Supporting Information

The Importance of the Compact Disordered State in the Fuzzy Interactions between Intrinsically Disordered Proteins

Dan Wang[1,§], Shaowen Wu[2,§], Dongdong Wang[3,§], Xingyu Song[1], Maohua Yang[1], Wolun Zhang[4], Shaohui Huang[5,6] Jingwei Weng[1]*, Zhijun Liu[7]*, Wenning Wang[1]*

Methods

FCS experiments and data analysis. Molecular hydrodynamic radii were measured using a bench-top FCS instrument (i.e., CorTector SX100; LightEdge Technologies Ltd., Zhongshan, China) equipped with two cw-lasers (488 nm and 638 nm) and an Olympus 60X NA1.2 water immersion objective. Briefly, the instrument was calibrated using a 10 nM solution of Atto655-carboxylic acid with a reported diffusion coefficient of 426 ± 8 $\mu m^2 \cdot s^{-1}$. [1] High precision coverslips (No. 1,5H, Deckglaser; Sigma Aldrich) were immersed in whole milk overnight, rinsed two times with ddH_2O, and air dried at room temperature. Such a coverslip was balanced directly on top of the Olympus objective, upon which a 100 µl sample of 25 nM Alexa647 labeled 4.1G-CTD protein (labeled at residue 982) was added. Subsequently, 100 FCS measurements were performed using the 638 nm laser with each experiment lasting 5 seconds; two more sets of such 100 FCS measurements were performed using two new coverslips. Auto-correlation curves from the above 300 FCS measurements were analyzed using the Correlation Analysis software (LgihtEdge Technologies Limited) and the following mathematic model:

$$G(\tau) = \frac{1}{N} \cdot \frac{(1 - T + Te^{-\frac{\tau}{\tau_T}})}{1 - T} \cdot \frac{1}{(1 + \frac{\tau}{\tau_D})\sqrt{1 + \frac{\tau}{\tau_D S^2}}}$$

N is the number of fluorescently labeled sample molecules in the FCS volume, τ_D is the characteristic diffusion correlation time of the sample molecule, S is the structure parameter, T is the fraction of fluorophores residing in the triplet state, and τ_T is the triplet lifetime. From each τ_D value derived from each 5s-FCS experiment, we can

calculate a corresponding hydrodynamic radius of the sample molecule using following equations:

$$\tau_D = \frac{r_0^2}{4D} \quad \text{and} \quad D = \frac{k_B T}{6\pi\eta R_H}$$

r_0 is the lateral dimension of the FCS volume, which is measurable using a calibration dye of known diffusion coefficient (see above); D is the diffusion coefficient of the sample molecule, k_B is Boltzmann constant, T is absolute experimental temperature, η is the viscosity of the sample buffer, and R_H is the hydrodynamic radius of the sample molecule. FCS is a quasi-single-molecule technique, thus each 5s experiment samples a slightly different pools of sample molecules of different hydrodynamic radii; Thus, 300 such FCS experiments yield a size distribution of the sample molecules (Supplemental Figure 12).

Table S1 ITC measurements of 4.1G-CD/NuMA binding affinity at various NaCl concentrations

[NaCl] (mM)	0	100	200	300	400	600
pH	8.0	8.0	8.0	8.0	8.0	8.0
K_D (□M)	0.92±0.13	1.96±0.33	2.93±0.53	3.99±0.27	4.84±0.37	4.40±0.28
ΔH kcal/mol	-9.66±0.19	-10.8±0.34	-11.1±0.46	-10.7±0.18	-10.8±0.23	-11.4±0.19
-$T\Delta S$ kcal/mol	1.43	2.97	3.59	3.30	3.51	4.13
n	0.578	0.539	0.605	0.646	0.664	0.658

Table S2. The number of traversals from the lowest to highest temperature and *vice versa* of replicas in the REMD simulation of 4.1G-CTD at pH3.6

Replica	Traversals low→high	Traversals high→low	Replica	Traversals low→high	Traversals high→low	Replica	Traversals low→high	Traversals high→low
0	3	2	16	3	2	32	1	1
1	4	3	17	0	0	33	3	3
2	3	2	18	2	2	34	0	0
3	1	1	19	1	1	35	1	1
4	0	0	20	2	1	36	1	1
5	1	1	21	2	3	37	2	2
6	0	0	22	2	2	38	4	5
7	4	4	23	0	1	39	1	1
8	2	1	24	0	0	40	2	3
9	1	0	25	0	0	41	1	1
10	2	1	26	2	3	42	1	0
11	0	1	27	3	2	43	1	1
12	2	2	28	2	2	44	2	3
13	2	2	29	0	0	45	0	1
14	2	1	30	0	1	46	0	1
15	0	0	31	2	2	47	2	3
Average time for the round trip						137.14 ns		

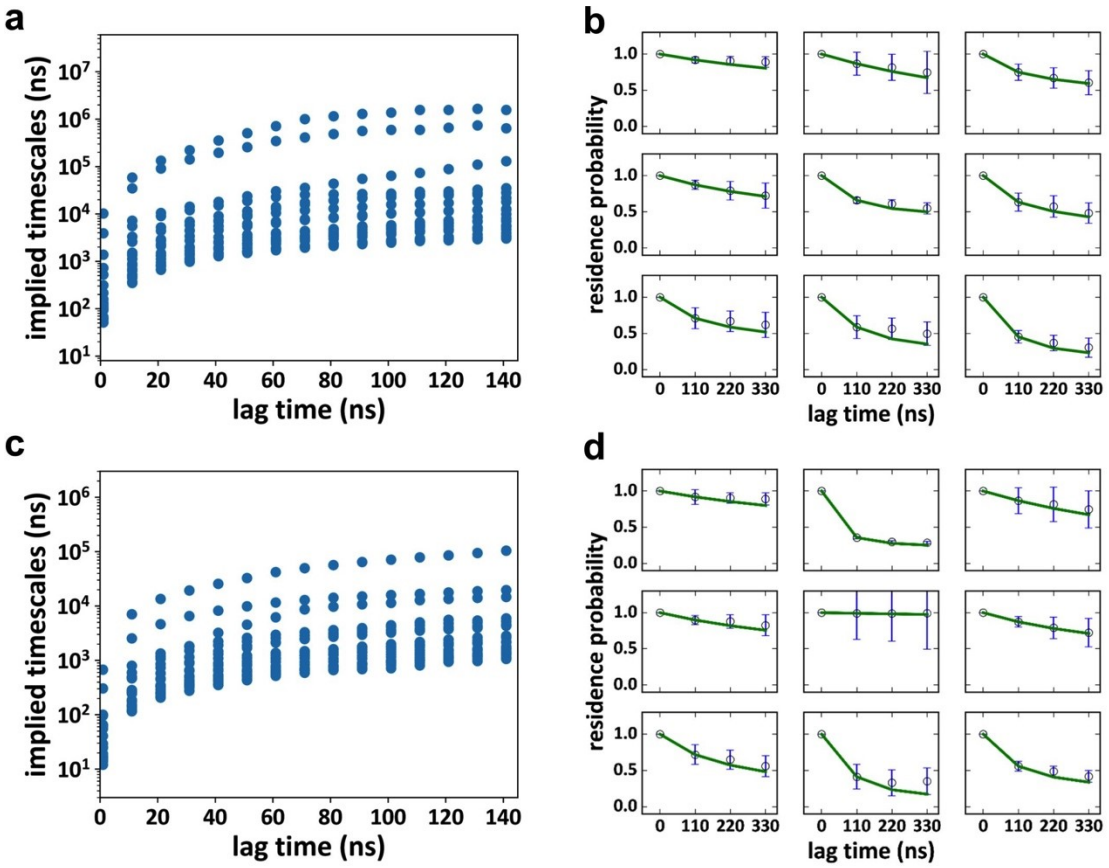

Figure S1. Validation of the Markov state models (MSMs). Variations of the implied timescale with the lag time in the 1200-state MSM (**a**) and 200-state MSM (**b**). CK tests of the 1200-state MSM (**c**) and 200-state MSM (**d**).

