The intrinsically disordered N-terminal domain of galectin-3 dynamically mediates multisite self-association of the protein through fuzzy interactions

Received for publication, June 18, 2017, and in revised form, August 23, 2017 Published, Papers in Press, September 11, 2017, DOI 10.1074/jbc.M117.802793

Yu-Hao Lin[‡], De-Chen Qiu[‡], Wen-Han Chang[‡], Yi-Qi Yeh[§], U-Ser Jeng[§¶], Fu-Tong Liu[‖], and Jie-rong Huang[‡1]**

*From the [‡]Institute of Biochemistry and Molecular Biology and the **Institute of Biomedical Informatics, National Yang-Ming University, Number 155 Section 2 Li-nong Street, Taipei 11221, Taiwan, the [§]National Synchrotron Radiation Research Center, Hsinchu 30076, Taiwan, the [¶]Department of Chemical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan, and the [‖]Institute of Biomedical Sciences, Academia Sinica, Taipei 11529, Taiwan*

Edited by Wolfgang Peti

Galectins are a family of lectins that bind β-galactosides through their conserved carbohydrate recognition domain (CRD) and can induce aggregation with glycoproteins or glycolipids on the cell surface and thereby regulate cell activation, migration, adhesion, and signaling. Galectin-3 has an intrinsically disordered N-terminal domain and a canonical CRD. Unlike the other 14 known galectins in mammalian cells, which have dimeric or tandem-repeated CRDs enabling multivalency for various functions, galectin-3 is monomeric, and its functional multivalency therefore is somewhat of a mystery. Here, we used NMR spectroscopy, mutagenesis, small-angle X-ray scattering, and computational modeling to study the self-association–related multivalency of galectin-3 at the residue-specific level. We show that the disordered N-terminal domain (residues ~20–100) interacts with itself and with a part of the CRD not involved in carbohydrate recognition (β-strands 7–9; residues ~200–220), forming a fuzzy complex via inter- and intramolecular interactions, mainly through hydrophobicity. These fuzzy interactions are characteristic of intrinsically disordered proteins to achieve liquid–liquid phase separation, and we demonstrated that galectin-3 can also undergo liquid–liquid phase separation. We propose that galectin-3 may achieve multivalency through this multisite self-association mechanism facilitated by fuzzy interactions.

Galectins are a family of lectins that bind to β-galactoside through their conserved carbohydrate recognition domain (CRD)[2] (1). Extracellularly, they can induce aggregation of or form lattices with glycoproteins or glycolipids on the cell surface and thus regulate cell activation, migration, adhesion, and signaling (1). Fourteen of the 15 known galectins have dimeric or tandem-repeated CRDs; this multivalency facilitates the above-described functions (1–3). Galectin-3 is the only chimeric member of this family (4), with its CRD being connected to its non-lectin N-terminal domain (NTD; residues 1–112, Fig. 1, *A–C*). This intrinsically disordered NTD is essential for the biological activity of galectin-3 (5–8). On the basis of the protein's co-precipitation ratio with pentasaccharides, it was reported that galectin-3 forms pentamers cross-linked through the NTD (9). Galectin-3, however, is mainly found in its monomeric form in solution (5, 6, 10, 11). How this protein becomes a functional oligomer is still under debate.

Depending on which domain is involved, the interactions are categorized as type-N (6, 9, 12–14) or type-C (15–18) for the NTD- or CRD-mediated self-associations. In most studies published to date, saccharide or asialofetuin has been used to induce self-association; indirect assays, such as chemical cross-linking (5, 19), fluorescence anisotropy (16), and solid-phase radioligand-binding assay (15), have also been used to study galectin-3 self-association. On the other hand, NMR spectroscopy provides residue-specific information and can detect a protein's intrinsic self-association propensity in the absence of additional ligands. Birdsall *et al.* (20) showed evidence of NTD and CRD interaction in an early NMR study. More recently, Mayo and co-workers (21) assigned the backbone chemical shifts of galectin-3 and demonstrated detailed studies of this protein (18, 22). They identified a transient α-helix in the NTD (residues 5–15) and showed that the PGA*X* motif is the epitope interacting with the CRD (by titrating synthetic peptides with the CRD alone). They also observed a concentration-dependent NMR chemical shift in the CRD part and concluded that the self-association is a type-C interaction. Their study also showed, however, that no self-association was observed in the absence of the NTD based on the results of the molecular diffusion coefficients. How the NTD mediates galectin-3 self-association is thus still unclear.

Here we performed NMR dynamics and paramagnetic relaxation enhancement (PRE) studies of more than 15 galectin-3 variants to depict a detailed model of how this intrinsically disordered NTD mediates its self-association. We found that the

This work was supported by Ministry of Science and Technology of Taiwan Grants 104-2113-M-010-001-MY2 and 106-2113-M-010-005-MY2 (to J. H.) and 106-2321-B-001-042 (to F.-T. L.). The authors declare that they have no conflicts of interest with the contents of this article.

This article contains supplemental Tables S1–S3 and Figs. S1–S9.

[1] To whom correspondence should be addressed: Institute of Biochemistry and Molecular Biology, National Yang-Ming University, No. 155 Section 2 Li-nong St., Taipei 11221, Taiwan. Tel.: 886-2-2826-7258; E-mail: jierongh@ym.edu.tw.

[2] The abbreviations used are: CRD or C, carbohydrate recognition domain; NTD or N, N-terminal domain; PRE, paramagnetic relaxation enhancement; hetNOE, heteronuclear NOE; HSQC, heteronuclear single quantum coherence; MTSL, (1-oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) methanethiosulfonate; IMAC, immobilized metal-ion affinity chromatography; SUMO, small ubiquitin-like modifier; SAXS, small-angle X-ray scattering.

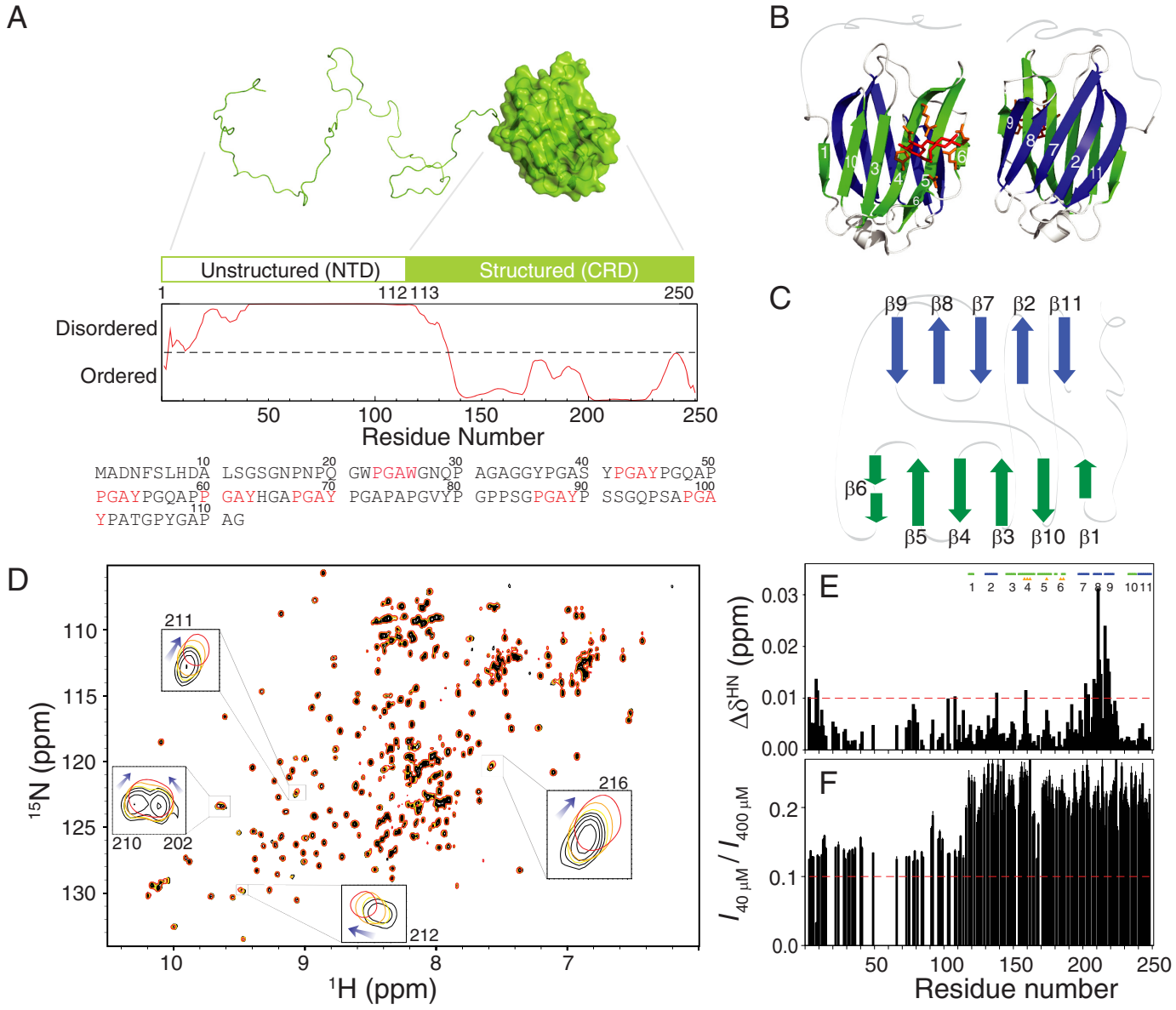

Figure 1. The structure of galectin-3 and its self-association ability. *A*, schematic representation of the structured and intrinsically disordered domains of galectin-3 (disorder predicted by IUPRED (79)). The amino acid sequence of the NTD is listed, and the PGA*X* motifs are shown in *red*. *B*, *ribbon model* of the structured CRD. The carbohydrate-binding side chains that interact with the ligand (in *red*) are shown in *orange*. *C*, topological connection of the β-strands in the CRD. *D*, overlay of HSQC spectra of 40 (*black*), 100 (*orange*), 200 (*yellow*), and 400 μM (*red*) samples of galectin-3. The clearest changes in chemical shifts (for residues 202, 210, 211, 212, and 216) are highlighted. Shown are chemical shift differences (*E*) and ratios of peak intensities (*F*) between 40 and 400 μM samples as a function of residue number.

NTD and the CRD interact in a "fuzzy" complex manner inter- and intramolecularly through both NTD–NTD and NTD–CRD interactions.

Results

The self-association of galectin-3 is concentration-dependent

We investigated galectin-3 self-association at a residue-specific level by comparing the ^{15}N-^{1}H HSQC spectra of 40–400 μM protein samples. Several cross-peaks moved as the protein concentration was increased (Fig. 1*D*). The largest changes occurred for peaks assigned to residues ~200–220, in β-strands 7, 8, and 9 (Fig. 1*E*), which are on the reverse side with respect to the carbohydrate-binding site (Fig. 1*B*). Accounting for the change in concentration alone, the peak intensity ratio between the 40 and 400 μM spectra should be ~0.1, but the mean ratios for the NTD and CRD are ~0.13 and ~0.25, respectively (Fig. 1*F*). This is because the corresponding peaks are weaker and/or broader in the spectrum of the higher concentration sample. The chemical shift changes and intensity ratios suggest that galectin-3 self-associates at higher protein concentrations, in agreement with previous observations (18).

NMR dynamics studies confirm the self-association

We used NMR spin relaxation experiments, which are widely used to study the dynamical properties of biomolecules, to further characterize the self-association. The overall tumbling rate, diffusion tensor, and internal motions of a molecule in the pico- to nanosecond timescale can be derived from the results

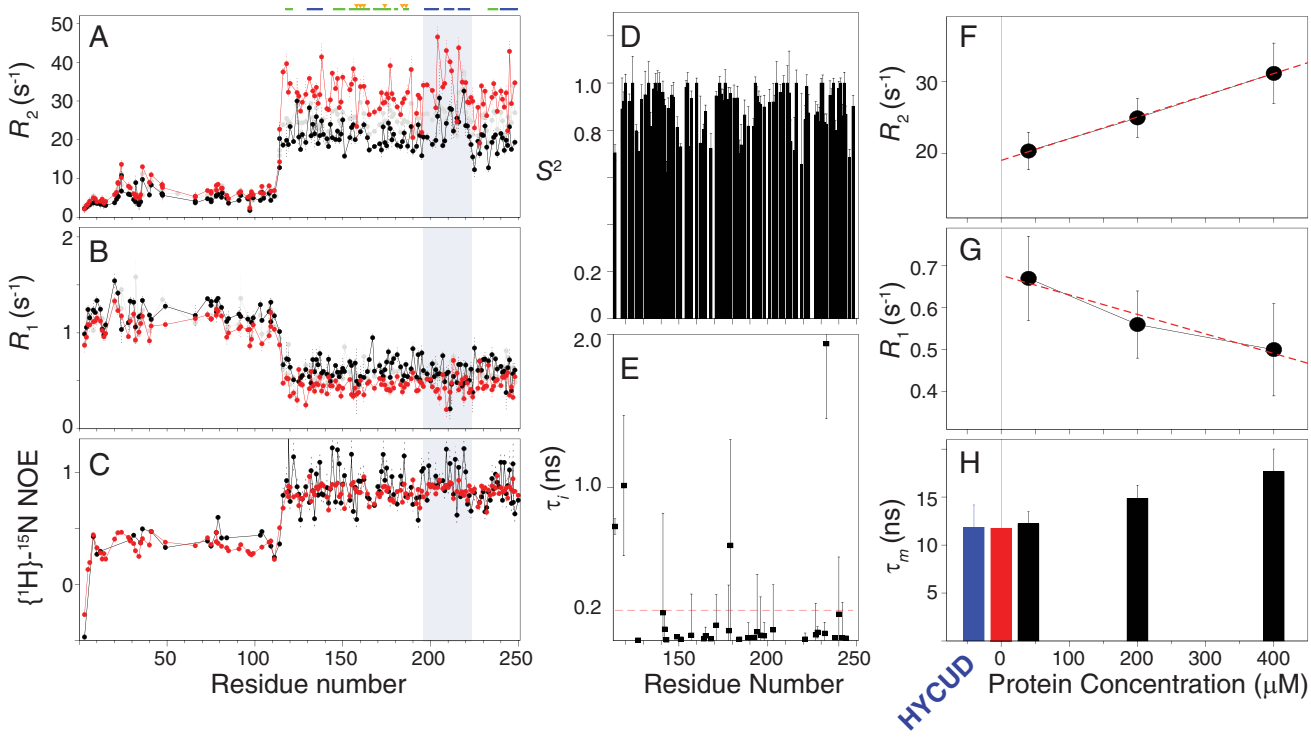

Figure 2. NMR dynamics analysis of galectin-3. Shown are transverse (R_2) (*A*) and longitudinal (R_1) (*B*) relaxation rates and heteronuclear NOEs (*C*) for 40 (*black*), 200 (*gray*), and 400 μM (*red*) samples of galectin-3. The *gray-shaded area* indicates β strands 7–9. Also shown are order parameters (S^2) (*D*) and the internal correlation times (*E*) of the N–H bond vectors (τ_i) in the CRD of galectin-3, calculated using the model-free approach. *F* and *G*, average relaxation rates in the CRD (average over residues 120–200 and 220–250) at different concentrations. The *dashed red lines* were obtained by linear regression to the experimental data. The *gray line* indicates a sample concentration of zero. *H*, rotational correlation times (τ_m) calculated from experimentally measured R_2 and R_1 values (*black bars*) compared with the ones calculated from values extrapolated to zero sample concentration (*red*) and predicted by HYCUD (*blue*). The corresponding values are listed in supplemental Table S2. *Error bars*, S.D.

