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

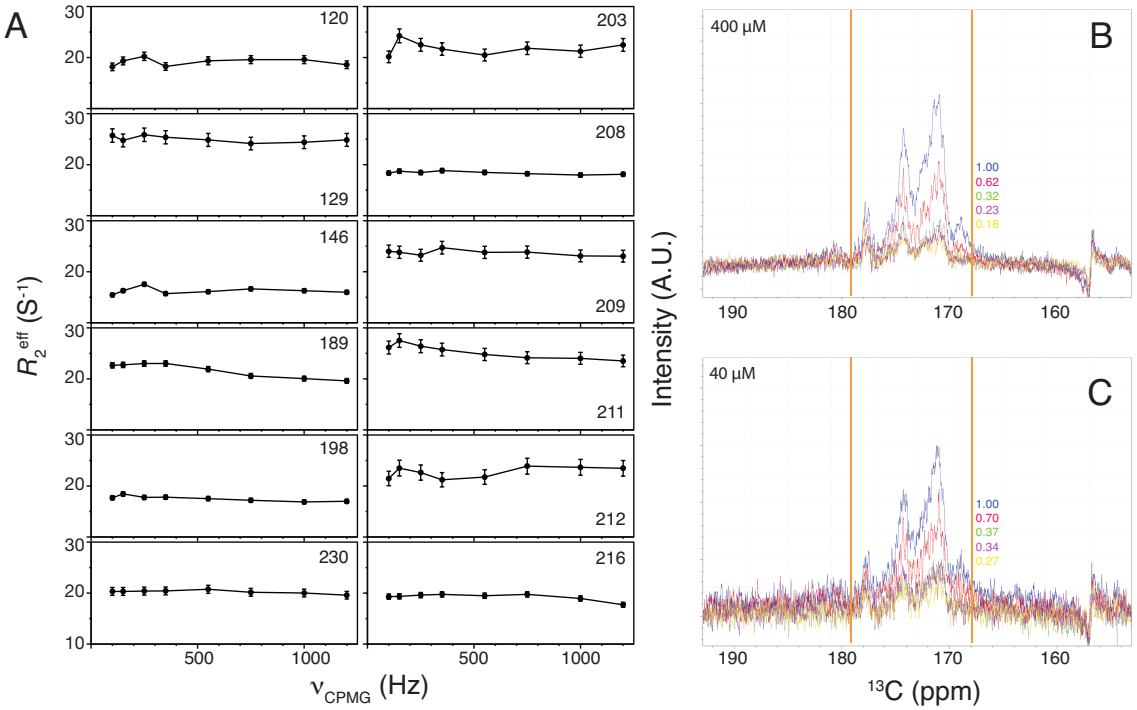

Figure S1. (*A*) Typical ^{15}N Carr-Purcell-Meiboom-Gill relaxation dispersion curves for 400 µM full-length galactin-3. The residue number is indicated in each panel. (*B,C*) The one-dimensional ^{13}C spectrum with different T_2 delays (blue: 20 ms, red: 40 ms, green: 60 ms; purple: 80 ms; yellow: 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.