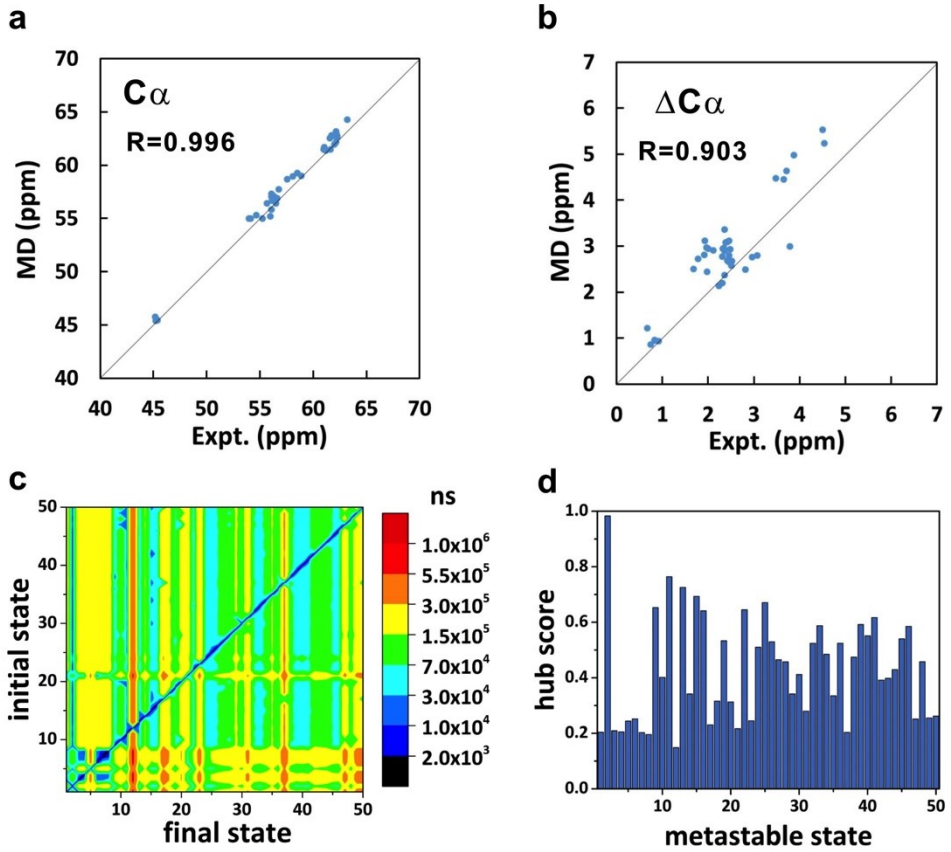

Figure S2. Agreements between the experimental and MSM-predicted Cα chemical shifts (**a**) and secondary chemical shifts (**b**). **c**. The MFPTs of the inter-state transitions among the first 50 macrostates of the 200-state MSM of 4.1G-CTD. **d**. The hub scores of the first 50 macrostates of the MSM.

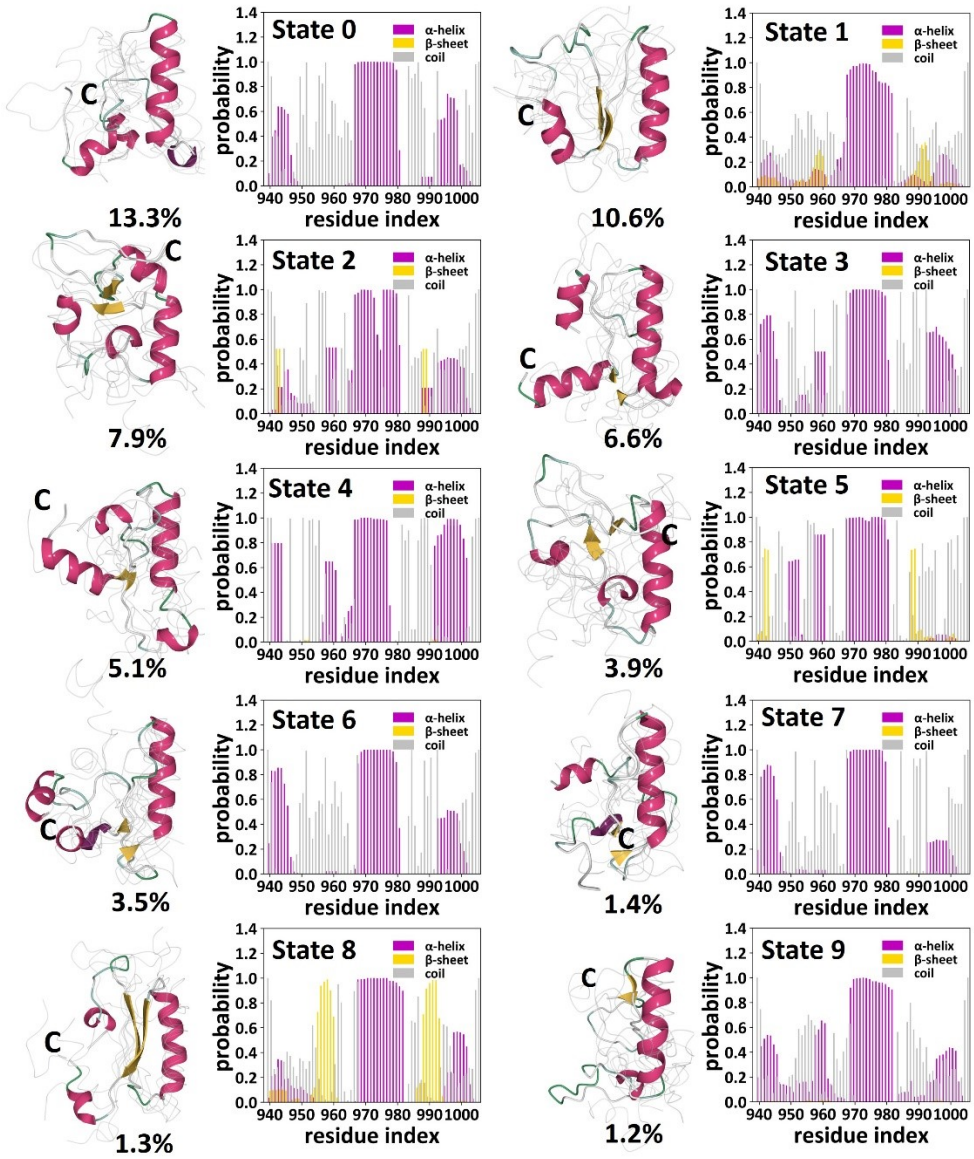

Figure S3. The central structure (cartoon representation) superimposed with other four structures (ribbon representation) in each macrostate and the secondary structure contents of the first ten macrostates of the 200-state MSM of 4.1G-CTD**.** The PDB files of each state are available through github (https://github.com/dongdawn/4.1G-NuMA/tree/master/4.1G-CTD-MSM). The structures are drawn using the web software Hermite (https://hermite.dp.tech/), α-helix, β-sheet, coil and turn are colored magenta, yellow, white and green, respectively.