of spin relaxation spectroscopy. The relaxation data are also indicative of chemical exchange due to micro- to millisecond timescale motions (23–25). We collected the longitudinal relaxation rates (R_1), transverse relaxation rates (R_2), and heteronuclear NOEs (hetNOEs) at different protein concentrations. The residues in the NTD have higher R_1 values, lower R_2 values, and lower hetNOEs than those in the CRD because the intrinsically disordered region is more dynamic (Fig. 2, *A*–*C*). More importantly, the values obtained for R_2 and R_1 are higher and lower, respectively, for the higher-concentration sample, which is generally interpreted as indicating equilibrium exchange between the monomeric form and a larger molecule (26–31). The hetNOEs vary little with concentration (Fig. 2*C*) because hetNOEs are relatively insensitive to the overall tumbling rate of a molecule (23–25). The same concentration-dependent trend from the ^{13}C transverse relaxation experiments and flat Carr-Purcell-Meiboom-Gill relaxation dispersion (32, 33) profiles indicate that the increase of ^{15}N R_2 is not due to chemical exchange (supplemental Fig. S1).

Model-free analysis with one or two types of motions can be applied to interpret the dynamics of a folded protein (34). This analysis may be applied to intrinsically disordered systems with additional motional modes by increasing the number of parameters for mathematical fitting, but more experimental data are required (35). It is possible to describe the dynamics of full-length galectin-3 using other types of motion, but this is beyond the scope of the current study. Because our analyses were mainly focused on the CRD part, we deduced the dynamic properties of full-length galectin-3 from the data collected for the CRD. We used the program Tensor2 (36) to calculate the generalized order parameters (S^2) and internal correlation times of the amide N–H bond vectors (τ_i) using the R_1, R_2, and hetNOE data measured in the 40 μM sample (34). Most of the S^2 values are >0.8, and the τ_i values are <200 ps (Fig. 2, *D* and *E*), indicating that the N–H bond vectors move on a timescale shorter than the overall tumbling rate of the molecule. These results are similar to those of an NMR dynamics study of the CRD only (37). Under such criteria, the overall tumbling rate of the CRD can be calculated directly using Equation 2 (see "Experimental procedures"). We averaged the relaxation rate constants (Fig. 2 (*F* and *G*) and supplemental Table S1) and used first-order extrapolation to estimate the corresponding values at zero protein concentration (31). The extrapolated R_1 and R_2 values give a rotational correlation time for the CRD of ~11.76 ns in the presence of the NTD. We then compared this approximation with the theoretical correlation time. The rotational diffusion of a rigid domain in a protein with an intrinsically disordered tail is hindered because of the presence of disordered peptides (38). The HYCUD algorithm has been used to predict the rotational tumbling dynamics of proteins with flexible linkers (39, 40) and of intrinsically disordered proteins (41). Using the same approach for the CRD in the presence of the NTD yields a correlation time of 11.84 ± 2.34 ns (supplemental Table S1 and Fig. 2*H*). This value agrees well with the one calculated using relaxation times extrapolated to zero protein concentration (11.76 ns), confirming that 40 μM galectin-3 is

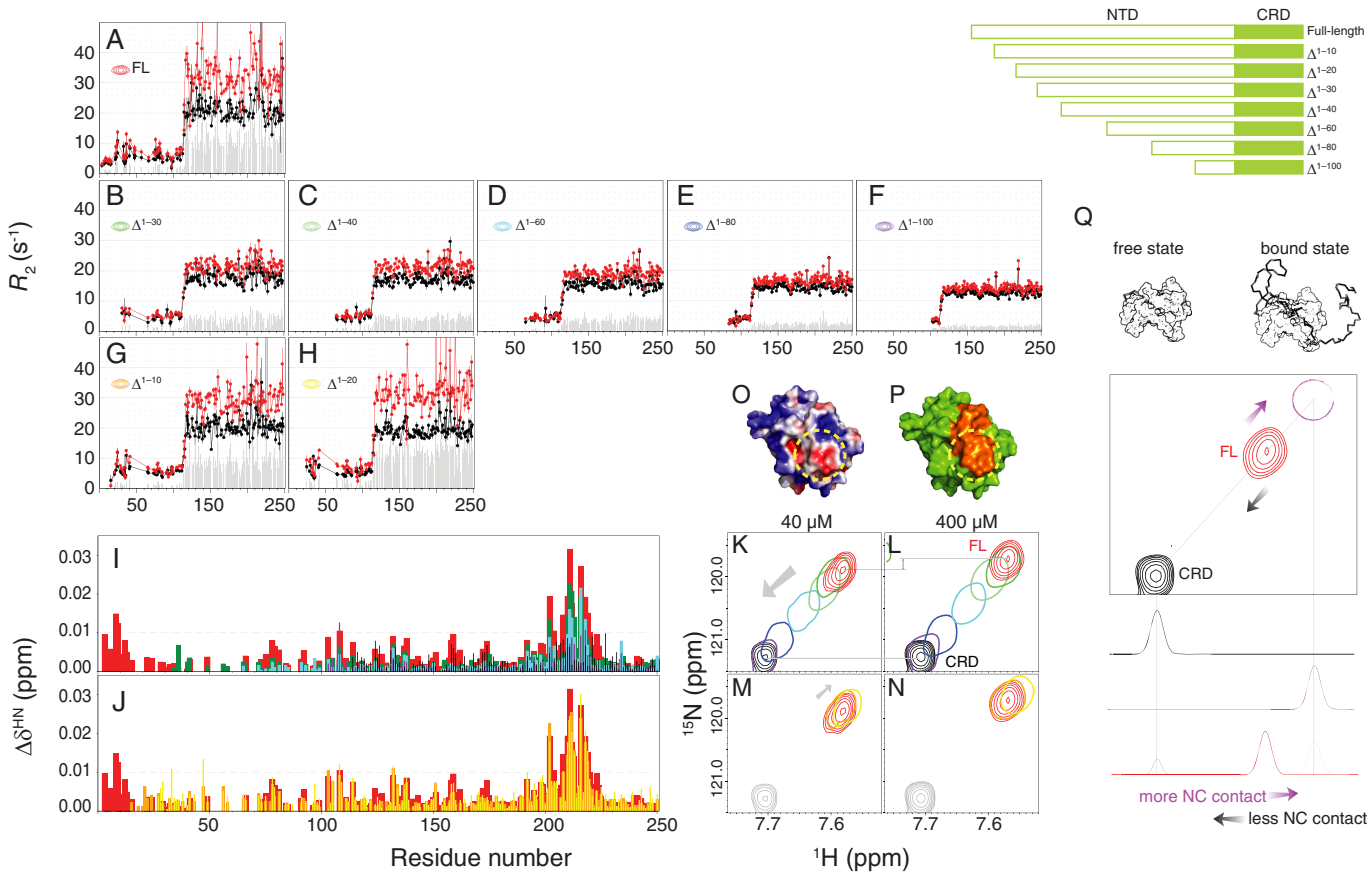

Figure 3. N-terminal truncation studies of galectin-3. *A–H*, transverse relaxation rates (R_2) for 40 (*black*) or 400 μM (*red*) samples of galectin-3 constructs of different lengths (as illustrated in the *top right corner*). The *gray bars* represent the R_2 difference between the high- and low-concentration samples. *I* and *J*, chemical shift differences between 40 and 400 μM samples of full-length (*red*), Δ^{1-10} (*orange*), Δ^{1-20} (*yellow*), Δ^{1-30} (*dark green*), Δ^{1-40} (*light green*), Δ^{1-60} (*cyan*), Δ^{1-80} (*blue*), and Δ^{1-100} (*purple*) constructs of galectin-3. *K–N*, with the same *color scheme*, the HSQC peak from residue 216 for the 40 μM (*K* and *M*) and 400 μM (*L* and *N*) samples of Δ^{1-30}–Δ^{1-100} galectin-3 constructs and the CRD alone (*K* and *L*) and the Δ^{1-10} and Δ^{1-20} constructs and the CRD alone (*M* and *N*). Shown are electrostatic (*O*) and simple surface representations (*P*) of the CRD modeled using APBS (43) with positive and negative charge shown in *blue* and *red*, respectively (*O*), and with residues 200–220 shown in *orange* (*P*). The frequently contacted and negatively charged area is *circled* in *yellow* in *both panels*. *Q*, a graphical explanation of the peak movements observed for the galectin-3 constructs (using residue Ala-216 as an example). The peak corresponding to what would be observed if this site were fully occupied by the NTD (the bound state, not observed experimentally) is shown in *purple*. The intermediate positions of the peaks observed for the full-length protein (in *red*) show that the corresponding residues are in fast exchange between the free and bound states. The more contact there is between the NTD and the CRD (*e.g.* at higher protein concentrations), the greater the population of the bound state is and the closer the experimental peak is to the bound position. The opposite is true when there is less contact between the two domains (*e.g.* shorter constructs). *Error bars*, S.D.

mainly monomeric. Accordingly, for all experiments, we used the data of the 40 μM sample as a monomeric reference to avoid the intrinsic effect on protein dynamics caused by different lengths of the construct, buffer conditions, and working temperatures.

The intrinsically disordered NTD interacts with the CRD in a fuzzy complex manner

We applied the same NMR approach to NTD-only and CRD-only constructs (supplemental Figs. S2 and S3). The absence of any variations in chemical shifts, HSQC peak intensity ratios, and R_2 rates confirms that the CRD alone does not self-associate, whereas the corresponding changes for the NTD alone show that it does (5, 6). Therefore, although the CRD in the full-length construct shows more pronounced chemical shift differences between protein concentrations (Fig. 1, *D–F*), these changes are mediated in the presence of the NTD.

We then systematically truncated the NTD to identify which segment is essential for self-association. The construct without the first 30 residues (Δ^{1-30}) shows the most obvious decrease in the R_2 difference (ΔR_2, *gray bars* in Fig. 3 (*A* and *B*); averaged ΔR_2 value for the CRD are 11.7 ± 1.9 and 4.0 ± 0.9 s^{-1}, respectively (supplemental Table S2)). The ΔR_2 values of Δ^{1-40} are similar to those of Δ^{1-30}, whereas for Δ^{1-40} to Δ^{1-100}, the ΔR_2 values gradually decrease (Fig. 3 (*C–F*), 4.0 ± 0.7, 3.6 ± 0.6, 2.1 ± 0.4, and 1.5 ± 0.3 s^{-1}). Similarly, the chemical shift differences between two concentrations for residues ~200–220 are smaller for the shorter constructs (Fig. 3*I*). These results indicate that all of these 10-residue-based segments (from 20 to 100) are involved to a certain level in self-association, probably because of the repeated PGA*X* motif (Fig. 1*A*; there is no such motif between residues 30 and 40, and this may explain the similar ΔR_2 values between the Δ^{1-30} and Δ^{1-40} constructs). Furthermore, the HSQC peaks from residues 200–220 shift systematically with the number of residues truncated (constructs Δ^{1-30} to Δ^{1-100}; Fig. 3 (*K* and *L*) and supplemental Fig. S4), toward their positions in the CRD-only spectrum (*i.e.* in the opposite direction from the one corresponding to increases in concentration for the full-length construct; compare the *arrows* in Figs. 1*D* and 3*K* for residue 216). These results indicate that

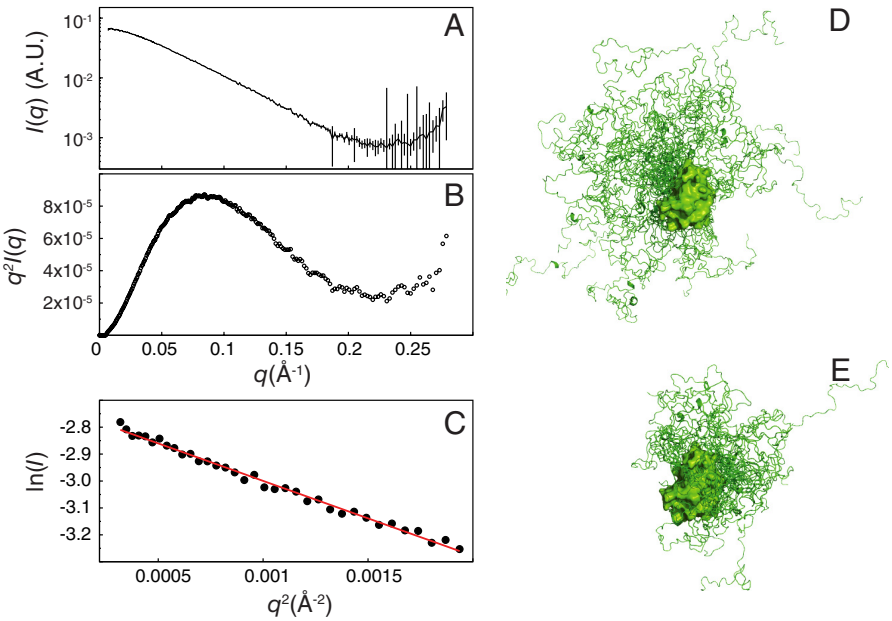

Figure 4. Small-angle X-ray scattering studies. *A*, SAXS intensities of galectin-3 as a function of the scattering vector q (Å^{-1}). *B*, the Kratky plot is typical of that of a partially folded protein (78). *C*, linear fitting of the Guinier plot shows (using Equation 4) that the radius of gyration (R_g) of galectin-3 is ~28.92 Å. The range chosen for the fitting is within the qR_g limit (1.28). This experimental R_g is smaller than that of a non-restrained ensemble (R_g = 34.61 Å) (*D*) but close to that of an ensemble with restrained spatial sampling (R_g = 27.66 Å) (*E*). *A.U.*, arbitrary units.

