and fall within the spectral region associated with random coil conformation. This region of the HSQC has been expanded in Figure 1B for better clarity. Overall, such a structure is in line with the occurrence of highly charged ions in the bimodal charge distribution of Gal-3 noted from nano-ESI MS (Kopitz et al. 2003). Also, the different degrees of organization can well underlie the two transition temperatures (T_m) at 56 °C/~39 °C in differential scanning calorimetry (DSC) of murine Gal-3 (Agrwal et al. 1993). Our NMR diffusion measurements (not shown) indicate that full length Gal-3 is primarily monomeric at the concentration of 400 μM used to perform our heteronuclear NMR experiments and the resonance assignments reported here. Nevertheless, the line width of proton signals within the CRD domain improved considerably at protein concentrations < 40 μM , suggesting some tendency of the CRD in Gal-3 to associate in solution.

Backbone resonances (HN, N, C, $\text{C}\alpha$, and $\text{H}\alpha$) for all residues within the CRD and N-terminal tail were fully assigned, with the exception of the fast exchangeable HN amides H64, G93, G152, N153, N166, S188, and K226, as well as CO of residues preceding prolines in the CRD. The major difficulty with assigning these resonances lay in the presence of multiple Pro residues (27 *in toto*) in the tandem-repeat section of the tail. This relatively large number of prolines also gave rise to the presence of multiple minor resonances associated with *cis*-peptide bonds in slow exchange on the chemical shift time scale with their respective *trans*-peptide isomers (Poznański et al. 1993). These inherent ambiguities notwithstanding, assignment of the major resonances of prolines in the *trans*-linkage was possible and have been reported here. In detail, the Pro-derived resonances from the CRD were readily assignable, because their

neighbouring residues were unique in the amino acid sequence. Within the N-terminal tail, however, making specific assignments of prolines and their neighbouring residues was more challenging, primarily because of the tandem-repeat nature of the collagen-like peptide segments, e.g. PGAYP segments and the multiple PG motifs.

For this reason, we synthesized two peptides (residues 1-50 and residues 51-107) covering the N-terminal tail to aid in the assignment process. This approach worked well because of the random coil and dynamic nature of the tail within Gal-3. Resonances assigned in these peptides could properly be compared with resonances of the same residues in Gal-3, even for the minor cis-proline peaks, many of which were simply confirmed using these peptides. Residues within PGAYP sequences were difficult to uniquely assign, because most resonances have near-identical proton and backbone carbon chemical shifts that fall within 0.1 ppm of each other. E.g. the segment $^{67}\text{PGAY}^{70}$ could not be uniquely assigned. Similarly, $^{42}\text{PGAYPGQAPP}^{51}$ has the same amino acid sequence as segment 51-60, leading to indistinguishable chemical shifts for resonances of each equivalent amino acid residue. Sequential assignments of other difficult proline-containing segments could however be resolved by recording additional ^{13}C -detected correlations between carbonyl/ $\text{C}\alpha$ and proline nitrogen atoms. Overall, greater than 90% of side-chain resonances were assigned in both the CRD and N-terminal tail. Chemical shift assignments have been deposited in the BioMagRes-Bank (<http://www.bmrb.wisc.edu>) under accession number 19491.

Analysis of the secondary structure based on the $^1\text{H}\alpha$, $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$ and ^{15}NH chemical shift index (CSI) (Wishart and Sykes 1994) is shown in Figure 2, with specific sequences classified as α -helix, β -strand, random coil (loops and turns), and poly-proline as indicated. Sequences within the CRD of full length Gal-3 have a predicted consensus for β -strand secondary structure interspersed with segments of random coil (loops and turns). The random coil curve within the CRD amino acid sequence has been omitted for clarity in Figure 2. These secondary structure predictions show a high degree of similarity with structural elements observed in the crystal structure of the truncated Gal-3 CRD (Seetharaman et al. 1998). Our NMR data also confirm the presence of a cis-conformation for P117 that was previously observed in the crystal structure of the CRD (Saraboji et al. 2012). For the N-terminal tail, CSI analysis predicts predominantly random coil conformation, but with some potential for formation of β -strand and/or poly-proline structure within residues 43-60 and 78-85. In fact, a certain degree of secondary structure within

this segment would explain the increased T_m -value (~ 48 °C) in DSC experiments on the separately purified N-terminal tail (Agrwal et al. 1993).