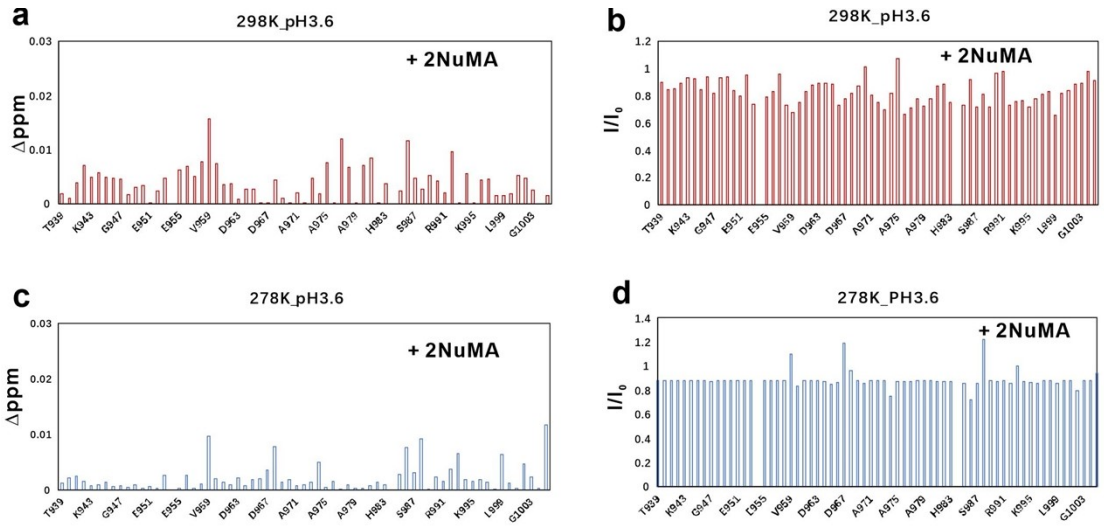

Figure S4. The changes of NMR peak intensities and chemical shifts of 4.1G-CTD upon addition of NuMA at pH 3.6 (a-d).

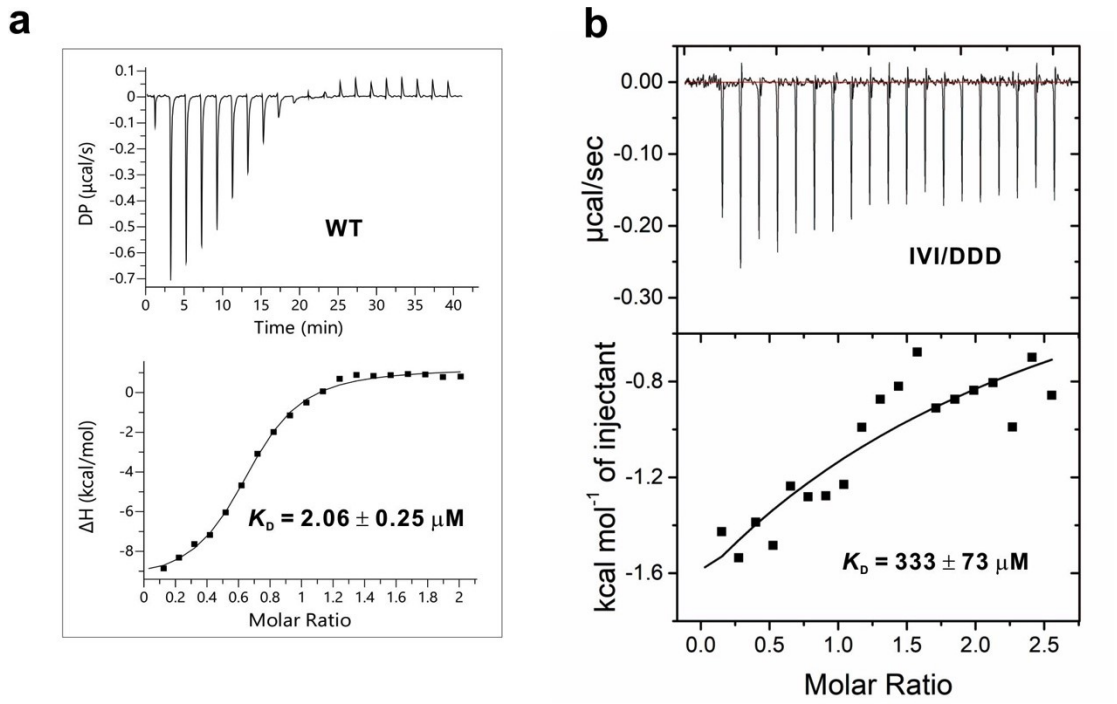

Figure S5. ITC measurements of NuMA binding affinities of WT and IVI/DDD mutant of 4.1G-CTD.

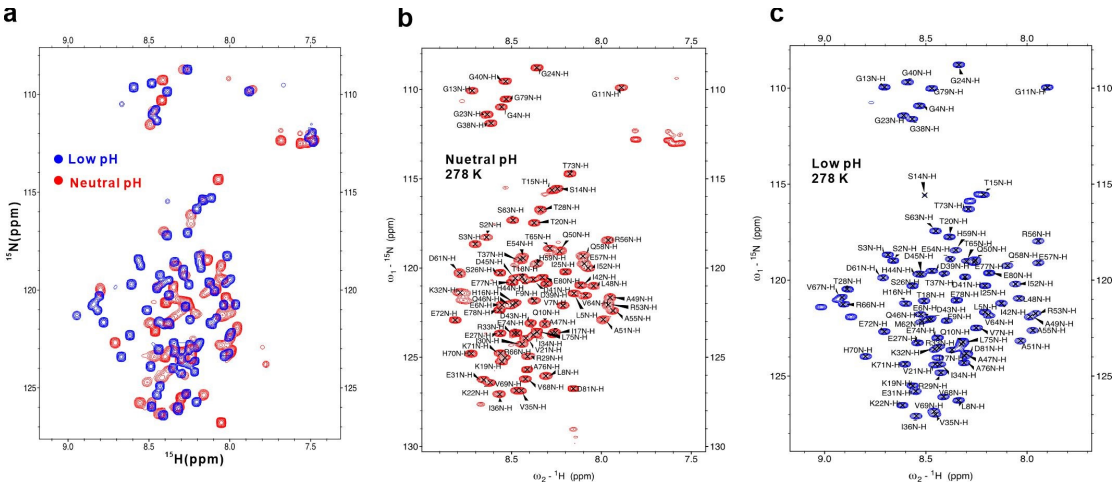

Figure S6. (a) ^{1}H-^{15}N HSQC spectra of 4.1G-CTD at two pH conditions and 278 K. ^{1}H-^{15}N HSQC spectrum and backbone assignment of 4.1G-CTD under neutral pH (**b**) and low pH (**c**) buffers at 278 K.