these residues (positions 200–220) are in the fast exchange regime between a free state (as in the CRD-only construct) and a fully occupied state (Fig. 3*Q*), thus suggesting that there is interaction between the NTD and residues 200–220. The *black peak* (the free state) in Fig. 3*Q* is observed when the surface of residues 200–220 is not in contact with the NTD (*i.e.* when the CRD is studied in isolation). The concentration dependence experiments and those on the systematically truncated constructs show that the more contact there is between the NTD and residues 200–220, the further the corresponding peaks are from their positions in the CRD-only spectrum (Fig. 3*Q*). In the 40 μM samples, intermolecular self-association is unlikely (Fig. 2), so these differences must be caused by intramolecular contacts between the NTD and CRD ("intra-NC"). In the 400 μM samples, these differences reflect intermolecular ("inter-NC") and intra-NC contacts. These results also suggest that many NTD sites can bind to a single CRD region, a typical characteristic of fuzzy interactions (42). Consistent with this fuzzy interaction model, SAXS data collected from samples with a low protein concentration (Fig. 4, *A* and *B*) show that full-length galectin-3 is smaller than a computationally modeled random NTD ensemble (radius of gyration of 28.92 *versus* 34.61 Å; Fig. 4, *C* and *D*); a model with restrained spatial sampling is closer in size (27.66 Å; Fig. 4*E*) to the experimentally determined value (see "Experimental procedures" for details of the analysis).

We also noticed that the constructs with the first 10 or 20 residues removed (Δ^{1-10} and Δ^{1-20}) behave differently from the others, with similar or slightly higher ΔR_2 values (9.4 ± 1.4 and 12.8 ± 1.4 s^{-1} (Fig. 3 (*G* and *H*)) compared with 11.7 ± 1.9 s^{-1} (Fig. 3*A*)) and chemical shift perturbations similar to those of full-length galectin-3 (Fig. 3*J*). The HSQC peaks from residues 200–220 in the Δ^{1-10} and Δ^{1-20} constructs are further away from their positions in the CRD-only spectrum than they are in full-length galectin-3 at both concentrations (Fig. 3 (*M* and *N*) and supplemental Fig. S4). This is the opposite trend from the one shown in Fig. 3 (*K* and *L*) and suggests that the first 20 residues hinder the interaction between the NTD and CRD. The only two charged residues in the NTD (Asp-3 and Asp-9, negatively charged) may induce electrostatic repulsion that interferes with N–C interactions. Indeed, modeling (using APBS (Adaptive Poisson–Boltzmann Solver) (43)) the surface charge distribution of the CRD reveals a negatively charged region that corresponds partly to the surface of residues 200–220 (Fig. 3 (*O* and *P*), *circled* in *yellow*). In addition, the presence of five polar residues between positions 11 and 20 (two serines, two asparagines, and one glutamine) may also hinder contact between the NTD and CRD, because their interaction is mainly driven by hydrophobicity (see below). The different behaviors following the deletion of the first 20 residues may relate to its biological importance; the secretion of galectin-3 is blocked when the first 11 amino acids are removed (44), and a phosphorylated Ser-6 (increasing the total negative charge) is required for its anti-apoptotic activity (45).

Paramagnetic relaxation enhancement experiments show both intermolecular N–N and N–C interactions

To investigate the role of NTD–NTD (supplemental Fig. S2) interactions in full-length galectin-3, we recorded HSQC spectra of ^{15}N-labeled wild-type protein mixed with an equal amount of NMR-inactive (^{14}N) protein with an oxidized (1-oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) methanethiosulfonate (MTSL) label to four different sites (A10C, A31C, A49C, and A100C) in the NTD and one (I250C) in the CRD (Fig. 5*A*) at high and low protein concentrations (I_{ox}). We also recorded HSQC spectra of samples with the same concentration of the wild-type or reduced MTSL protein for normaliza-

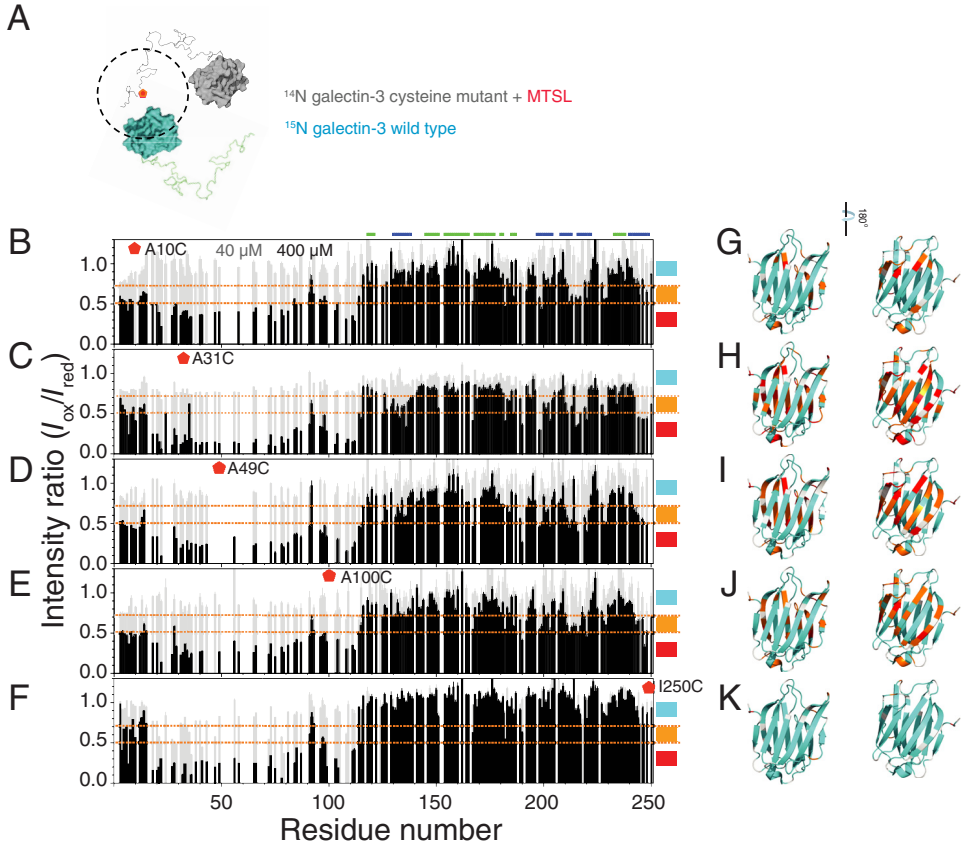

Figure 5. NMR paramagnetic relaxation enhancement studies. *A*, schematic representation of the intermolecular PRE experiments. The *blue* and *gray* molecules represent the ^{15}N-labeled (NMR-active) and NMR-inactive (^{14}N) MTSL-labeled proteins, respectively. The MTSL label is shown in *red*, and the *dashed line* is its putative effective radius. *B–F*, ratios of HSQC peak intensities between samples (*gray bars*, 40 μM; *black bars*, 400 μM) with an oxidized (I_{ox}) and reduced (I_{red}) MTSL label attached to the A10C (*B*), A31C (*C*), A49C (*D*), A100C (*E*), or I250C (*F*) mutants. *G–K*, models of the carbohydrate recognition domain in the same orientation as in Fig. 1*B*, with residues *colored* according to the intensity ratios shown in *B–F*: <0.5 (*red*); between 0.5 and 0.7 (*orange*); >0.7 (*cyan*).

tion (I_{red}). The signal ratios (I_{ox}/I_{red}) between the samples with oxidized and reduced MTSL are shown in Fig. 5 (*B–F*). For the 40 μM samples (*gray bars*), the I_{ox}/I_{red} ratio is close to one for all residues, indicating negligible intermolecular interaction at this concentration (also confirming that MTSL did not introduce an extra intermolecular effect). In the high-concentration samples (*black bars* in Fig. 5 (*B–F*)), the signals from the NTD are weaker when labeled with oxidized MTSL on the NTD, suggesting that the NTDs interact intermolecularly ("inter-NN" interactions) (Fig. 5, *B–E*). A broad range of intensity bleaching in the NTD at these four well-separated spin-label sites suggests that the contact is fuzzy between the NTDs. In the CRD, on the contrary, the peaks whose signal is damped the most come from residues opposite to the carbohydrate-binding site (β-strands 2, 7, 8, 9, and 11) in the A31C, A49C, and A100C instances (the level of bleached intensity is mapped onto the crystal structure in Fig. 5 (*H–J*)), indicating the presence of inter-NC contacts as well (Figs. 1*D* and 3). The MTSL label on the A10C mutant has less contact with the CRD (Fig. 5*G*), in agreement with our conclusion that the first 20 residues inhibit contact between the NTD and CRD (Fig. 3, *M–P*). Nevertheless, this part of the NTD is still involved in the inter-NN interactions, as shown in Fig. 5*B*. The MTSL label on the CRD (I250C; the β-strand 11) bleaches a broad range of intensity in the NTD, consistent with the model that the NTD contacts with the CRD fuzzily and that the first ~20 residues have less contact with this domain (Fig. 5*F*). Furthermore, no obvious intensity bleaching in the CRD indicates that intermolecular interaction between the CRDs is negligible (Fig. 5, *F* and *K*).

Hydrophobicity is the main force for the self-association

To investigate the driving force mediating the observed interactions, we tested the dynamics of the protein in different buffer conditions. Galectin-3 still self-associates when the concentrations of salt (100 mM NaCl) or ligand (250 mM glucose) are high (Fig. 6 (*A–C*); with averaged ΔR_2 values of 11.7 ± 1.9, 13.6 ± 4.1, and 17.3 ± 3.4 s^{-1}, respectively). In the presence of 0.8 M urea, on the contrary, the ΔR_2 values between high- and low-protein concentration samples are substantially reduced (11.7 ± 1.9 s^{-1} *versus* 7.0 ± 1.4 s^{-1}; Fig. 6*D* and supplemental Table S2); the HSQC peak intensity ratios are also closer to the molar ratio (0.1; Fig. 6*H*). In addition, the peaks for residues 200–220 appear at positions closer to those observed in the CRD-only spectrum for both protein concentrations (Fig. 6*I*). These results indicate that hydrophobicity is the driving force for self-association because urea is known to disrupt the hydrophobic effect (46). We tried to increase the urea concentration (up to 4 M) to entirely disrupt these interactions, but the resulting spectra were of poor quality, with the protein having transitioned into the so-called "molten globule" (47) state (the *purple spectrum* in supplemental Fig. S6*A*). Nevertheless, the peaks in Fig. 6*I* are closer to the positions of the CRD-only construct

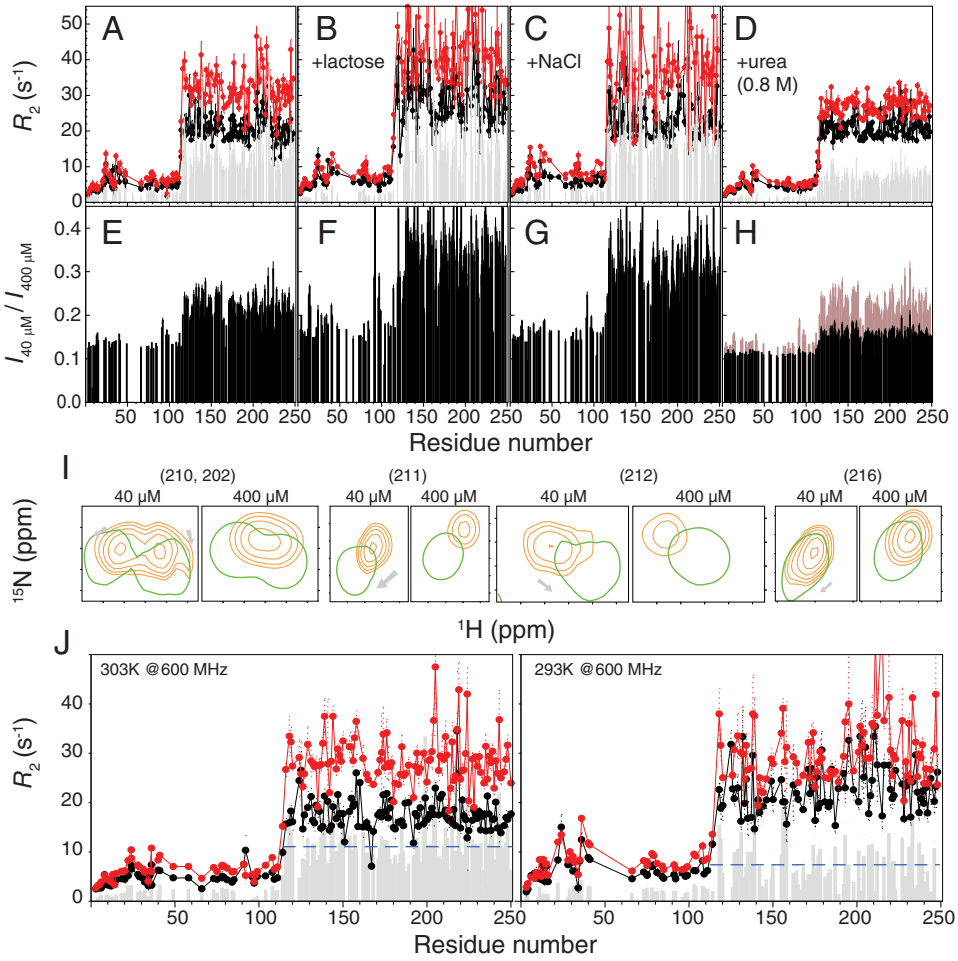

Figure 6. Hydrophobicity drives the self-association of galectin-3. Shown are transverse relaxation rates (R_2) for 40 (*black*) or 400 μM (*red*) samples of galectin-3 in buffer (*A*), with 250 mM lactose (*B*), with 100 mM NaCl (*C*), or with 0.8 M urea (*D*). Shown are ratios of HSQC peak intensities between 40 and 400 μM samples in buffer (*E*), with 250 mM lactose (*F*), with 100 mM NaCl (*G*), or with 0.8 M urea (*H*). *E* is overlaid in *H* for ease of comparison. *I*, representative HSQC peaks (residues 202, 210, 211, 212, and 216) from 40 and 400 μM samples in the absence (*orange*) or presence (*green*) of 0.8 M urea. *J*, R_2 measured in a 14.1-tesla magnet at 303 K (*left*) or 293 K (*right*). The averaged ΔR_2 values for both cases are indicated as *blue dashed lines*. A comparison between samples with and without 250 mM lactose at different protein concentration is shown in supplemental Fig. S5.

from 0.8 to 4 M urea (supplemental Fig. S6, *B–E*), indicating weakened interactions between the NTD and CRD. We also compared ΔR_2 values at different temperatures because the hydrophobic interaction is enhanced at higher temperatures (48). In keeping with our interpretation, the ΔR_2 values are smaller in the low-temperature experiments, indicating a weaker self-association (Fig. 6*J*). The results from these temperature-dependent experiments are in agreement with those obtained using pulse field gradient NMR to measure the diffusion coefficient of self-associated galectin-3 (18).