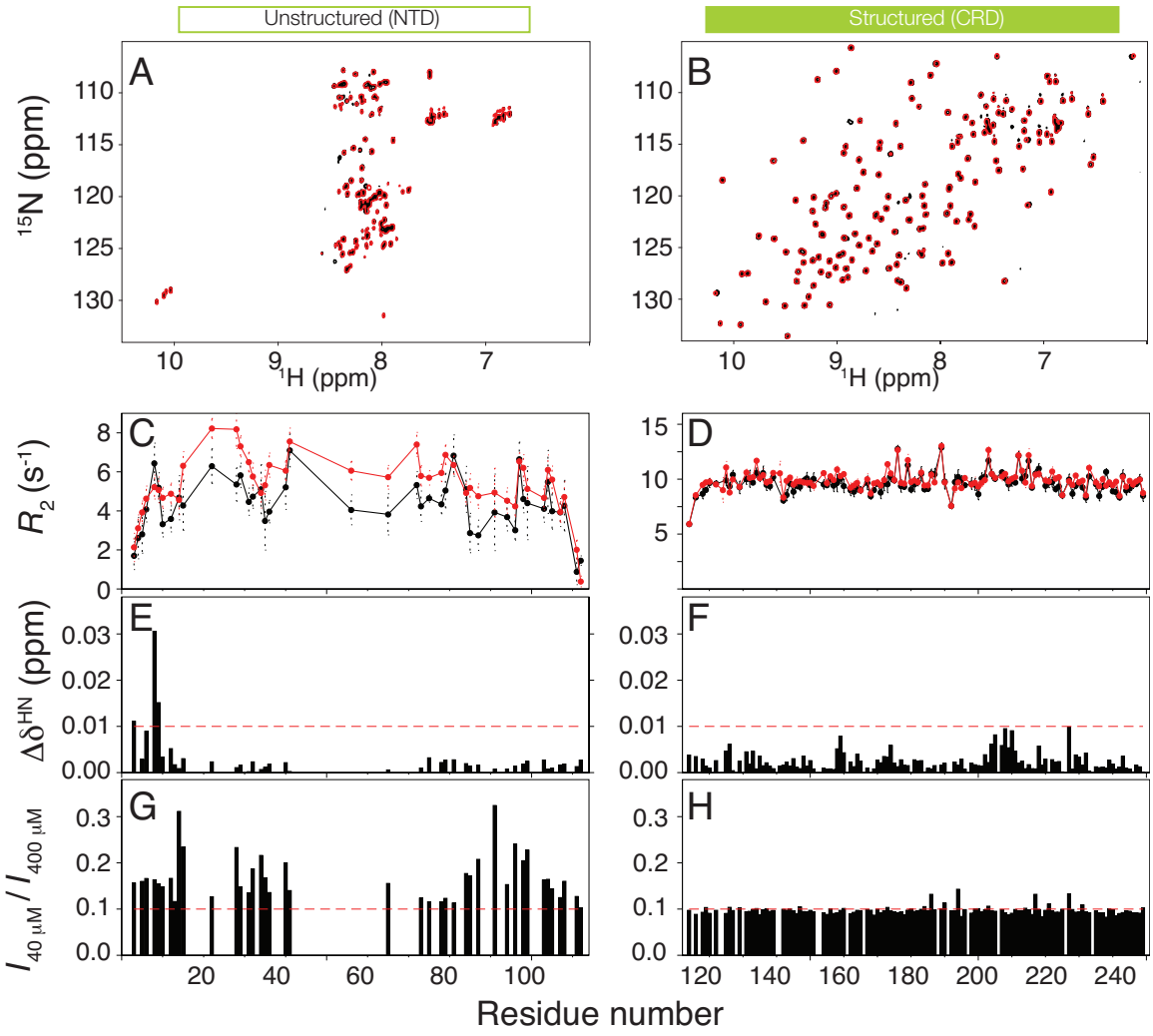

FIGURE S2. (*A,B*) HSCQ spectra, (*C,D*) transverse relaxation rate constants, (*E,F*) chemical shift perturbation, and (*G,H*) ratios of peak intensities ratio for (*A,C,E,G*) the N-terminal domain (NTD) of galectin-3 alone, and (*B,D,F,G*) the carbohydrate recognition domain (CRD) alone. In parts (*A–D*),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