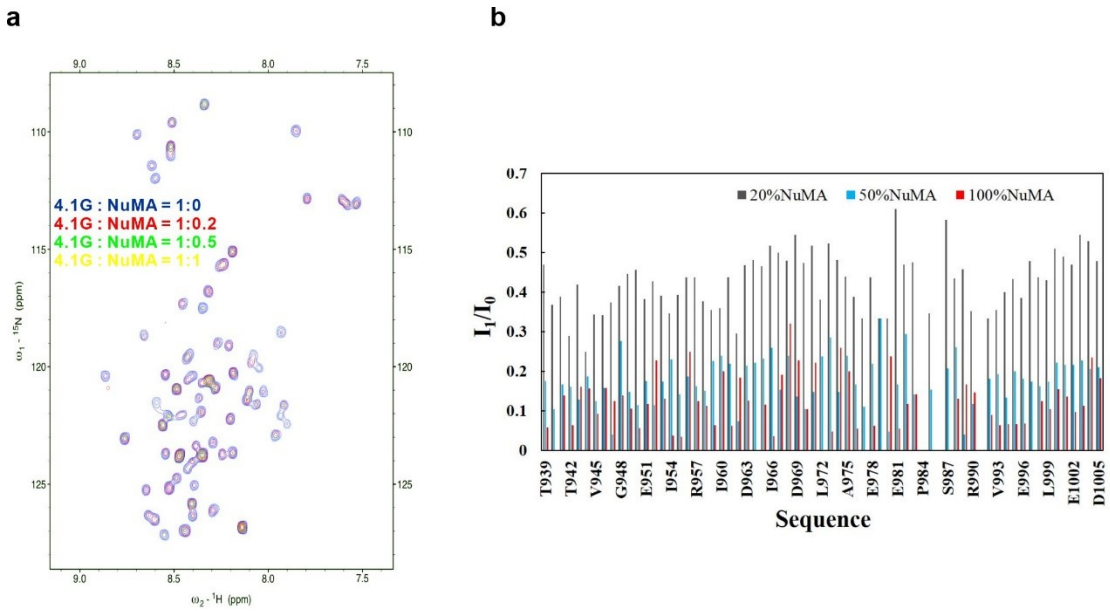

Figure S7. Interaction between 4.1G-CTD and NuMA at 278 K. (a) ^{1}H-^{15}N HSQC spectra of 4.1G-CTD titrated with various molar ratio of NuMA. (b) Changes of peak intensities of 4.1G-CTD upon NuMA addition at various molar ratios.

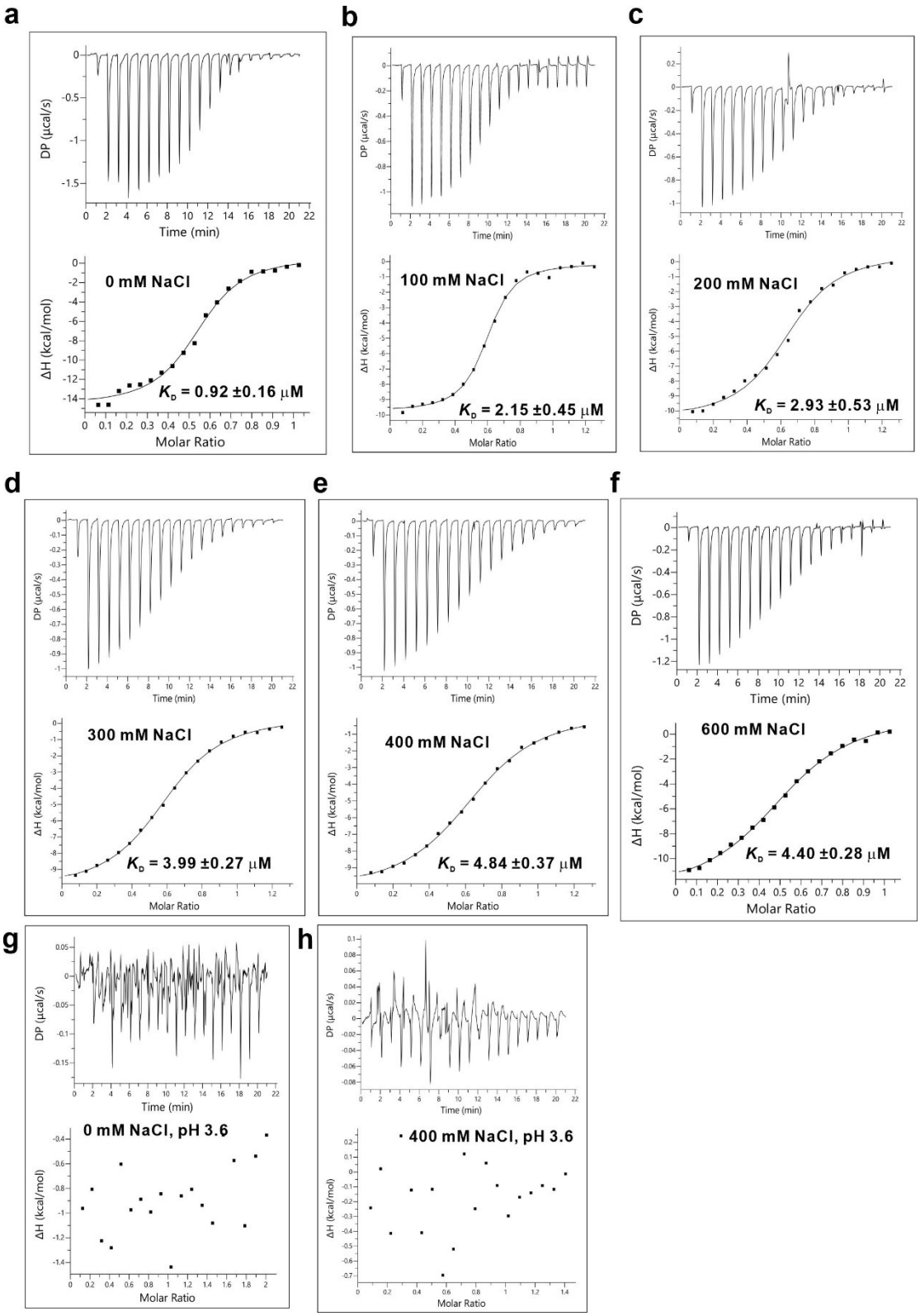

Figure S8. Interaction between 4.1G-CTD and NuMA with different ion strengths and pH conditions measured by ITC titrations. (a)~(f) Measurement at neutral pH and NaCl concentrations of 0, 100, 200, 300, 400 and 600 mM, respectively. (g) and (h) Measurement at pH 3.6 and NaCl concentrations of 0 and 400 mM, respectively.