Discussion

The fuzzy interaction between the NTD and CRD

Unlike other members of the galectin family, galectin-3 is monomeric in solution with only one carbohydrate recognition site. How this molecule achieves its functional multivalency and which domain mediates self-association are subjects of much debate (9, 12, 14–18). Mayo and co-workers' NMR studies (18, 22) provide the opportunity to investigate this protein at a residue-specific level and offer the chance to close these debates. We recorded HSQC spectra at different galectin-3 concentrations to confirm the findings of Ippel *et al.* (18) that the chemical shifts of the residues around 200–220 are concentration-dependent (Fig. 1*D*) and that the intensity ratio between different concentrations is not as expected (Fig. 1*E*). Accordingly, Ippel *et al.* (18) concluded that galectin-3 self-association occurs in the CRD rather than the NTD. This contradicts our findings that the NTD interacts with the CRD through a many-to-one binding, inter- or intramolecularly (Figs. 3 and 5). The chemical shift perturbation is undetected in the NTD, probably because every contact to one site of the CRD is dispersed to many sites of the NTD.

We used NMR spin dynamics experiments to characterize galectin-3 self-association (Fig. 2). The results obtained for NTD-truncated constructs prove that the intramolecular interaction between the NTD and the CRD is fuzzy in character (Fig. 3); no single site of the NTD is critical for its interaction with the CRD. SAXS analysis and computational modeling support this fuzzy interaction model (Fig. 4). PRE experiments show both intermolecular NTD–NTD and NTD–CRD interactions (Fig. 5). The similar PRE patterns obtained with the spin labels placed at four different sites are also consistent with fuzzy inter-

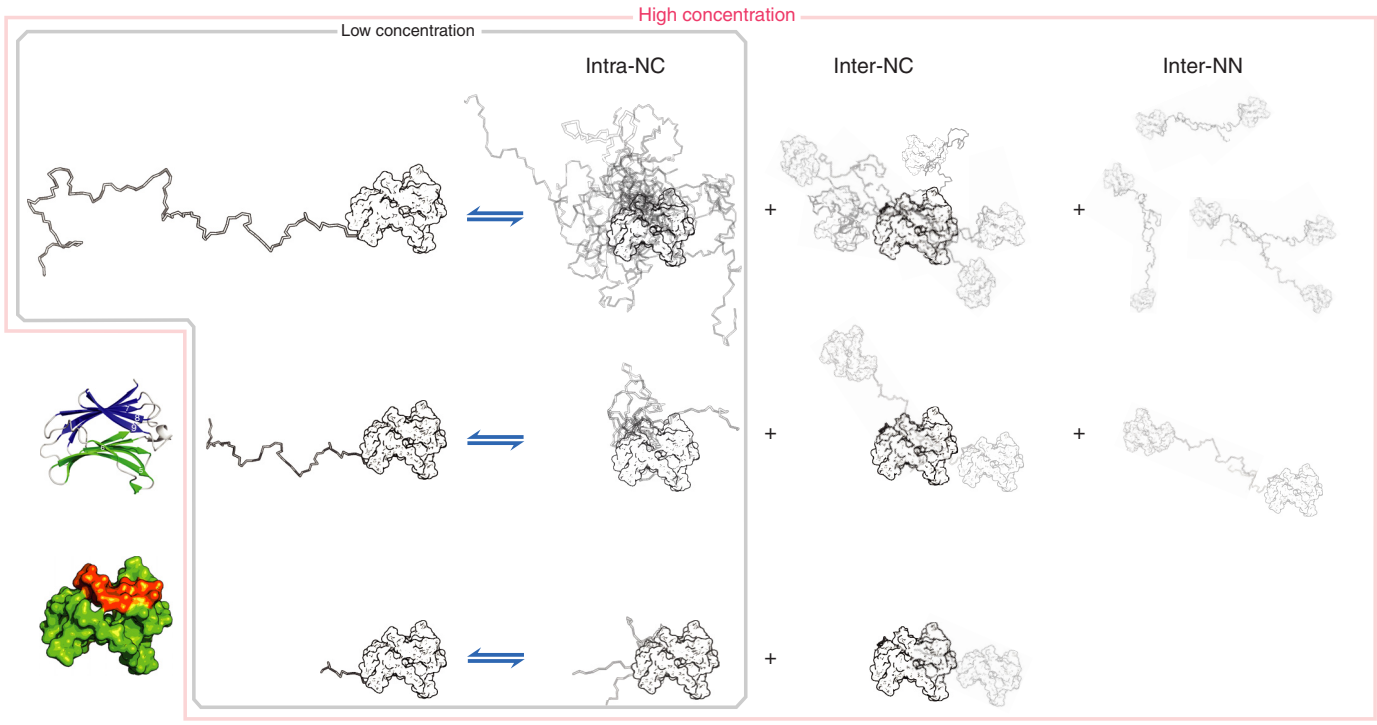

Figure 7. Schematic description of the proposed model of galectin-3 self-association. The *ribbon diagram* in the *bottom left corner* is shown using the *same colors* as in Fig. 1*B*, whereas in the *surface representation below*, shown in the same orientation, residues 200–220 (the interaction region) are shown in *orange*. The remaining surface figures with NTDs of different lengths are shown in the same orientation, in *white*. At low protein concentrations, only intramolecular interactions between the NTD and the CRD occur. The shorter the construct is, the lower the chances are that the NTD interacts with the CRD (Fig. 3, *K* and *Q*). At high concentrations, there are both intra- and intermolecular contacts between the NTD and residues 200–220 of the CRD (Figs. 1*D* and 3*Q*). Furthermore, the PRE experiments show that there are also intermolecular NTD–NTD contacts (Fig. 5). These contacts occur less frequently when the constructs are shorter, as evidenced by smaller chemical shift perturbations (Fig. 3*I*).

actions between the NTDs (Fig. 5). The interaction can be disrupted using mild concentration of urea or be enhanced by increasing the temperature, suggesting that self-association is driven by hydrophobicity (Fig. 6). A model that summarizes these interactions is illustrated in Fig. 7.

The fuzzy interaction and functional implications may be linked through the liquid–liquid phase separation property of galectin-3

The low sequence complexity of the structurally disordered NTD is reminiscent of those recently reported proteins whose low-complexity domains mediate their liquid–liquid phase separation behavior in various cellular functions (49–56). Indeed, we found that the NTD (as does a phosphomimetic mutant S6E) undergoes temperature-dependent phase separation (supplemental Fig. S7). As discussed in the recent review by Wu and Fuxreiter (57), fuzzy contact promotes reversible and dynamic assembly via transient and direct interactions and increases interaction affinities. The fuzzy interactions between the NTD and CRD or between the NTD and NTD might thus assist the formation of a higher-order assembly (*i.e.* may favor liquid–liquid phase separation) in galectin-3. The multivalency of galectin-3 to form galectin–glycan lattice may stem from its ability to undergo liquid–liquid phase separation. Regarding the pentameric model for galectin-3, we note that the situation in which galectin-3 precipitated in the study by Ahmad *et al.* (9) is similar to those in which proteins coacervated in the presence of counterions or nucleic acids (58). The phase-separated form observed here may be an alternative explanation for this protein's multivalency. These results do not discredit the pentamer model, however; in fact, liquid–liquid phase separation, in bringing the proteins together, may also assist the formation of a pentamer.

Conclusion

Our results show that galectin-3 self-associates via inter- and intramolecular NTD–CRD interactions and intermolecular NTD–NTD contacts. To the best of our knowledge, this is the first time these three types of interactions have been shown to occur in a fuzzy manner and to be driven by hydrophobicity. The fact that galectin-3 undergoes liquid–liquid phase separation sheds new light on the protein's function in its self-associated form. Our study also provides new insight on how galectin-3 self-associates when it binds to glycoconjugates, such as those present on the cell surface, resulting in aggregation of these glycoconjugates or the formation of galectin lattices.

Experimental procedures

DNA constructs

The plasmid containing a hexahistidine-tagged small ubiquitin-like modifier protein (His_6-SUMO) in a pHD vector and the protease His_6-Ulp1(403–621) were provided as a gift by Dr. T. F. Wang (Institute of Molecular Biology, Academia Sinica); the DNA construct design and protein purification protocol were adapted from Lee *et al.* (59). The full-length human galectin-3 gene was appended to the His_6-SUMO tag using SfoI (at

the 5′-end) and XhoI (at the 3′-end) cutting sites and designed primers (see supplemental Table S3). The fusion construct (His_6-SUMO-Gal3) was used as a basis for the following construct design. cDNA coding for the NTD alone was created by inserting a stop codon at the site corresponding to residue number 113. Truncated constructs (Δ^{1-10}, Δ^{1-20}, etc.) were designed based on the FastCloning method (60) with designed primers (supplemental Table S3). Cysteine mutations were introduced using appropriate primers (supplemental Table S3). All of the resulting plasmids were fully sequenced.

Protein expression and purification

The protein expression scheme for the fusion construct is the same as our protocol for histidine-tagged proteins described in a previous publication (61). The supernatant of cell lysate was filtered (0.45 μm) and loaded into a nickel-charged immobilized metal-ion affinity chromatography (IMAC) column (Qiagen, Inc.). The column was washed using 10 column volumes of 50 mM Tris-HCl with 300 mM NaCl at pH 7.5, and the bonded protein was eluted using 5 column volumes of the same buffer with an additional 500 mM imidazole. Imidazole was removed using a PD-10 column (GE Healthcare). His_6-Ulp1(403–621) protease was added to the protein solution with a final concentration of 30 μM and left at 4 °C for 2 h to detach the His_6-SUMO tag and galectin-3. The protease-digested solution was loaded into a nickel-charged IMAC column, and the flow-through was collected (supplemental Fig. S8*A*). Protein purity was checked using SDS-polyacrylamide gels (supplemental Fig. S8*B*). The collected protein was also loaded into a G75 gel-filtration column (GE Healthcare) with a FPLC system (supplemental Fig. S8*C*) or a PD-10 column to switch the buffer to a phosphate buffer (20 mM) at pH 6.8. The purified sample was frozen with liquid nitrogen and stored at −80 °C until needed. Protease inhibitor (Roche Applied Science) was added before each experiment.

Our purification protocol differs from most of those published previously (5, 18), in which galectin-3 is captured by lactosylated beads and eluted using a high concentration of lactose (250–500 mM). In those instances, extensive dilution is required to avoid any spurious effects from trace amounts of lactose, or a small amount of lactose (*e.g.* 25 mM) was added to saturate its effect (18).

NMR experiments

All NMR experiments were performed at 303 K, unless otherwise stated, on Bruker AVIII 850- or 600-MHz spectrometers, both equipped with a TCI cryogenic probe. The ^{1}H-^{15}N HSQC spectra were collected using a standard pulse sequence with WATERGATE solvent suppression (62, 63). The spin relaxation experiments were measured using a standard pulse sequence (23, 64) with delay times of 17, 34, 51, 68, 85, and 102 ms to determine the ^{15}N R_2 and 100, 200, 300, 600, 800, and 1000 ms to determine the ^{15}N R_1. Peak intensities were fitted to exponential decays with a Monte Carlo procedure to estimate fitting error. All dynamics data were collected in an interleaved manner with an interscan delay of 3 s.

All NMR data were processed using NMRPipe (65) and analyzed with SPARKY (66). Peak intensities and their errors were measured using non-linear line-shape analysis in NMRPipe (65). The peak intensities from a particular spectra were normalized to the corresponding number of scans for calculating their ratio with standard error propagation (67). The average chemical shift perturbation for a given peak was calculated using the following equation (68),

$$\Delta\delta_{av} = \sqrt{\frac{(\Delta\delta_H)^2 + \left(\frac{1}{5}\Delta\delta_N\right)^2}{2}} \quad \text{(Eq. 1)}$$

where $\Delta\delta_H$ and $\Delta\delta_N$ are the chemical shift differences, respectively, in the proton and nitrogen dimensions, between two HSQC peaks.