Acknowledgments. This work was generously supported by a research grant from the National Cancer Institute (CA-096090) to KHM; EC funding (GlycoHIT, contract no. 260600; GLYCOPHARM, contract no. 217297) to JJB & HJG, and European Research Council funding (ERC AdG 249929) to CW. NMR instrumentation was provided with funds from the National Science Foundation (BIR-961477), the University of Minnesota Medical School, and the Minnesota Medical Foundation.

References

- Agrwal N, Sun Q, Wang S-Y, Wang JL (1993) Carbohydrate-binding protein 35. I. Properties of the recombinant protein and the individuality of the domains. *J Biol Chem* 268:14932-14939
- Amano M, Eriksson H, Manning JC, Detjen KM, André S, Nishimura S-I, Lehtiö J, Gabius H-J (2012) Tumour suppressor p16^{INK4a}: anoikis-favouring decrease in N/O-glycan/cell surface sialylation by down-regulation of enzymes in sialic acid biosynthesis in tandem in a pancreatic carcinoma model. *FEBS J* 279:4062-4080
- Balan V, Nangia-Makker P, Jung YS, Wang Y, Raz A (2010) Galectin-3: a novel substrate for c-Abl kinase. *Biochim Biophys Acta* 1803:1198-1205
- Camilloni C, De Simone A, Vranken W, Vendruscolo M (2012) Determination of Secondary Structure Populations in Disordered States of Proteins using NMR Chemical Shifts. *Biochemistry* 51:2224-2231.
- Dawson H, André S, Karamitopoulou E, Zlobec I, Gabius H-J (2013) The growing galectin network in colon cancer and clinical relevance of cytoplasmic galectin-3 reactivity. *Anticancer Res* 33:3053-3059
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) Nmrpipe: a multidimensional spectral processing system based on unix pipes. *J Biomol NMR* 6:277-293
- Díez-Revuelta N, Velasco S, André S, Kaltner H, Kübler D, Gabius H-J, Abad-Rodríguez J (2010) Phosphorylation of adhesion- and growth-regulatory human galectin-3 leads to the induction of axonal branching by local membrane L1 and ERM redistribution. *J Cell Sci* 123:671-681
- Gabius H-J, André S, Jiménez-Barbero J, Romero A, Solís D (2011) From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem Sci* 36:298-313
- Hackeng TM, Rosing J, Spronk HMH, Vermeer C (2001) Total chemical synthesis of human matrix Gla protein. *Protein Sci* 10:864-870
- Haudek KC, Spronk KJ, Voss PG, Patterson RJ, Wang JL, Arnoys EJ (2010) Dynamics of galectin-3 in the nucleus and cytoplasm. *Biochim Biophys Acta* 1800:181-189
- Hughes RC (1994) Mac-2: a versatile galactose-binding protein of mammalian tissues. *Glycobiology* 4:5-12
- Kaltner H, Gabius H-J (2012) A toolbox of lectins for translating the sugar code: the galectin network in phylogenesis and tumors. *Histol Histopathol* 27:397-416
- Keller RLJ (2004) The computer-aided resonance assignment tutorial. (Cantina Verlag)
- Kopitz J, Bergmann M, Gabius H-J (2010) How adhesion/growth-regulatory galectins-1 and -3 attain cell specificity: case study defining their target on neuroblastoma cells (SK-N-MC) and marked affinity regulation by affecting microdomain organization of the membrane. *IUBMB Life* 62:624-628

Kopitz J, André S, von Reitzenstein C, Versluis K, Kaltner H, Pieters RJ, Wasano K, Kuwabara I, Liu FT, Cantz M, Heck AJ, Gabius HJ (2003) Homodimeric galectin-7 (p53-induced gene 1) is a negative growth regulator for human neuroblastoma cells. *Oncogene* 22:6277-6288

Li X, Ma Q, Wang J, Liu X, Yang Y, Zhao H, Wang Y, Jin Y, Zeng J, Li J, Song L, Li X, Li P, Qian X, Cao C (2010) c-Abl and Arg tyrosine kinases regulate lysosomal degradation of the oncoprotein galectin-3. *Cell Death Differ* 17:1277-1287