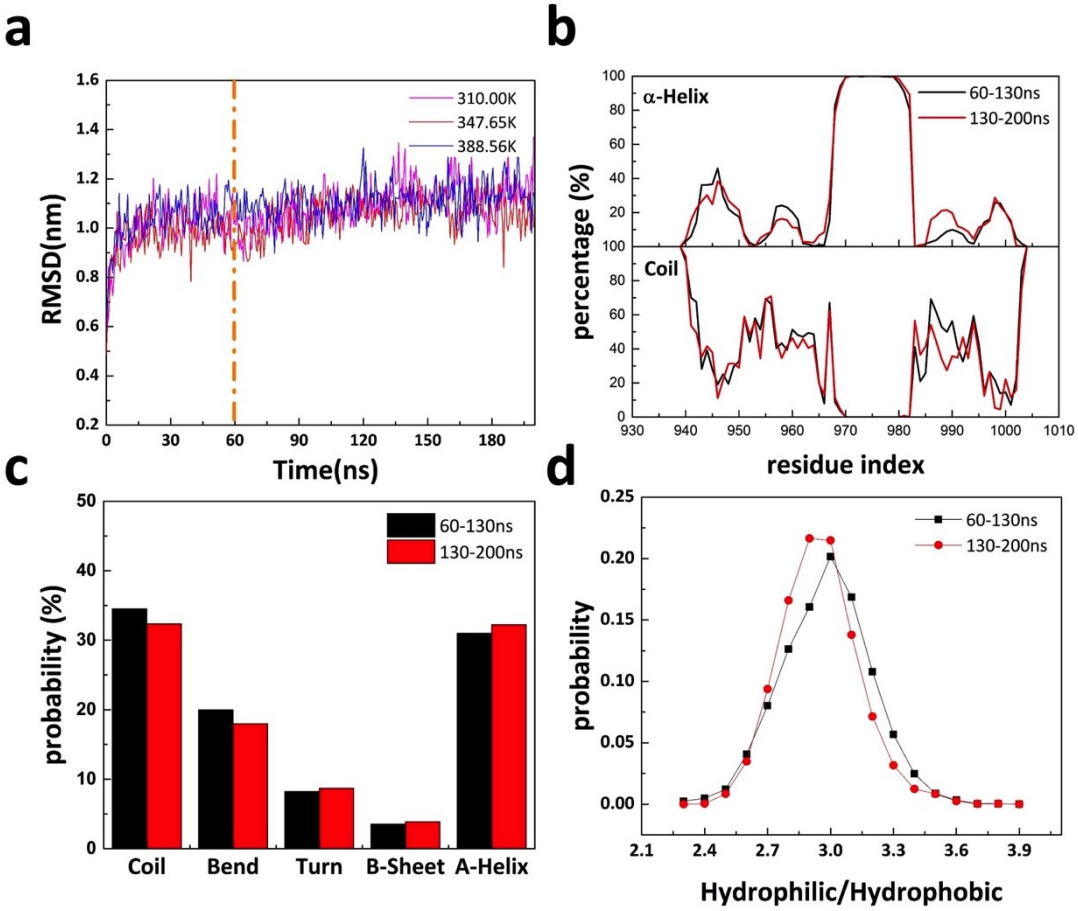

Figure S9. Convergence of the REMD simulation of 4.1G-CTD at pH 3.6. (a) Time evolutions of backbone root mean square deviations (RMSDs) relative to the initial structure at three temperatures. (**b**) The percentages of α-helix and coil as a function of residues at two simulation time intervals (60-130 ns and 130-200 ns). (**c**) Probability of main secondary structures. (**d**) The probability distributions of hydrophilic to hydrophobic SASA ratio.

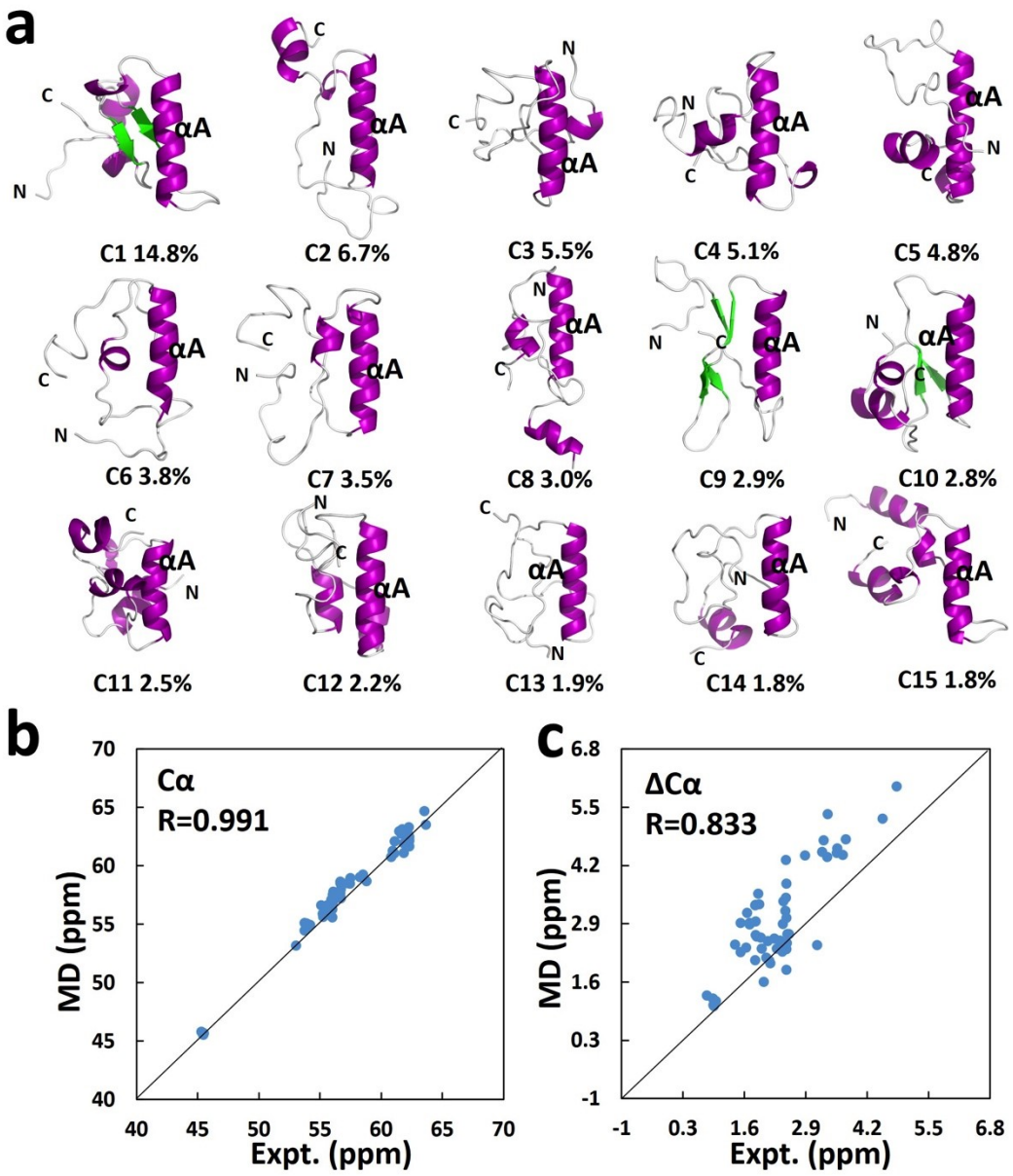

Figure S10. REMD simulation derived conformational ensemble of 4.1G-CTD at pH3.6. (**a**) Representative structures of the first 15 clusters of the structure ensemble of 4.1G-CTD based on REMD simulations at pH 3.6. The structure α-helix, β-sheet and coil are colored magenta, green and white, respectively. The correlations between experimental and calculated NMR Cα chemical shifts (**b**) and secondary chemical shifts **(c).**