MTSL labeling

The nitroxide spin-label MTSL (Toronto Research Chemicals) was attached to the thiol group of cysteine mutations. Using the same purification protocol as described above, after the His_6-SUMO-Gal3 cysteine mutant was eluted from IMAC column, 5 mM tris(2-carboxyethyl)phosphine or 10 mM DTT was added to oxidize the thiol group. After the tris(2-carboxyethyl)phosphine or DTT was diluted 10 times, the sample was concentrated and loaded into a desalting PD-10 column to remove the reducing agent. The MTSL was added immediately in the flow-through of the desalting column to a final concentration of ~600 μM (~10 times the molar amount of protein). After reacting for 30 min, the MTSL was washed out using the PD-10 column. The MTSL-labeling efficiency was checked using Ellman's reagent (69, 70). Protease Ulp1(403–621) was then used to disconnect His_6-SUMO and MTSL-labeled galectin-3, and the subsequent purification steps were identical to those described above. The reduced-MTSL-labeled sample was prepared by adding 3 μl of ascorbic acid from a stock solution to the NMR sample (with a final concentration 2 times that of the protein) or using the same concentration of wild-type sample as normal. The MTSL label cannot interact with the native cysteine at position 173 (in the middle of β-stand 5) because this cysteine's thiol group is buried in the CRD (confirmed by the Ellman's reagent (69, 70)). We used the same MTSL-labeling procedure on ^{15}N-labeled wild-type galectin-3. The identical HSQC spectrum and intensity profile also prove that the native cysteine cannot be spin-labeled. Furthermore, the similar HSQC spectrum from a reduced-MTSL-labeled sample confirm that no extra interaction is introduced by the MTSL (supplemental Fig. S9). The NMR intensity ratios between the reduced- and oxidized-MTSL samples were mapped onto the crystal structure (Protein Data Bank code 2NMO) using an in-house written script.

Dynamics analysis

The squared generalized order parameters (S^2) and internal correlation times of the amide N–H vectors (τ_i) in the folded CRD were determined using the program Tensor2 (36).

The rotational correlation time of the well-structured CRD was calculated using Equation 2 (23),

$$\tau_c = \frac{1}{4\pi\upsilon_N}\sqrt{6\frac{R_2}{R_1} - 7} \quad \text{(Eq. 2)}$$

where υ_N is the ^{15}N resonance frequency in hertz. The values used for R_1 and R_2 were the means of the corresponding values for residues 120–200 and 200–250, and the associated S.D. values were used as error estimates. The error for τ_c was calculated using the standard error propagation procedure for functions of several variables (67), namely the following,

$$\Delta\tau_c = \frac{1}{4\pi\upsilon_N}\frac{3}{R_1}\sqrt{\frac{R_1}{6R_2 - 7R_1}\left(\frac{R_2^2}{R_1^2}(\delta R_1)^2 + (\delta R_2)^2\right)} \quad \text{(Eq. 3)}$$

where δR_1 and δR_2 are the errors derived from the experimental data.

The theoretical τ_c of the CRD in full-length galectin-3 was predicted using the program HYCUD (39), starting from a structural model of the CRD taken from the Protein Data Bank (accession code 2NMO). The disordered domain was constructed using *flexible*-MECCANO (71–73), a program that builds up random coil structures using amino acid-specific conformational potentials. Two thousand conformers were created for this HYCUD analysis. The segmental size for the disordered domain was set to 10 amino acids in HYCUD. The program HYDROPRO (74) was used in conjunction with HYCUD to predict the hydrodynamic properties of the solute. The radius of the primary elements was set to 2.9 Å (74). The viscosity of the solvent was set to 0.8 mPa·s, the viscosity of water at 303 K (75).

SAXS experiments and analysis

SAXS data were collected at the 23A SWAXS end-station equipped with an on-line size exclusion–HPLC system (Agilent chromatographic system 1260 series) at the National Synchrotron Radiation Research Center (NSRRC), Taiwan (76). The sample solution (100 μl of 10 mg ml^{-1}; corresponding to the low protein concentration instances after HPLC dilution) was injected into the HPLC column with a flow rate of 0.35 ml min^{-1} and directed through a quartz capillary (2.0-mm diameter) thermostatted at 303 K for simultaneous SAXS and UV-visible absorption measurements at the same sample position with orthogonal incidences. SAXS data were collected with 1 data frame per 30 s using a Pilatus 1M-F area detector. Buffer solutions were measured under identical conditions for background scattering subtraction. With 15-keV X-rays (wavelength λ = 0.8266 Å) and a sample-to-detector distance of 3.17 m, the scattering vector q, defined by $4\pi\lambda^{-1}\sin\theta$ with scattering angle 2θ, covered 0.007–0.25 Å^{-1}. Data were evaluated for radiation damage, background subtraction quality, and sample concentration effects, and well-overlapped SAXS profiles collected over the sample elution peak of HPLC were integrated for improved data statistics.

The data were analyzed using the ATSAS software package (77). The scattering intensity I as a function of the scattering vector q (Å^{-1}) was replotted using $q^2{\cdot}I(q)$ *versus* q, forming Kratky plots (78), to demonstrate the degree of compactness of the sample. The scattering data were also replotted using the Guinier approximation ($\ln[I(q)]$ *versus* q^2) (78),

$$\ln(I(q)) \cong \ln(I(0)) - \frac{1}{3}R_g^2q^2 \quad \text{(Eq. 4)}$$

where R_g is the radius of gyration of the molecule, which can be obtained by linear regression (Fig. 4*C*). The Guinier approximation is valid only for very small angles (*i.e.* for $q < 1.3/R_g$) (78).

Modeling of conformational ensembles

Ten thousand randomized NTD conformers attached to the CRD were generated using *flexible*-MECCANO (71–73). The R_g was calculated based on the root-mean-squared distances of all of the Cα atoms from the center of mass of each conformer,

$$R_g = \sqrt{\frac{1}{N}\sum_i^N(\vec{r}_i - \langle\vec{r}\rangle)^2} \quad \text{(Eq. 5)}$$

where N is the total number of the Cα atoms, and the vector r_i represents the coordinates of atom i. Among these 10,000 conformers, around 400 having at least one pair of Cα atoms <8 Å apart, one from residues 20–100 and the other from residue 200–220, were selected as the ensemble representing restricted spatial sampling.

Author contributions—Y.-H. L. collected and analyzed the NMR dynamics and PRE data and created part of the constructs. D.-C. Q. collected and analyzed the PRE data. W.-H. C. designed and created part of the constructs. Y.-Q. Y. collected and analyzed the SAXS data. U.-S. J. collected and analyzed the SAXS data. F.-T. L. conceived the study, wrote the paper, and obtained funding. J. H. conceived the study, collected the data, analyzed the data, wrote the paper, and obtained funding.

Acknowledgments—We acknowledge Chun-Yi Lee and Wei-Jie Chen for initial work on this project. We thank Dr. Chi-Fon Chang and the Core Facility for Protein Structural Analysis at Academia Sinica for technical NMR support, Cho-Ying Chiang and Dr. Huan-Yuan Chen (GlycoCore, Academia Sinica) for help in preparing galectin-3, and Professor Won-Jing Wang (National Yang-Ming University) for help in using the microscope.

References

1. Yang, R. Y., Rabinovich, G. A., and Liu, F. T. (2008) Galectins: structure, function and therapeutic potential. *Expert Rev. Mol. Med.* **10,** e17
2. Nabi, I. R., Shankar, J., and Dennis, J. W. (2015) The galectin lattice at a glance. *J. Cell Sci.* **128,** 2213–2219
3. Rabinovich, G. A., Toscano, M. A., Jackson, S. S., and Vasta, G. R. (2007) Functions of cell surface galectin-glycoprotein lattices. *Curr. Opin. Struct. Biol.* **17,** 513–520
4. Dumic, J., Dabelic, S., and Flögel, M. (2006) Galectin-3: an open-ended story. *Biochim. Biophys. Acta* **1760,** 616–635
5. Hsu, D. K., Zuberi, R. I., and Liu, F. T. (1992) Biochemical and biophysical characterization of human recombinant IgE-binding protein, an S-type animal lectin. *J. Biol. Chem.* **267,** 14167–14174
6. Massa, S. M., Cooper, D. N., Leffler, H., and Barondes, S. H. (1993) L-29, an endogenous lectin, binds to glycoconjugate ligands with positive cooperativity. *Biochemistry* **32,** 260–267
7. Kuwabara, I., and Liu, F. T. (1996) Galectin-3 promotes adhesion of human neutrophils to laminin. *J. Immunol.* **156,** 3939–3944
8. Yamaoka, A., Kuwabara, I., Frigeri, L. G., and Liu, F. T. (1995) A human lectin, galectin-3 (epsilon bp/Mac-2), stimulates superoxide production by neutrophils. *J. Immunol.* **154,** 3479–3487
9. Ahmad, N., Gabius, H. J., André, S., Kaltner, H., Sabesan, S., Roy, R., Liu, B., Macaluso, F., and Brewer, C. F. (2004) Galectin-3 precipitates as a penta-

mer with synthetic multivalent carbohydrates and forms heterogeneous cross-linked complexes. *J. Biol. Chem.* **279,** 10841–10847
10. Morris, S., Ahmad, N., André, S., Kaltner, H., Gabius, H. J., Brenowitz, M., and Brewer, F. (2004) Quaternary solution structures of galectins-1, -3, and -7. *Glycobiology* **14,** 293–300
11. Roff, C. F., and Wang, J. L. (1983) Endogenous lectins from cultured cells. Isolation and characterization of carbohydrate-binding proteins from 3T3 fibroblasts. *J. Biol. Chem.* **258,** 10657–10663
12. Nieminen, J., Kuno, A., Hirabayashi, J., and Sato, S. (2007) Visualization of galectin-3 oligomerization on the surface of neutrophils and endothelial cells using fluorescence resonance energy transfer. *J. Biol. Chem.* **282,** 1374–1383
13. Fermino, M. L., Polli, C. D., Toledo, K. A., Liu, F. T., Hsu, D. K., Roque-Barreira, M. C., Pereira-da-Silva, G., Bernardes, E. S., and Halbwachs-Mecarelli, L. (2011) LPS-induced galectin-3 oligomerization results in enhancement of neutrophil activation. *PLoS One* **6,** e26004
14. Halimi, H., Rigato, A., Byrne, D., Ferracci, G., Sebban-Kreuzer, C., ElAntak, L., and Guerlesquin, F. (2014) Glycan dependence of galectin-3 self-association properties. *PLoS One* **9,** e111836
15. Kuklinski, S., and Probstmeier, R. (1998) Homophilic binding properties of galectin-3: involvement of the carbohydrate recognition domain. *J. Neurochem.* **70,** 814–823
16. Lepur, A., Salomonsson, E., Nilsson, U. J., and Leffler, H. (2012) Ligand induced galectin-3 protein self-association. *J. Biol. Chem.* **287,** 21751–21756
17. Yang, R. Y., Hill, P. N., Hsu, D. K., and Liu, F. T. (1998) Role of the carboxyl-terminal lectin domain in self-association of galectin-3. *Biochemistry* **37,** 4086–4092
18. Ippel, H., Miller, M. C., Vértesy, S., Zheng, Y., Cañada, F. J., Suylen, D., Umemoto, K., Romanò, C., Hackeng, T., Tai, G., Leffler, H., Kopitz, J., André, S., Kübler, D., Jiménez-Barbero, J., *et al.* (2016) Intra- and intermolecular interactions of human galectin-3: assessment by full-assignment-based NMR. *Glycobiology* **26,** 888–903
19. Mehul, B., Bawumia, S., Martin, S. R., and Hughes, R. C. (1994) Structure of baby hamster kidney carbohydrate-binding protein CBP30, an S-type animal lectin. *J. Biol. Chem.* **269,** 18250–18258
20. Birdsall, B., Feeney, J., Burdett, I. D., Bawumia, S., Barboni, E. A., and Hughes, R. C. (2001) NMR solution studies of hamster galectin-3 and electron microscopic visualization of surface-adsorbed complexes: evidence for interactions between the N- and C-terminal domains. *Biochemistry* **40,** 4859–4866
21. Ippel, H., Miller, M. C., Berbís, M. A., Suylen, D., André, S., Hackeng, T. M., Cañada, F. J., Weber, C., Gabius, H. J., Jiménez-Barbero, J., and Mayo, K. H. (2015) ^{1}H, ^{13}C, and ^{15}N backbone and side-chain chemical shift assignments for the 36 proline-containing, full length 29 kDa human chimera-type galectin-3. *Biomol. NMR Assign.* **9,** 59–63
22. Berbís, M. Á., André, S., Cañada, F. J., Pipkorn, R., Ippel, H., Mayo, K. H., Kübler, D., Gabius, H. J., and Jiménez-Barbero, J. (2014) Peptides derived from human galectin-3 N-terminal tail interact with its carbohydrate recognition domain in a phosphorylation-dependent manner. *Biochem. Biophys. Res. Commun.* **443,** 126–131
23. Kay, L. E., Torchia, D. A., and Bax, A. (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28,** 8972–8979
24. Akke, M. (2012) Conformational dynamics and thermodynamics of protein-ligand binding studied by NMR relaxation. *Biochem. Soc. Trans.* **40,** 419–423
25. Fawzi, N. L., Ying, J., Ghirlando, R., Torchia, D. A., and Clore, G. M. (2011) Atomic-resolution dynamics on the surface of amyloid-β protofibrils probed by solution NMR. *Nature* **480,** 268–272
26. Fushman, D., Cahill, S., and Cowburn, D. (1997) The main-chain dynamics of the dynamin pleckstrin homology (PH) domain in solution: analysis of ^{15}N relaxation with monomer/dimer equilibration. *J. Mol. Biol.* **266,** 173–194
27. Pfuhl, M., Chen, H. A., Kristensen, S. M., and Driscoll, P. C. (1999) NMR exchange broadening arising from specific low affinity protein self-association: analysis of nitrogen-15 nuclear relaxation for rat CD2 domain 1. *J. Biomol. NMR* **14,** 307–320
28. Korchuganov, D. S., Nolde, S. B., Reibarkh, M. Y., Orekhov, V. Y., Schulga, A. A., Ermolyuk, Y. S., Kirpichnikov, M. P., and Arseniev, A. S. (2001) NMR study of monomer-dimer equilibrium of barstar in solution. *J. Am. Chem. Soc.* **123,** 2068–2069
29. Akerud, T., Thulin, E., Van Etten, R. L., and Akke, M. (2002) Intramolecular dynamics of low molecular weight protein tyrosine phosphatase in monomer-dimer equilibrium studied by NMR: a model for changes in dynamics upon target binding. *J. Mol. Biol.* **322,** 137–152
30. Baryshnikova, O. K., and Sykes, B. D. (2006) Backbone dynamics of SDF-1alpha determined by NMR: interpretation in the presence of monomer-dimer equilibrium. *Protein Sci.* **15,** 2568–2578
31. Jensen, M. R., Kristensen, S. M., Keeler, C., Christensen, H. E., Hodsdon, M. E., and Led, J. J. (2008) Weak self-association of human growth hormone investigated by nitrogen-15 NMR relaxation. *Proteins* **73,** 161–172
32. Mulder, F. A., Mittermaier, A., Hon, B., Dahlquist, F. W., and Kay, L. E. (2001) Studying excited states of proteins by NMR spectroscopy. *Nat. Struct. Biol.* **8,** 932–935
33. Sugase, K., Dyson, H. J., and Wright, P. E. (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447,** 1021–1025
34. Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104,** 4546–4559
35. Abyzov, A., Salvi, N., Schneider, R., Maurin, D., Ruigrok, R. W., Jensen, M. R., and Blackledge, M. (2016) Identification of dynamic modes in an intrinsically disordered protein using temperature-dependent NMR relaxation. *J. Am. Chem. Soc.* **138,** 6240–6251
36. Dosset, P., Hus, J. C., Blackledge, M., and Marion, D. (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR* **16,** 23–28
37. Diehl, C., Genheden, S., Modig, K., Ryde, U., and Akke, M. (2009) Conformational entropy changes upon lactose binding to the carbohydrate recognition domain of galectin-3. *J. Biomol. NMR* **45,** 157–169
38. Bae, S. H., Dyson, H. J., and Wright, P. E. (2009) Prediction of the rotational tumbling time for proteins with disordered segments. *J. Am. Chem. Soc.* **131,** 6814–6821
39. Rezaei-Ghaleh, N., Klama, F., Munari, F., and Zweckstetter, M. (2013) Predicting the rotational tumbling of dynamic multidomain proteins and supramolecular complexes. *Angew. Chem. Int. Ed. Engl.* **52,** 11410–11414
40. Rezaei-Ghaleh, N., Klama, F., Munari, F., and Zweckstetter, M. (2015) HYCUD: a computational tool for prediction of effective rotational correlation time in flexible proteins. *Bioinformatics* **31,** 1319–1321
41. Parigi, G., Rezaei-Ghaleh, N., Giachetti, A., Becker, S., Fernandez, C., Blackledge, M., Griesinger, C., Zweckstetter, M., and Luchinat, C. (2014) Long-range correlated dynamics in intrinsically disordered proteins. *J. Am. Chem. Soc.* **136,** 16201–16209
42. Sharma, R., Raduly, Z., Miskei, M., and Fuxreiter, M. (2015) Fuzzy complexes: specific binding without complete folding. *FEBS Lett.* **589,** 2533–2542
43. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 10037–10041
44. Gong, H. C., Honjo, Y., Nangia-Makker, P., Hogan, V., Mazurak, N., Bresalier, R. S., and Raz, A. (1999) The NH_2 terminus of galectin-3 governs cellular compartmentalization and functions in cancer cells. *Cancer Res.* **59,** 6239–6245
45. Yoshii, T., Fukumori, T., Honjo, Y., Inohara, H., Kim, H. R., and Raz, A. (2002) Galectin-3 phosphorylation is required for its anti-apoptotic function and cell cycle arrest. *J. Biol. Chem.* **277,** 6852–6857
46. Zangi, R., Zhou, R., and Berne, B. J. (2009) Urea's action on hydrophobic interactions. *J. Am. Chem. Soc.* **131,** 1535–1541
47. Redfield, C. (2004) Using nuclear magnetic resonance spectroscopy to study molten globule states of proteins. *Methods* **34,** 121–132
48. Chandler, D. (2005) Interfaces and the driving force of hydrophobic assembly. *Nature* **437,** 640–647
49. Nott, T. J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T. D., Bazett-Jones, D. P., Pawson, T., Forman-Kay, J. D., and Baldwin, A. J. (2015) Phase transition of a disordered nuage protein gen-