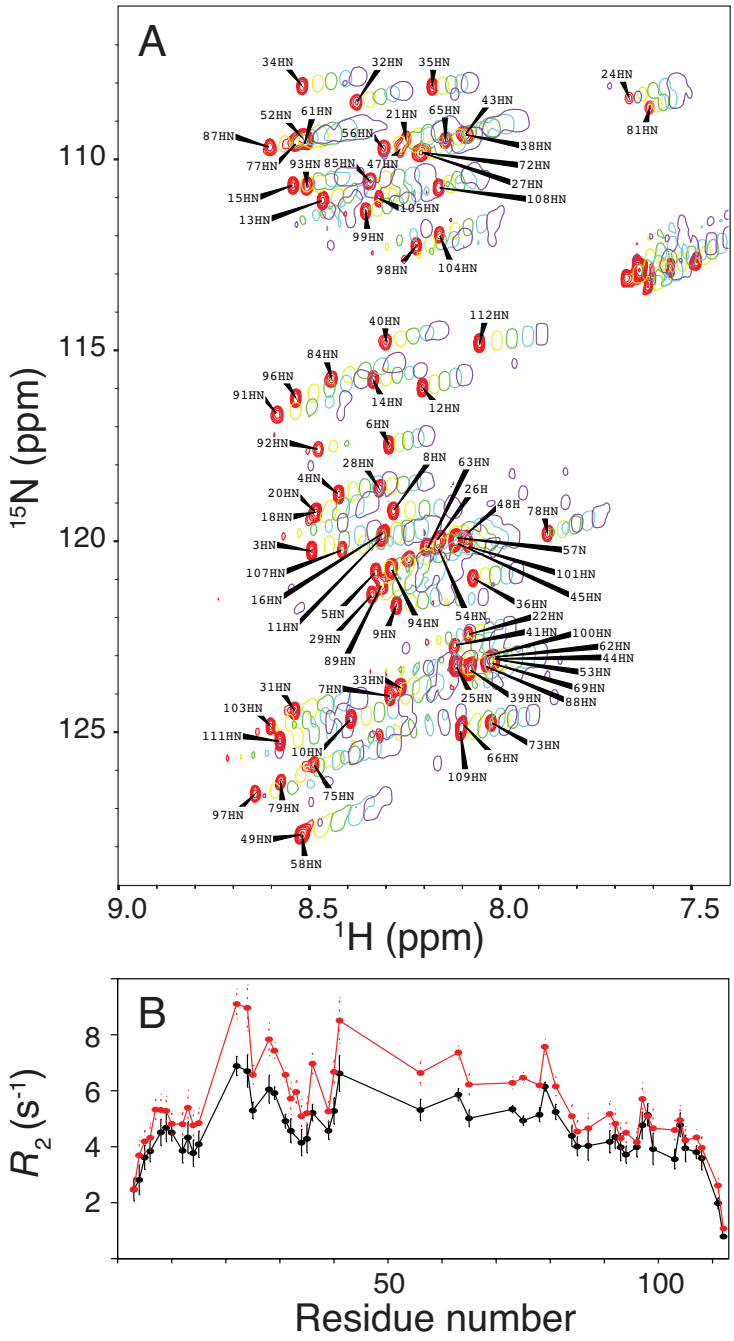

Figure S3. (*A*) Assigned HSQC spectrum of the N-terminal domain (NTD) of galectin-3 at 283 K (red) overlaid on the spectra recorded at 288 (yellow), 293 (green), 298 (cyan), and 303 K (purple). (*B*) NMR transverse relaxation rates (R_2) of the NTD collected at 283 K on a 850 MHz spectrometer for 40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