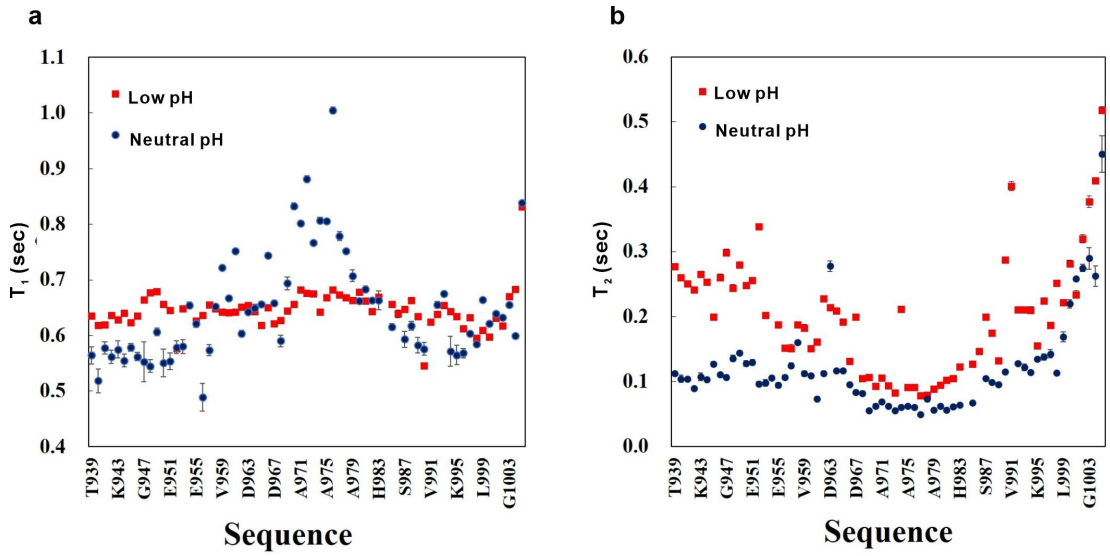

Figure S11. Relaxation time T_1 (a) and T_2 (b) of 4.1G-CTD at both pH conditions measured at 278 K.

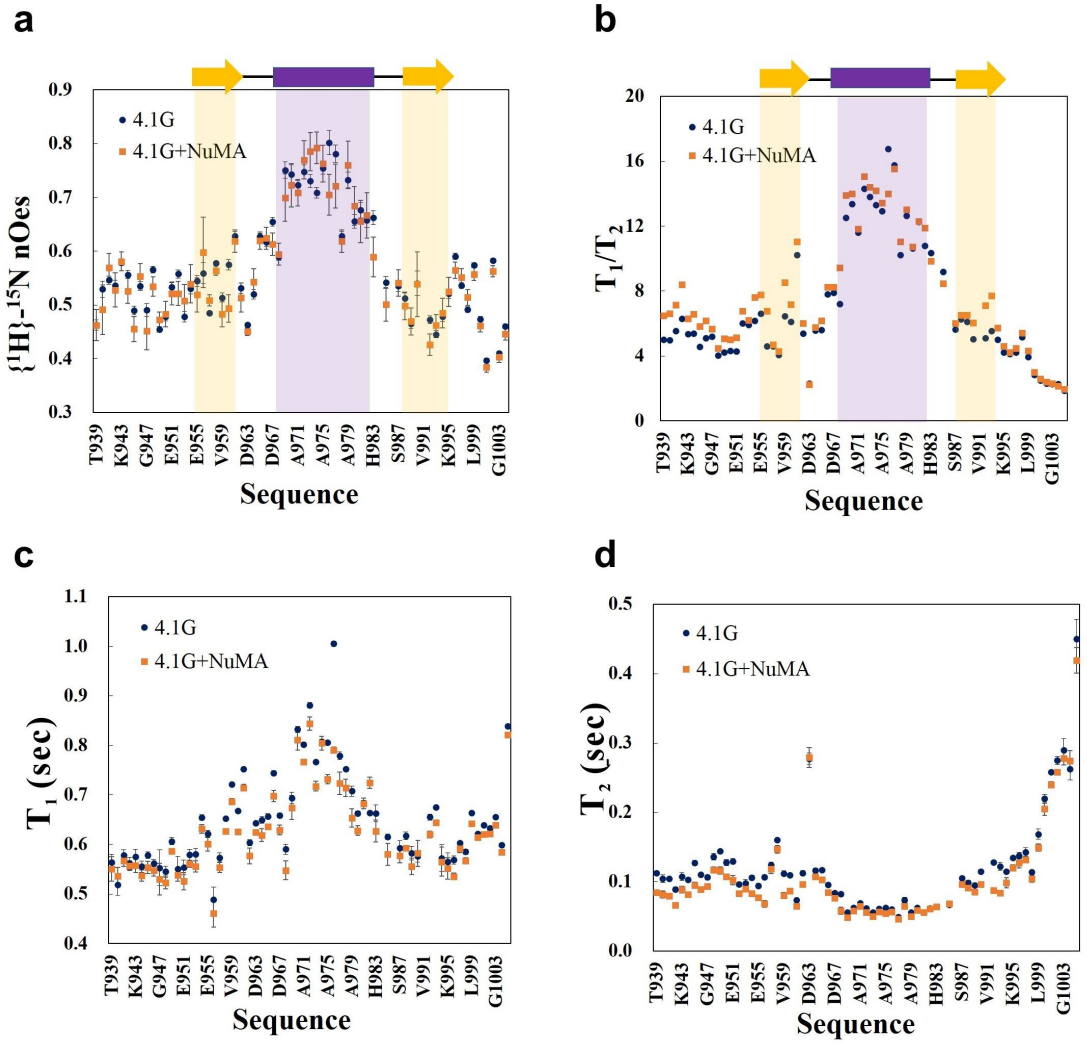

Figure S12. Comparison of the backbone dynamics of 4.1G-CTD and 4.1G-CTD/NuMA complex. a. Heteronuclear ^{1}H-^{15}N NOEs of 4.1G-CTD and 4.1G-CTD/NuMA complex**. b**. T_1/T_2 of 4.1G-CTD and 4.1G-CTD/NuMA complex. c. T_1 values of 4.1G-CTD and 4.1G-CTD/NuMA complex. **d**. T_2 values of 4.1G-CTD and 4.1G-CTD/NuMA complex

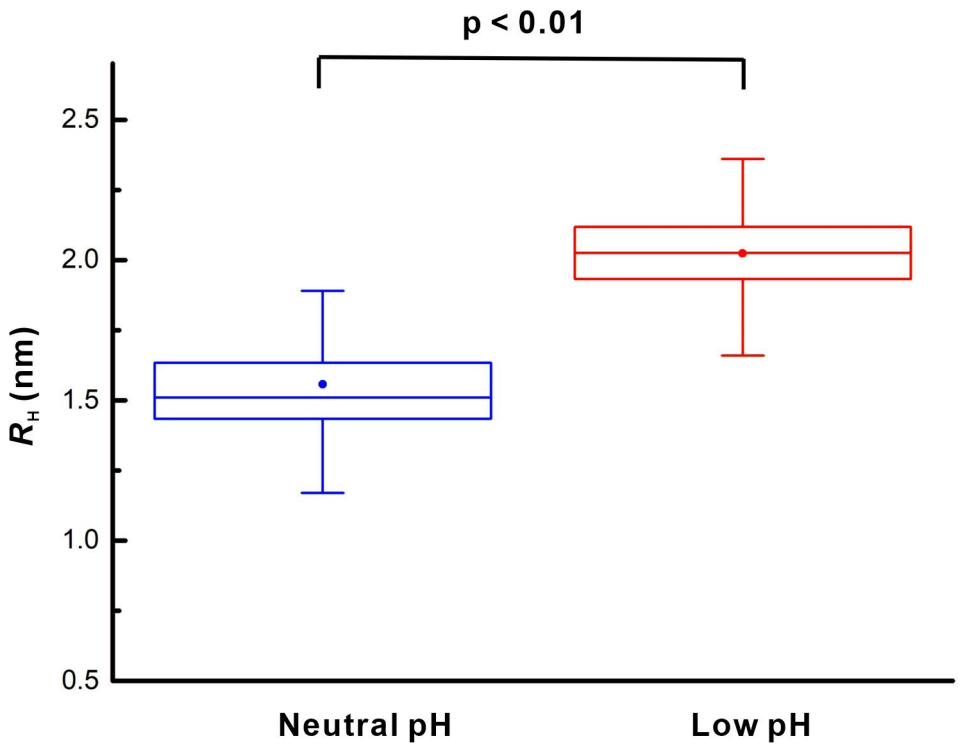

Figure S13. Hydrodynamics radius of 4.1G-CTD at neutral and low pH measured by FCS. The box plot shows the R_H distributions, with the dots denoting mean values. p = 6.57E-69 < 0.01, one-way ANOVA and Bonferroni test.