erates environmentally responsive membraneless organelles. *Mol. Cell* **57,** 936–947
50. Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T. M., Pozniakovski, A., Poser, I., Maghelli, N., Royer, L. A., Weigert, M., *et al.* (2015) A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162,** 1066–1077
51. Jiang, H., Wang, S., Huang, Y., He, X., Cui, H., Zhu, X., and Zheng, Y. (2015) Phase transition of spindle-associated protein regulate spindle apparatus assembly. *Cell* **163,** 108–122
52. Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A. P., Kim, H. J., Mittag, T., and Taylor, J. P. (2015) Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163,** 123–133
53. Zhang, H., Elbaum-Garfinkle, S., Langdon, E. M., Taylor, N., Occhipinti, P., Bridges, A. A., Brangwynne, C. P., and Gladfelter, A. S. (2015) RNA controls polyQ protein phase transitions. *Mol. Cell* **60,** 220–230
54. Lin, Y., Protter, D. S., Rosen, M. K., and Parker, R. (2015) Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol. Cell* **60,** 208–219
55. Xiang, S., Kato, M., Wu, L. C., Lin, Y., Ding, M., Zhang, Y., Yu, Y., and McKnight, S. L. (2015) The LC domain of hnRNPA2 adopts similar conformations in hydrogel polymers, liquid-like droplets, and nuclei. *Cell* **163,** 829–839
56. Burke, K. A., Janke, A. M., Rhine, C. L., and Fawzi, N. L. (2015) Residue-by-residue view of *in vitro* FUS granules that bind the C-terminal domain of RNA polymerase II. *Mol. Cell* **60,** 231–241
57. Wu, H., and Fuxreiter, M. (2016) The structure and dynamics of higher-order assemblies: amyloids, signalosomes, and granules. *Cell* **165,** 1055–1066
58. Pak, C. W., Kosno, M., Holehouse, A. S., Padrick, S. B., Mittal, A., Ali, R., Yunus, A. A., Liu, D. R., Pappu, R. V., and Rosen, M. K. (2016) Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. *Mol. Cell* **63,** 72–85
59. Lee, C. D., Sun, H. C., Hu, S. M., Chiu, C. F., Homhuan, A., Liang, S. M., Leng, C. H., and Wang, T. F. (2008) An improved SUMO fusion protein system for effective production of native proteins. *Protein Sci.* **17,** 1241–1248
60. Li, C., Wen, A., Shen, B., Lu, J., Huang, Y., and Chang, Y. (2011) FastCloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. *BMC Biotechnol.* **11,** 92
61. Chen, T. C., Hsiao, C. L., Huang, S. J., and Huang, J. R. (2016) The nearest-neighbor effect on random-coil NMR chemical shifts demonstrated using a low-complexity amino-acid sequence. *Protein Pept. Lett.* **23,** 967–975
62. Piotto, M., Saudek, V., and Sklenár, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **2,** 661–665
63. Bodenhausen, G., and Ruben, D. J. (1980) Natural abundance N-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69,** 185–189
64. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* **33,** 5984–6003
65. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–293
66. Goddard, T. D., and Kneller, D. G. (2005) *Sparky 3*, University of California, San Francisco
67. Taylor, J. R. (1997) *An Introduction to Error Analysis*, 2nd Ed., University Science Books, Sausalito, CA
68. Grzesiek, S., Stahl, S. J., Wingfield, P. T., and Bax, A. (1996) The CD4 determinant for downregulation by HIV-1 Nef directly binds to Nef: mapping of the Nef binding surface by NMR. *Biochemistry* **35,** 10256–10261
69. Gillespie, J. R., and Shortle, D. (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J. Mol. Biol.* **268,** 158–169
70. Ellman, G. L. (1959) Tissue sulfhydryl groups. *Arch. Biochem. Biophys.* **82,** 70–77
71. Ozenne, V., Bauer, F., Salmon, L., Huang, J. R., Jensen, M. R., Segard, S., Bernadó, P., Charavay, C., and Blackledge, M. (2012) Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* **28,** 1463–1470
72. Huang, J. R., Gentner, M., Vajpai, N., Grzesiek, S., and Blackledge, M. (2012) Residual dipolar couplings measured in unfolded proteins are sensitive to amino-acid-specific geometries as well as local conformational sampling. *Biochem. Soc. Trans.* **40,** 989–994
73. Bernadó, P., Blanchard, L., Timmins, P., Marion, D., Ruigrok, R. W., and Blackledge, M. (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 17002–17007
74. García De La Torre, J., Huertas, M. L., and Carrasco, B. (2000) Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* **78,** 719–730
75. Kestin, J. S., Mordechai; Wakeham William, A. (1978) Viscosity of liquid water in the range −8 °C to 150 °C. *J. Phys. Chem. Ref. Data* **7,** 941
76. Yeh, Y. Q., Liao, K. F., Shih, O., Shiu, Y. J., Wu, W. R., Su, C. J., Lin, P. C., and Jeng, U. S. (2017) Probing the acid-induced packing structure changes of the molten globule domains of a protein near equilibrium unfolding. *J. Phys. Chem. Lett.* **8,** 470–477
77. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D., Konarev, P. V., and Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45,** 342–350
78. Putnam, C. D., Hammel, M., Hura, G. L., and Tainer, J. A. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40,** 191–285
79. Dosztányi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21,** 3433–3434

Supplemental Data

The Intrinsically Disordered N-terminal Domain of Galectin-3 Dynamically Mediates Its Multisite Self-association of the Protein through Fuzzy Interactions

Yu-Hao Lin[a], De-Chen Qiu[a], Wen-Han Chang[a], Yi-Qi Yeh[c], U-Ser Jeng[c,d], Fu-Tong Liu[e], and Jie-rong Huang[a,b,*]

From the [a]Institute of Biochemistry and Molecular Biology, [b]Institute of Biomedical Informatics, National Yang-Ming University, No. 155 Section 2 Li-nong Street, Taipei, Taiwan
[c]National Synchrotron Radiation Research Center, Hsinchu 30076, Taiwan
[d]Department of Chemical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan
[e]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan

Running title: *Galectin-3 self-association*

To whom correspondence should be addressed: Prof. Jie-rong Huang, Institute of Biochemistry and Molecular Biology, National Yang-Ming University, No. 155 Section 2 Li-nong Street, Taipei, Taiwan; Telephone: (+886)-2-2826-7258; Email: jierongh@ym.edu.tw

Supplemental Figures

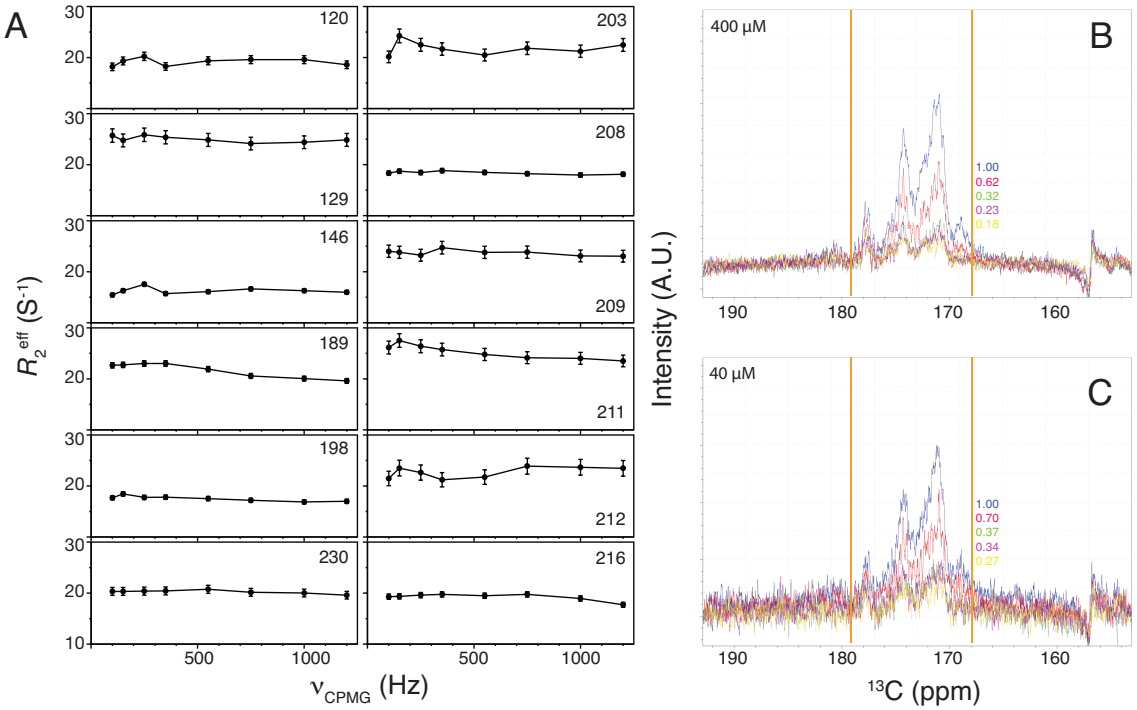

Figure S1. (*A*) Typical ^{15}N Carr-Purcell-Meiboom-Gill relaxation dispersion curves for 400 µM full-length galactin-3. The residue number is indicated in each panel. (*B,C*) The one-dimensional ^{13}C spectrum with different T_2 delays (blue: 20 ms, red: 40 ms, green: 60 ms; purple: 80 ms; yellow: 100 ms) for (*B*) 400 µM and (*C*) 40 µM samples. The numbers of scans were 64 and 1024 respectively. The integrated area below each spectrum (between 168 to 179 ppm, indicated as orange lines) were normalized to the one of the first decay and listed in the same color scheme.