Muhandiram DR, Kay LE (1994) Gradient-enhanced triple-resonance 3-dimensional NMR experiments with improved sensitivity. *J Magn Reson Ser B* 103:203-216

Nesmelova IV, Pang M, Baum LG, Mayo KH (2008) ^1H , ^{13}C , and ^{15}N backbone and side-chain chemical shift assignments for the 29 kDa human galectin-1 protein dimer. *Biomol NMR Assign* 2:203-205

Nesmelova IV, Berbís MÁ, Miller MC, Cañada FJ, André S, Jiménez-Barbero J, Gabius H-J, Mayo KH (2012) ^1H , ^{13}C , and ^{15}N backbone and side-chain chemical shift assignments for the 31 kDa human galectin-7 (p53-induced gene 1) homodimer, a pro-apoptotic lectin. *Biomol NMR Assign* 6:127-129

Poznański J, Ejchart A, Wierzchowski KL, Ciurak M. (1993) ^1H and ^{13}C -NMR investigations on *cis-trans* isomerization of proline peptide bonds and conformation of aromatic side chains in H-Trp-(Pro)_n-Tyr-OH peptides. *Biopolymers* 33:781-795

Quattroni P, Li Y, Lucchesi D, Lucas S, Hood DW, Herrmann M, Gabius H-J, Tang CM, Exley RM (2012) Galectin-3 binds *Neisseria meningitidis* and increases interaction with phagocytic cells. *Cell Microbiol* 14:1657-1675

Saraboji K, Hakansson M, Genheden S, Diehl C, Qvist J, Weininger U, Nilsson UJ, Leffler H, Ryde U, Akke M, Logan DT (2012) The carbohydrate-binding site in galectin-3 is preorganized to recognize a sugar-like framework of oxygens: ultra-high-resolution structures and water dynamics. *Biochem* 51:296-306

Seetharaman J, Kanigsberg A, Slaaby R, Leffler H, Barondes SH, Rini JM (1998) X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-Å resolution. *J Biol Chem* 273:13047-13052

Umemoto K, Leffler H (2001) Assignment of ^1H , ^{15}N and ^{13}C resonances of the carbohydrate recognition domain of human galectin-3. *J Biomol NMR* 20:91-92

Wishart DS, Sykes BD (1994) The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J Biomol NMR* 4:171-180

Wolfgang Bermel, Isabella C. Felli, Rainer Kümmerle, Roberta Pierattelli (2008) ^{13}C direct-detection biomolecular NMR. *Concepts in Magnetic Resonance Part A* 32:183-200

Figure Legends

Figure 1. (A) ^{15}N - ^1H HSQC 700 MHz spectrum recorded for 20 μM Gal-3 in 20 mM potassium phosphate buffer, 8 mM DTT, 0.1 mM EDTA (pH 6.8), acquired at 30°C is shown. Cross peaks are labeled with specific assignments. Protons labeled “H” stand for amide protons, and “HE” stands for Trp HE or Arg side-chain HE protons. HD21/HD22 and HE21/HE22 correspond to side chain amino protons of Asn and Gln, respectively. Lower case letters b, c, d, and e correspond to minor peaks induced by the cis-proline isomer state. Normally, these minor peaks are observed only for sequentially proximal proline residues within the N-terminal tail of Gal-3. However, through-space conformational effects resulting from the presence of cis-prolines can also result in resonance doubling within the CRD, e.g. HN of L135 and HN of S244. (B) Expansion of the more crowded part of the HSQC spectrum (boxed-in in panel A) is shown. This expansion contains primarily backbone amide and side chain amino resonances of the N-terminal tail.

Figure 2. Calculation of the secondary structure of Gal-3 was performed by using the $\delta 2\text{D}$ software program (Camilloni et al. 2012). The population of secondary structure (helix, β -strand, coil, poly-proline II) in $\delta 2\text{D}$ is based on the chemical shift index (CSI) consensus (Wishart and Sykes 1994) of all six experimentally determined $^1\text{H}\alpha$, $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, ^{13}CO and $^{15}\text{N}/^1\text{H}$ chemical shifts of residues in Gal-3. P117 has been defined as cis O117 in the $\delta 2\text{D}$ calculation. To avoid confusion from overlap of beta and coil lines within the CRD, the line for the coil population has been truncated at residue 115 in this plot. Therefore, the coil line is only shown through the N-terminal tail segment.