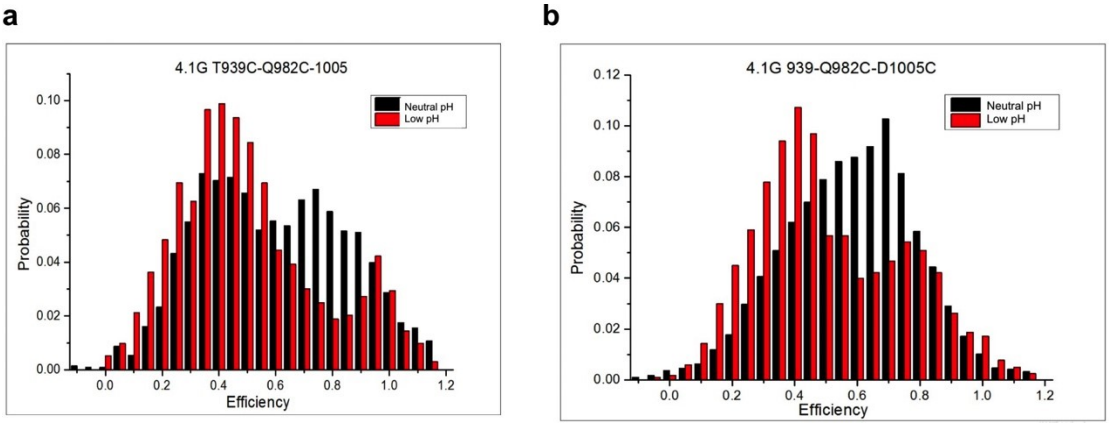

Figure S14. smFRET efficiency distributions of 4.1G-CTD at neutral and low pH conditions. a. System with dyes labeled on residue 939 and 982. **b**. System with dyes labeled on residue 982 and 1005.

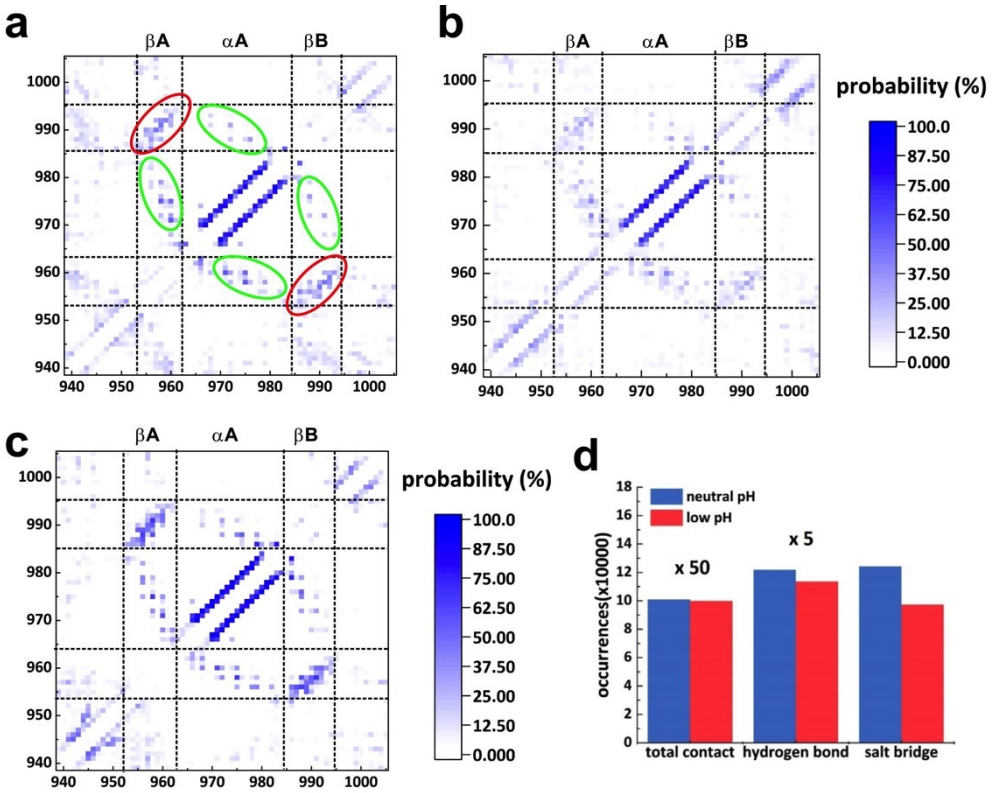

Figure S15. Intra-molecular contact maps of 4.1G-CTD at neutral pH (a), low pH (b) and 4.1G-CTD/NuMA complex (c) based on MD simulation trajectories. Red ovals highlight the contact probabilities between βA and βB, and green ovals highlight the contact probabilities between βA/βB and αA. (d) Comparison of intramolecular interaction numbers at two pH conditions.

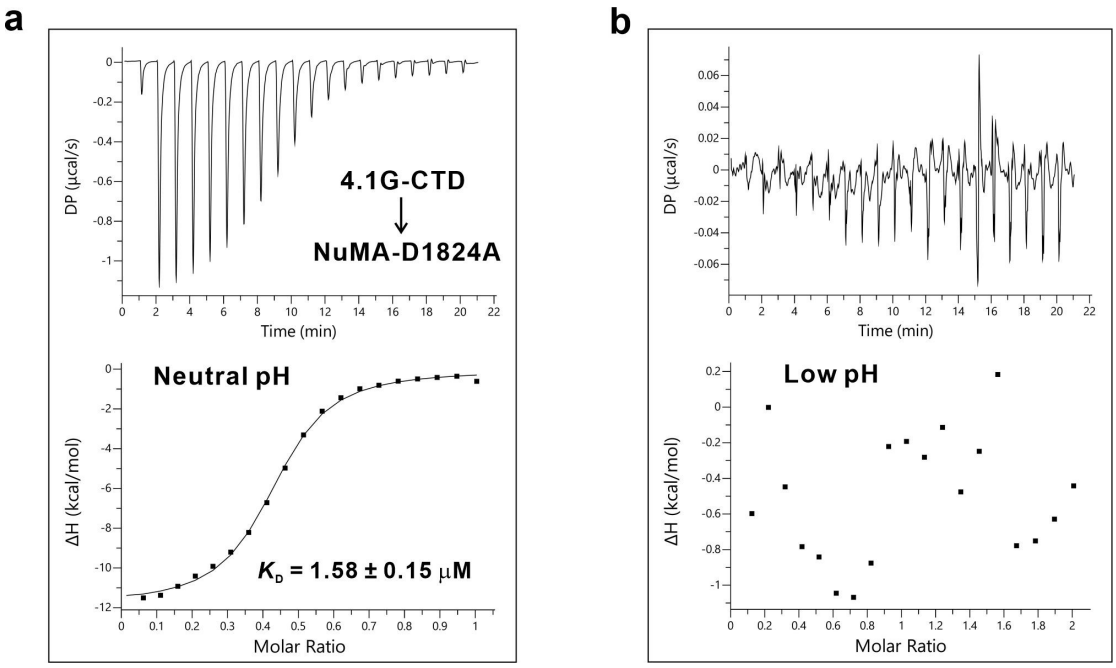

Figure S16. The interaction between 4.1G-CTD and NuMA-D1824A mutant at neutral (a) and low pH (b) conditions measured by ITC.