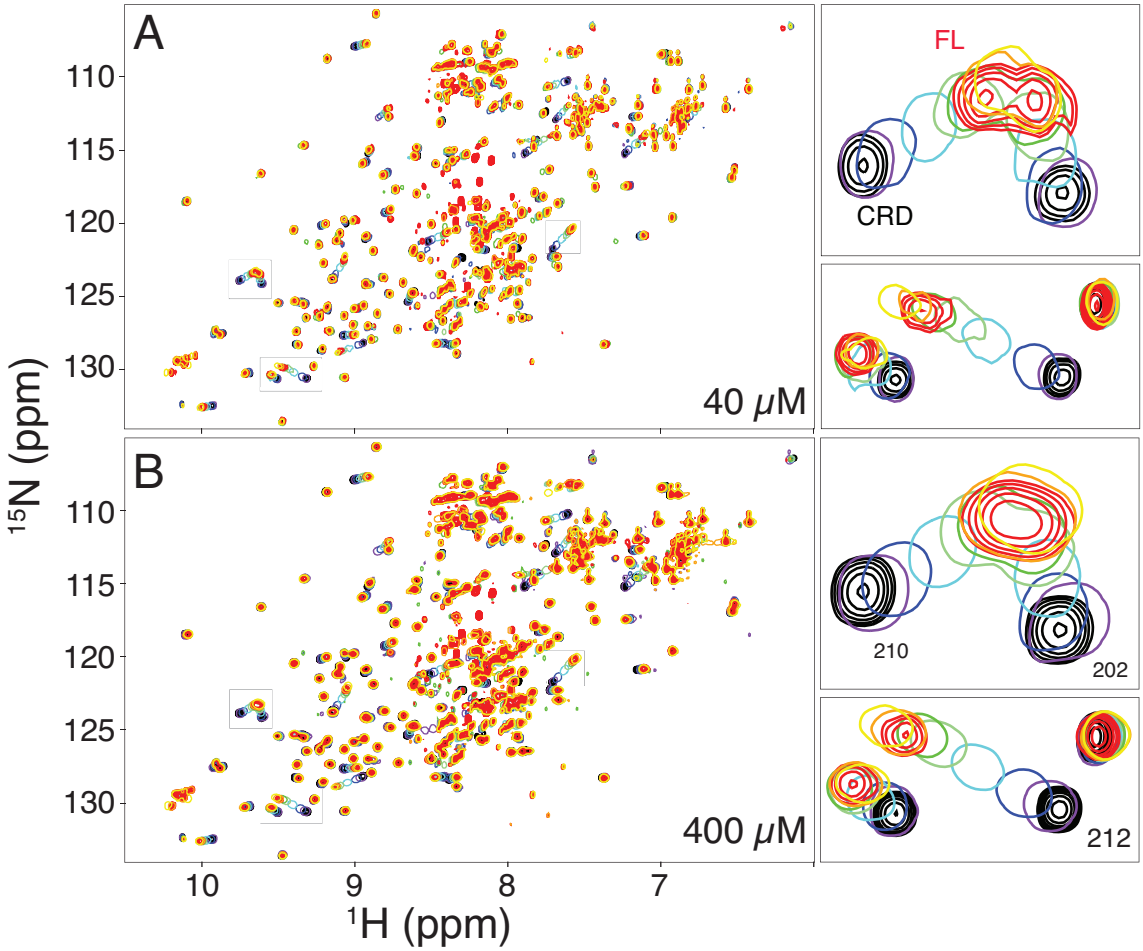

Figure S4. Overlay of the HSQC spectra of (*A*) 40 and (*B*) 400 μM samples of full-length (red), $\Delta^{1\text{-}10}$ (orange), $\Delta^{1\text{-}20}$ (yellow), $\Delta^{1\text{-}30}$ (dark green), $\Delta^{1\text{-}40}$ (light green), $\Delta^{1\text{-}60}$ (cyan), $\Delta^{1\text{-}80}$ (blue), 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.

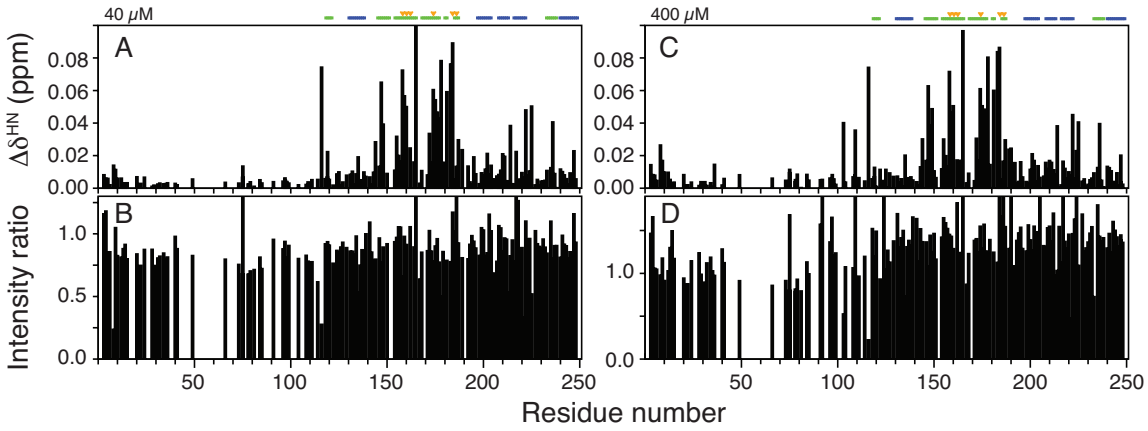

Figure S5. (*A,C*) Chemical shift differences and (*B,D*) ratios of HSQC peak intensities between samples of galectin-3 with and without 250 mM lactose at a protein concentration of (*A,B*) 40 and (*C,D*) 400 μM. The most prominent chemical shift perturbations as expected occur around the carbohydrate-binding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