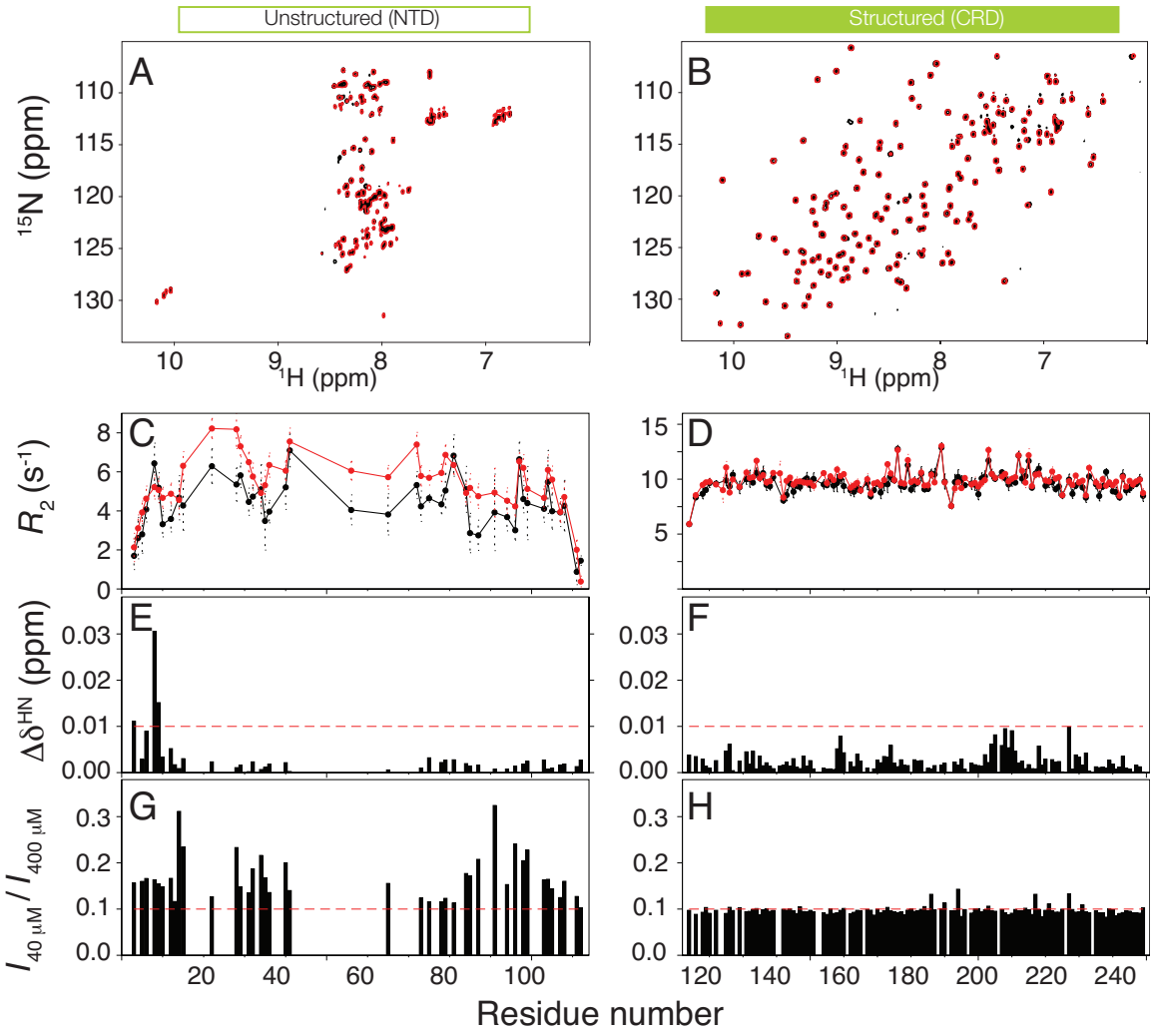

FIGURE S2. (*A,B*) HSCQ spectra, (*C,D*) transverse relaxation rate constants, (*E,F*) chemical shift perturbation, and (*G,H*) ratios of peak intensities ratio for (*A,C,E,G*) the N-terminal domain (NTD) of galectin-3 alone, and (*B,D,F,G*) the carbohydrate recognition domain (CRD) alone. In parts (*A–D*), the data from 40 or 400 μM samples are shown in black and red respectively.

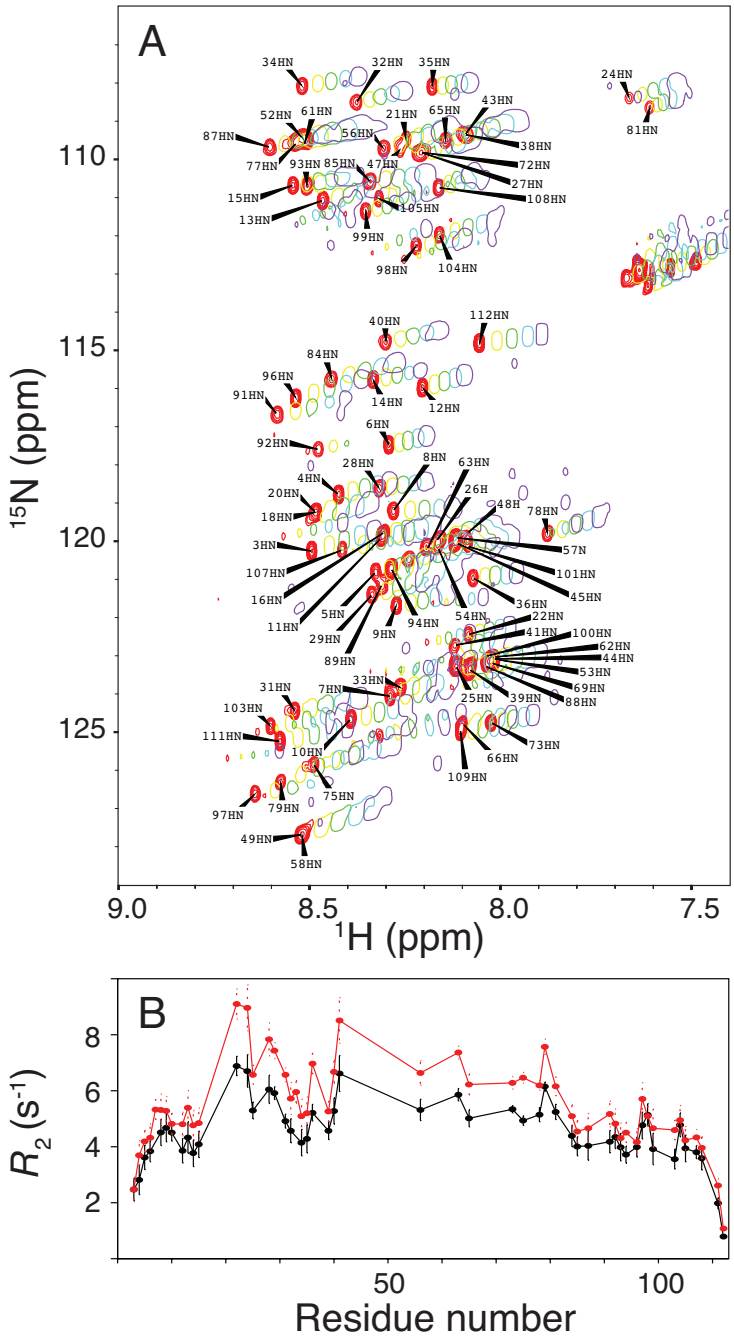

Figure S3. (*A*) Assigned HSQC spectrum of the N-terminal domain (NTD) of galectin-3 at 283 K (red) overlaid on the spectra recorded at 288 (yellow), 293 (green), 298 (cyan), and 303 K (purple). (*B*) NMR transverse relaxation rates (R_2) of the NTD collected at 283 K on a 850 MHz spectrometer for 40 (black) and 400 μM (red) samples.

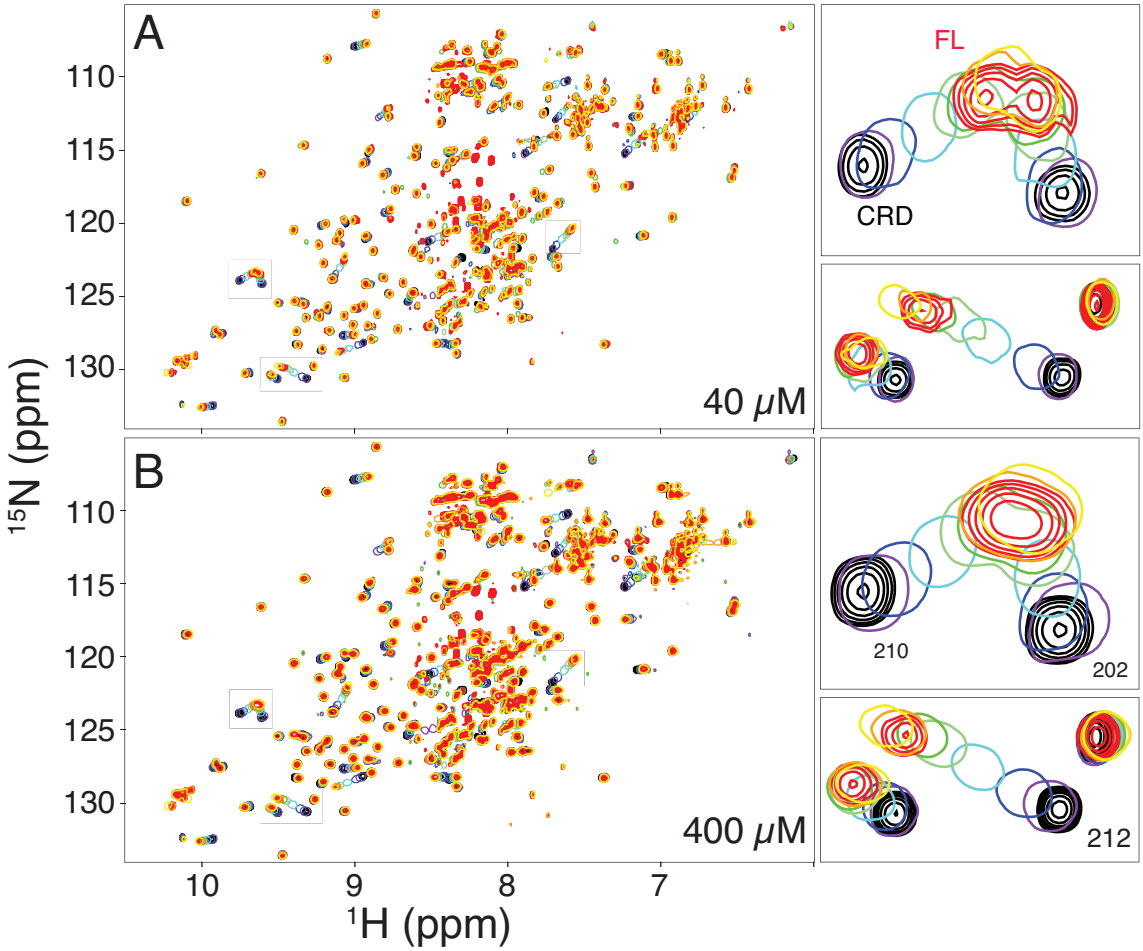

Figure S4. Overlay of the HSQC spectra of (*A*) 40 and (*B*) 400 μM samples of full-length (red), $\Delta^{1\text{-}10}$ (orange), $\Delta^{1\text{-}20}$ (yellow), $\Delta^{1\text{-}30}$ (dark green), $\Delta^{1\text{-}40}$ (light green), $\Delta^{1\text{-}60}$ (cyan), $\Delta^{1\text{-}80}$ (blue), and $\Delta^{1\text{-}100}$ (purple) constructs of galectin-3. Expanded views are shown on the right of the peaks whose positions change the most between constructs.

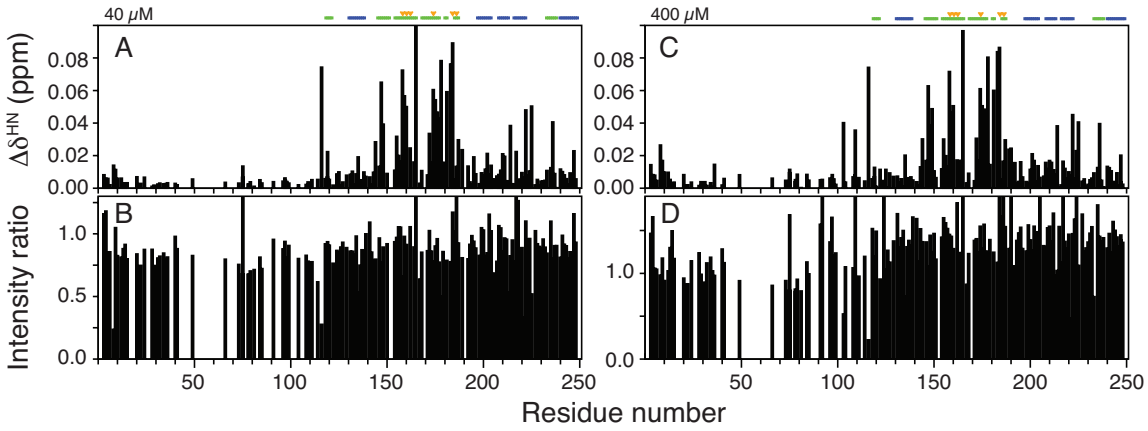

Figure S5. (*A,C*) Chemical shift differences and (*B,D*) ratios of HSQC peak intensities between samples of galectin-3 with and without 250 mM lactose at a protein concentration of (*A,B*) 40 and (*C,D*) 400 μM. The most prominent chemical shift perturbations as expected occur around the carbohydrate-binding site (green bars and orange triangles).