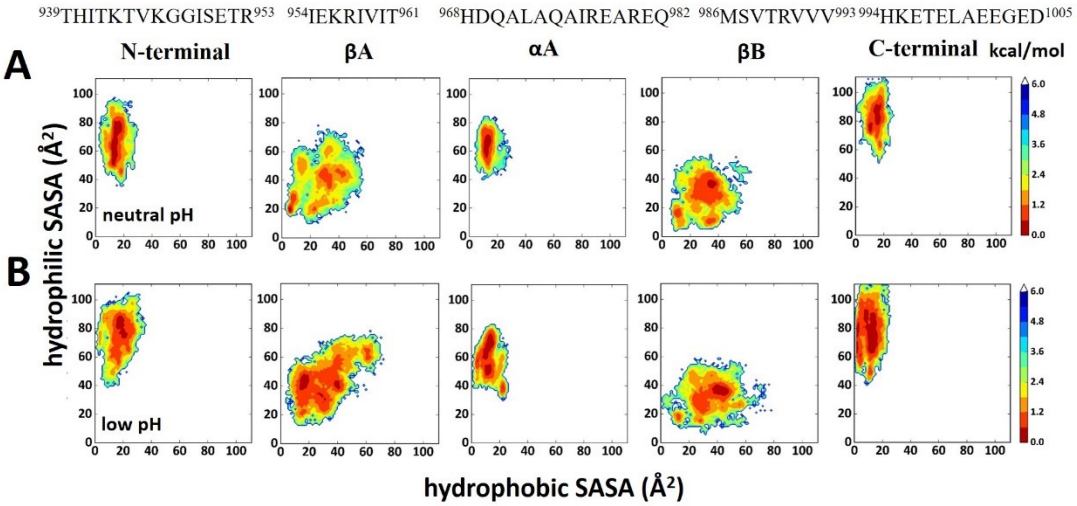

Figure S17. Free energy landscapes of each region as a function of hydrophilic and hydrophobic SASAs at neutral pH (A) and low pH (B). The SASA is an average value, which means dividing the total SASA of each region by the number of residues of this region.

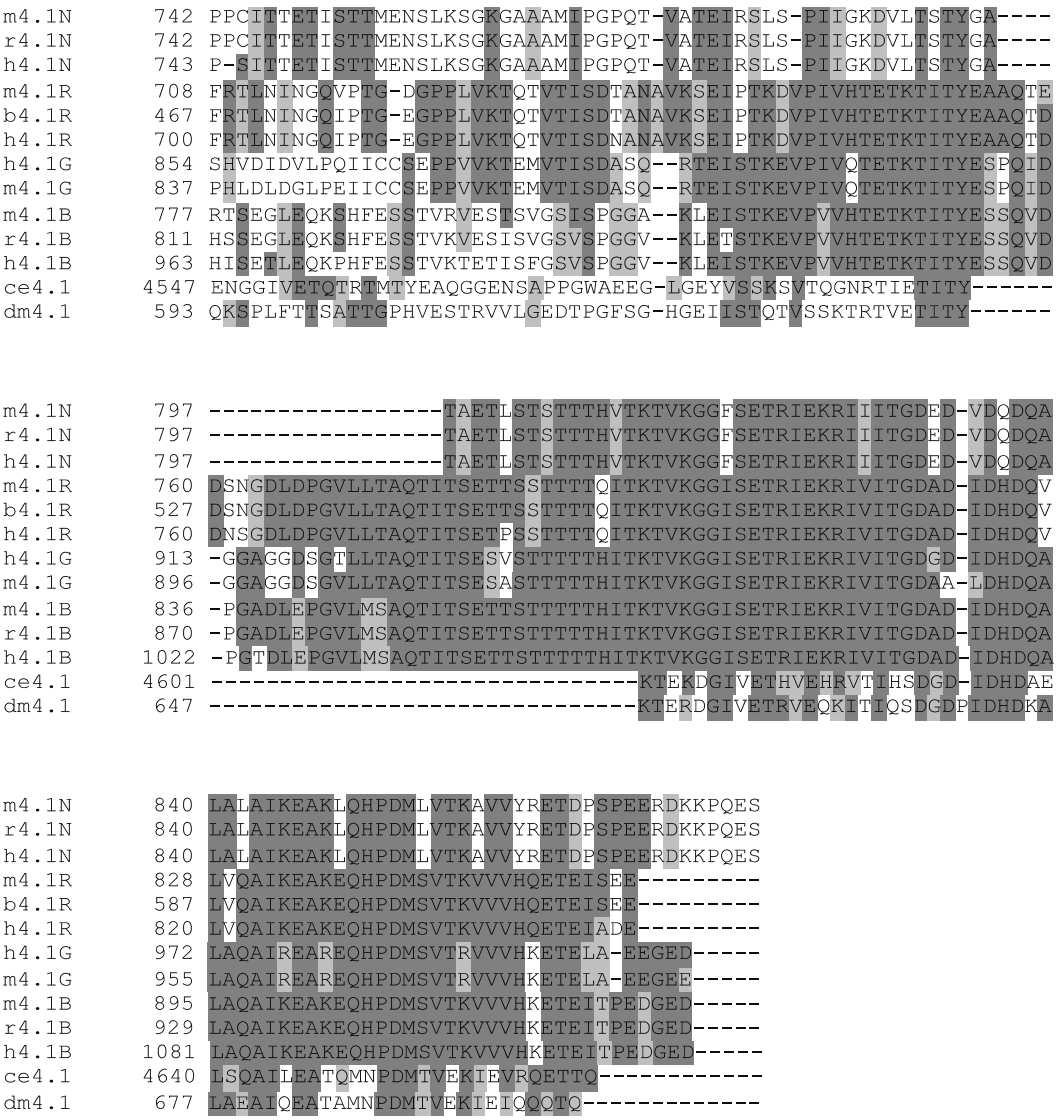

Figure S18. **Multiple sequence alignment of 4.1 CTDs**. This alignment is adapted from the work of Scott et al. [2], which indicates that the most C-terminal half of the domain is well conserved, but the N-terminal half is much more weakly conserved. Species of origin are indicated as h (human), m (mouse), r (rat), b (cow), dm (D. melanogaster), ce (C. elegans).

Reference:

1. T. Dertinger, V. Pacheco, I. von der Hocht, R. Hartmann, I. Gregor and J. Enderlein, *Chemphyschem*, 2007, **8**, 433-443.

2. C. Scott, G. W. Phillips and A. J. Baines, *Eur J Biochem*, 2001, **268**, 3709-3717.