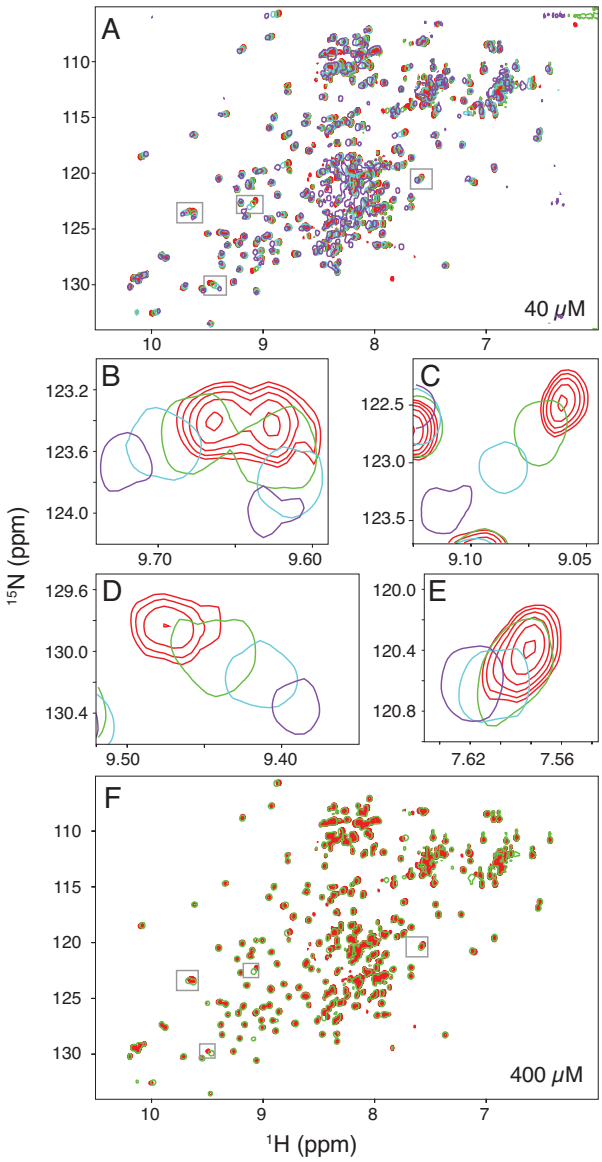

Figure S6. (*A*) Overlay of HSQC spectra of 40 μM samples of galectin-3 without urea (red) or in the presence of 0.8 (green), 2 (cyan), and 4 M (purple) urea (the latter spectrum is of lesser quality because the protein becomes denatured). (*B–E*) Expanded views of the peaks whose positions change the 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*.

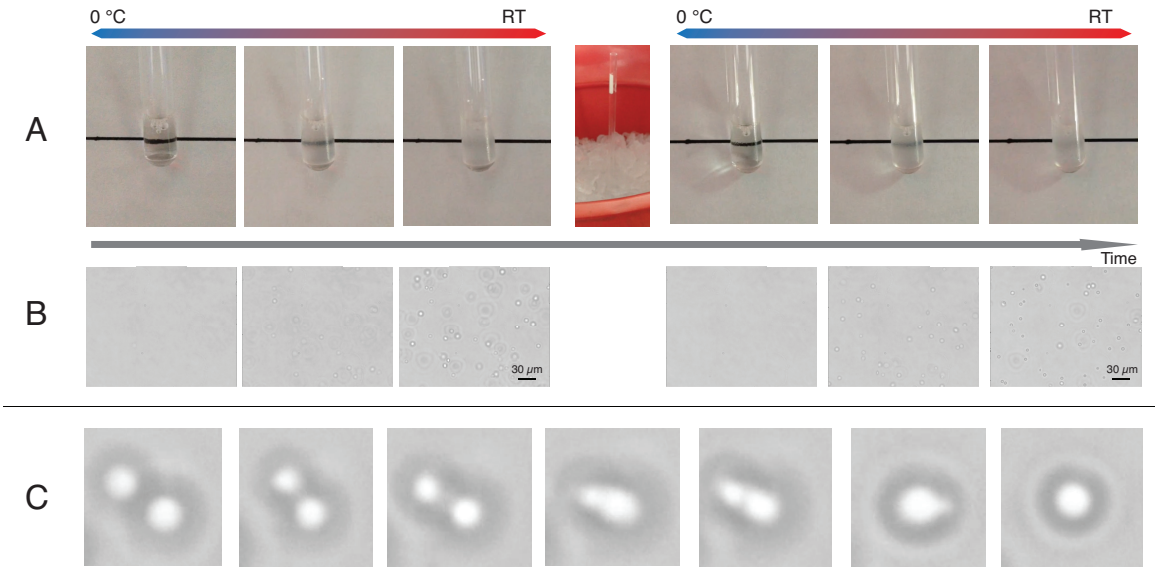

Figure S7. (*A*) Photographs and (*B*) microscope images of liquid-liquid phase separation of the N-terminal domain of galectin-3 (400 μM protein sample in the presence of 300 mM NaCl). The sample, initially transparent at ~0 °C (in iced water), becomes clouded when left for ~90 s at room temperature.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