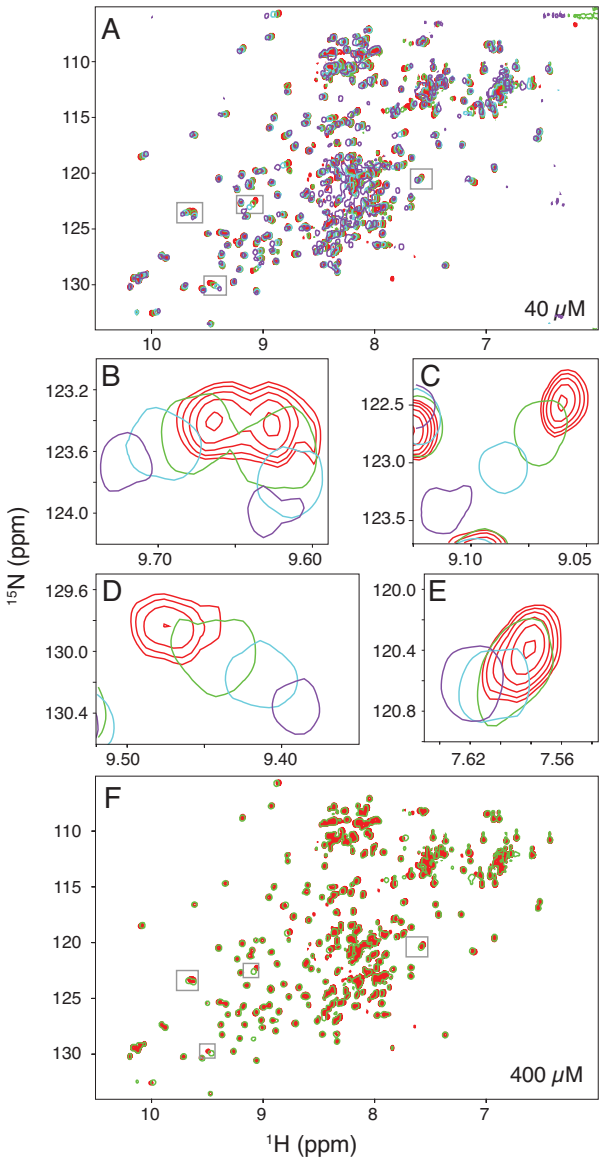

Figure S6. (*A*) Overlay of HSQC spectra of 40 μM samples of galectin-3 without urea (red) or in the presence of 0.8 (green), 2 (cyan), and 4 M (purple) urea (the latter spectrum is of lesser quality because the protein becomes denatured). (*B–E*) Expanded views of the peaks whose positions change the most, namely those assigned to (*B*) residue 210 and 202, (*C*) residue 211, (*D*) residue 212, and (*E*) residue 216. (*F*) Overlaid HSQC spectra of 400 μM samples of galectin-3 in the absence (red) and presence of 0.8 M urea (green). The squares indicate the peaks highlighted in Fig. 6*I*.

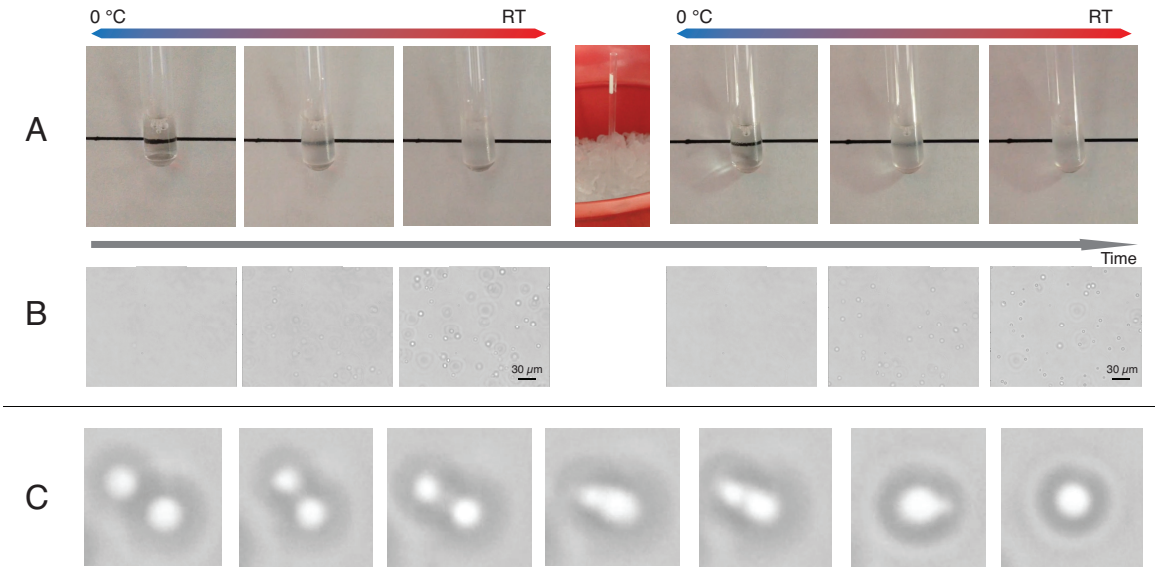

Figure S7. (*A*) Photographs and (*B*) microscope images of liquid-liquid phase separation of the N-terminal domain of galectin-3 (400 μM protein sample in the presence of 300 mM NaCl). The sample, initially transparent at ~0 °C (in iced water), becomes clouded when left for ~90 s at room temperature. This process is reversible. (*C*) An example of two droplets fusion event (less than ten seconds) demonstrates their liquid-like property.

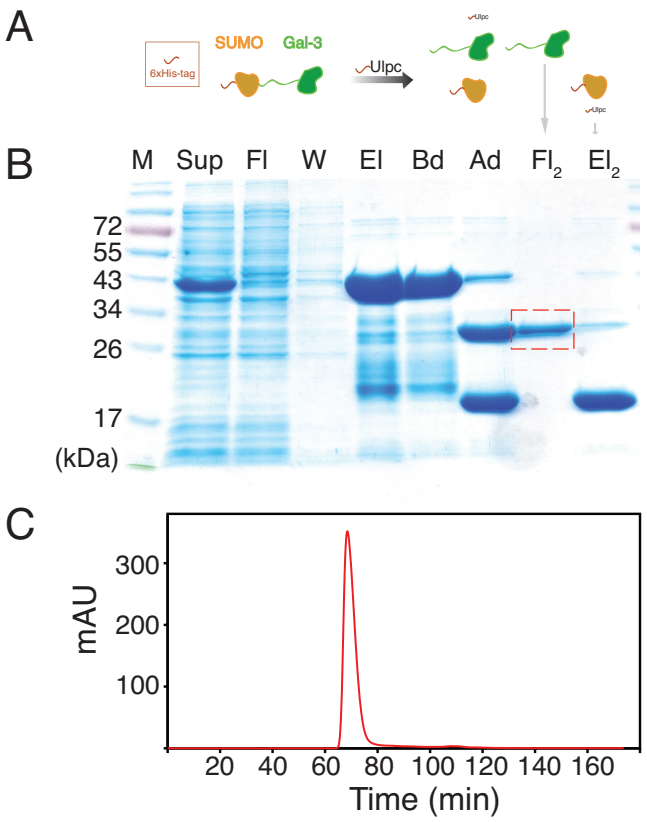

Figure S8. Purification of galectin-3. (*A*) Schematic flowchart of the purification process. (*B*) A typical SDS-PAGE gel used to confirm the purity of the sample. The lanes from left to right are: protein weight marker (M); the cell lysate supernatant (Sup); the flow-through of the nickel-charged IMAC column (Fl); the wash-through (W); the elution (El), before digestion using Ulp1 protease (Bd) and after protease digestion (Ad); the second flow-through of the protease-digested solution (Fl_2) containing the target protein; and the final elution (El_2). (*C*) The FPLC profile of the Fl_2 loaded into a gel-filtration (G75) column.

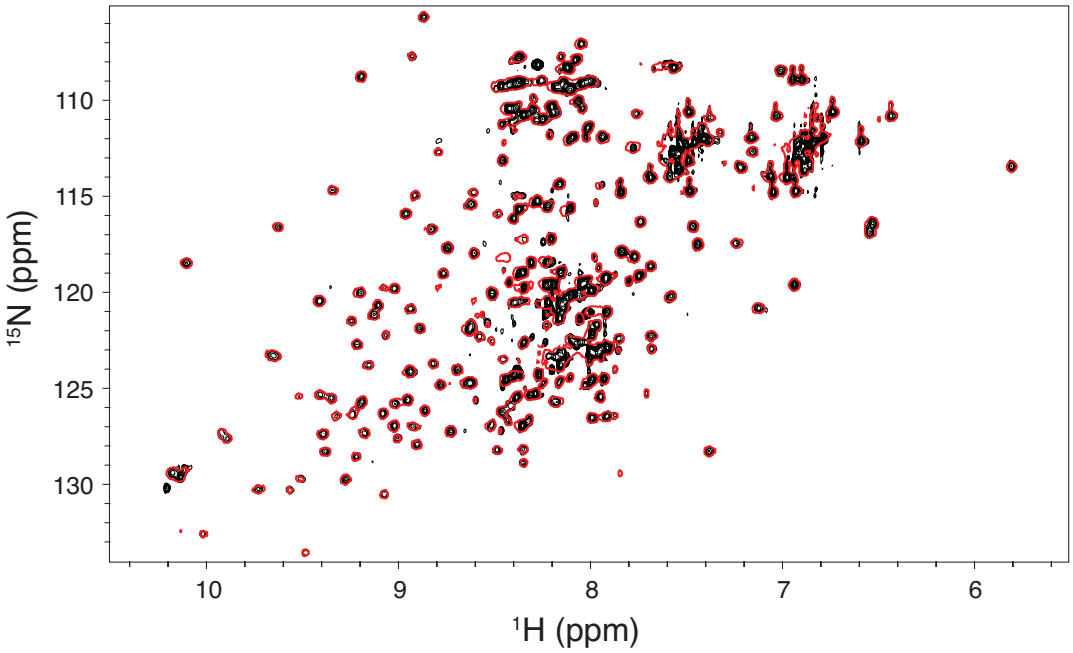

Figure S9. Comparison of the reduced-MTSL-labeled A31C sample (red) and the wild-type (black). The only differences are the mutation site and its nearest neighbors.

Table S1. Experimentally measured and extrapolated NMR dynamics parameters.[a]

Concentration (μM)	R_2 (s^{-1})	R_1 (s^{-1})	τ_c (ns)
400	31.1±4.2	0.50±0.11	17.68±2.33
200	24.9±2.7	0.56±0.08	14.86±1.37
40	20.3±2.6	0.67±0.10	12.22±1.25
Extrapolated	19.03	0.68	11.76
HYCUD	–	–	11.84±2.34

Table S2. Averaged ΔR_2 in different conditions[a] (s^{-1})

Fl[b]	$\Delta^{1\text{-}10}$	$\Delta^{1\text{-}20}$	$\Delta^{1\text{-}30}$	$\Delta^{1\text{-}40}$	$\Delta^{1\text{-}60}$	$\Delta^{1\text{-}80}$	$\Delta^{1\text{-}100}$
11.7±1.9	9.4±1.4	12.8±1.4	4.0±0.9	4.0±0.7	3.6±0.6	2.1±0.4	1.5±0.3
	+lactose[c]	+NaCl	+urea				
	13.6±4.1	17.3±3.4	7.0±1.4				
	303 K[d]	293 K					
	11.4±1.9	9.9±3.3					

[a]ΔR_2s were averaged from the values of the CRD parts. The errors were derived in two steps: (1) for each residue, its error was propagated from the original R_2's errors, which were derived from a Monte Carlo fitting procedure according to the spectrum noise; (2) the error in the table is the root-mean-squared value calculated from the error of each residue; those with large errors were removed (criteria: the error must be smaller then the subtracted value).

[b]This row shows the averaged ΔR_2 of the full-length and different truncated constructs measured at 303K using a 850 MHz spectrometer.

[c]Sample in different buffers at 303K using a 850 MHz spectrometer.

[d]This row shows the full-length construct using a 600MHz spectrometer at 303K and 293K.

Table S3. Primers used in this study.

Construct name	Template	Primer Sequence	
6xHis-SUMO-Gal3	pET-21-hGal3	Fw:	5' GGCATGGCAGACAATTTTTCGCTC 3'
		Rv:	5' ATTACTCGAGTTATATCATGGTATATGAAGC 3'
6xHis-SUMO-Gal3NTD	pET-21-hGal3	Fw:	5' GGCATGGCAGACAATTTTTCGCTC 3'
		Rv:	5' ATATCTCGAGTTACCCAGCAGGGG 3'
Δ^{1-10}	6xHis-SUMO-Gal3	Fw:	5' TTGGCGGCTTATCTGGGTCTGGAAACCCAAA 3'
		Rv:	5' CCAGATAAGCCGCCAATCTGTTCTCTGTG 3'
Δ^{1-20}	6xHis-SUMO-Gal3	Fw:	5'ATTGGCGGCGGATGGCCTGGCGCATGGG 3'
		Rv:	5'AGGCCATCCGCCGCCAATCTGTTCTCTGTGA 3'
Δ^{1-30}	6xHis-SUMO-Gal3	Fw:	5' TTGGCGGCGCTGGGGCAGGGGGCTAC 3'
		Rv:	5' GCCCCAGCGCCGCCAATCTGTTCTCTGTGAG 3'
Δ^{1-40}	6xHis-SUMO-Gal3	Fw:	5' TTGGCGGCTATCCTGGGGCCTACCCCGG 3'
		Rv:	5'CCAGGATAGCCGCCAATCTGTTCTCTGTGAGCCTCA 3'
Δ^{1-60}	6xHis-SUMO-Gal3	Fw:	5' TTGGCGGCGGCGCCTACCCTGGAGCAC 3'
		Rv:	5' TAGGCGCCGCCGCCAATCTGTTCTCTGT 3'
Δ^{1-80}	6xHis-SUMO-Gal3	Fw:	5' TTGGCGGCGGGCACCCAGCGGC 3'
		Rv:	5' GGTGGCCCGCCGCCAATCTGTTCTCTGT 3'
Δ^{1-100}	6xHis-SUMO-Gal3	Fw:	5' TTGGCGGCTACCCTGCCACTGGCCC 3'
		Rv:	5' GCAGGGTAGCCGCCAATCTGTTCTCTGT 3'
A10C	6xHis-SUMO-Gal3	Fw	5' CGCTCCATGATTGTTTATCTGGGTCTG 3'
		Rv	5' CAGACCCAGATAAACAATCATGGAGCG 3'
A31C	6xHis-SUMO-Gal3	Fw:	5' ACCAGCCTTGTGGGGCAGG 3'
		Rv:	5' CCTGCCCCACAAGGCTGGT 3'
A49C	6xHis-SUMO-Gal3	Fw:	5' CCTACCCCGGGCAGTGTCCCCCAGGGGCTTA 3'
		Rv:	5' TAAGCCCCTGGGGGACACTGCCCGGGGTAGG 3'
A100C	6xHis-SUMO-Gal3	Fw	5' GTGCCACCGGATGTTACCCTGCCAC 3'
		Rv	5' GTGGCAGGGTAACATCCGGTGGCAC 3'
I250C	6xHis-SUMO-Gal3	Fw	5' TCATATACCATGTGCTAACTCGAGCAC 3'
		Rv	5' GGTATATGAAGCACTGGTGAGGTCTAT 3'