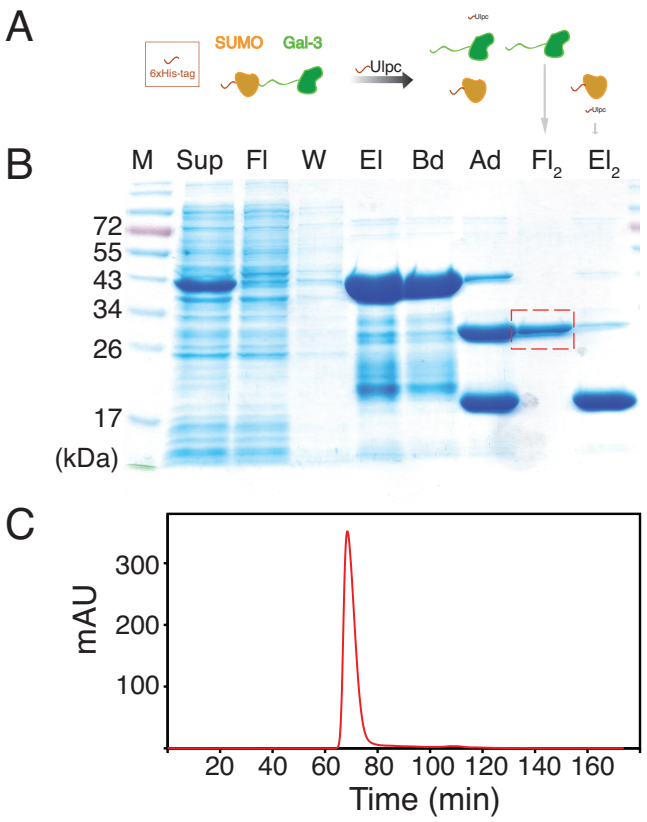

Figure S8. Purification of galectin-3. (*A*) Schematic flowchart of the purification process. (*B*) A typical SDS-PAGE gel used to confirm the purity of the sample. The lanes from left to right are: protein weight marker (M); the cell lysate supernatant (Sup); the flow-through of the nickel-charged IMAC 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.

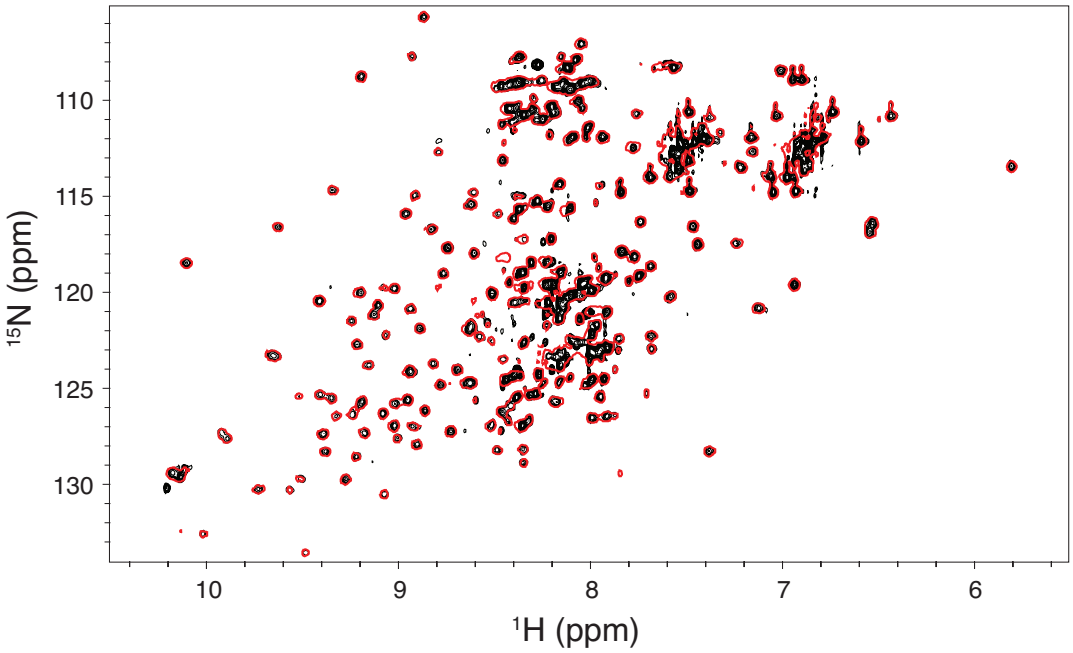

Figure S9. Comparison of the reduced-MTSL-labeled A31C sample (red) and the wild-type (black). The only differences are the mutation site and its nearest neighbors.

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'