Support Information

The Role of Calcium in Regulating the Conformational Dynamics of D-Galactose/D-Glucose-Binding Protein Revealed by Markov State Model Analysis

*Maohua Yang[1, #], Yegen Tang[1, #], Jingwei Weng[1, *], Zhijun Liu[2, *], Wenning Wang[1, *]*

[1]Department of Chemistry, Multiscale Research Institute of Complex Systems and Institute of Biomedical Sciences, Fudan University, Shanghai 200438, China

[2]National Facility for Protein Science in Shanghai, Zhangjiang Lab, Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China

# these authors contribute equally to the work

* Corresponding authors

Jingwei Weng: jwweng@fudan.edu.cn

Zhijun Liu: liuzhijun@sari.ac.cn

Wenning Wang: wnwang@fudan.edu.cn

Method

Calculation of the bending and twisting angle

The angle between the intersecting lines connecting the center of mass of the hinge region and those of the C-terminal domain and N-terminal domain is defined as bending angle, while the twisting angle formed by the centers of mass of the N-terminal domain, the N-terminal part of hinge region, the C-terminal part of hinge region and the C-terminal domain defines the twisting angle.[1]

Calculation of generalized correlation coefficients

Generalized correlations coefficients for the residues in the GGBP were calculated based on mutual information between all C_α atoms in the protein using the generalized correlation analysis approach.[2] The g_correlation module in the GROMACS package[3] was applied for the analysis.

NMR experiments

All NMR experiments were carried out at 25 °C on Bruker Avance III HD 900 MHz spectrometer equipped with cryogenic probe. ^{1}H-^{15}N HSQC spectra of apo-GGBP and GGBP saturated with Ca^{2+} (1:20 molar ratio) were recorded respectively. NMR spectra were processed with the NMRpipe software package[4] and analyzed with PIPP[5] and Sparky.[6]

Isothermal calorimetric titration

Isothermal calorimetric titration (ITC) measurements were performed on an MicroCal PEAQ-ITC calorimeter (Malvern Instruments Ltd, United Kingdom) at 20 °C. All protein samples were dissolved in a buffer containing 20 mM Tris, pH 8.0, 100 mM NaCl. The titrations were carried out by injecting 40-μl aliquots of apo-GGBP or GGBP/Ca^{2+} (0.47 mM) into glucose (0.05 mM) at time intervals of 2 min to ensure that the titration peak returned to the baseline. ITC data was analyzed using the Malvern MicroCal PEAQ ITC software (Malvern Instruments Ltd) and fitted by the the Wiseman isotherm binding model.[7-9]

Markov state model construction

To validate our model, we also used the tICA[10, 11] method combined with K-Centers,[12, 13] Mini Batch K-Medoids,[14] and Mini Batch K-means algorithms[15] to divide the conformational space. The backbone dihedral angles (Φ, Ψ) of 305 residues are used as the feature for tICA. The cosine and sine values of all the dihedral angles were also calculated and used together with the dihedral angles as features for every structure (totally 305 × 4 = 1220). The tICA method was adopted to reduce the 1220-dimension and capture the slow motions. Optimal parameters for tICA reduction and MSM construction were selected using a generalized matrix Rayleigh quotient (GMRQ)[10, 16] and parameters including tICA number, tICA lag time, cluster method and the number of clusters (Figure S3). Here, we performed 5 folds of cross validation by randomly selecting 80% number of trajectories as the training set for learning

MSM and used the remaining trajectories as test for scoring. According to the GMRQ scores in Figure S3, K-Centers and Mini Batch K-Medoids algorithms give the optimal number of states 50 for both apo-GGBP and GGBP/Ca^{2+} systems. Therefore, MSMs with 50 states were constructed and the 2D-PMFs were calculated from the MSMs (Figure S4).

Classification of macro-sates and mean first passage time calculation

To calculate the mean first passage time (MFPT), we lumped microstates into two macro-states according to,

$$f(S_i) =$$

$$\begin{cases} closed\ state, & 18 * x_i^{bending} - 2376 - y_i^{twisting} < 0 \\ open\ state, & 18 * x_i^{bending} - 2376 - y_i^{twisting} \geq 0 \end{cases}$$

,

where S_i is the center of every micro-states. $\left(x_i^{bending}, y_i^{twisting}\right)$ is the bending and twisting angle value of the center of the micro-state S_i.

Then, we followed the procedure in Ref[17] to compute the mean first passage time (MFPT) from initial state *i* to final state *f*, i.e. the average time taken to get from state *i* to state *f* for the first time.

$$MFPT_{if} = \sum_j P_{ij}(\tau + MFPT_{jf})$$

where P_{ij} is the transition probability from state i to state j, τ is the lag time of the transition probability matrix $\boldsymbol{T}$, and $MFPT_{jf}$ is the mean first passage time of the state j to final state f. The boundary condition is $MFPT_{ff} = 0$.

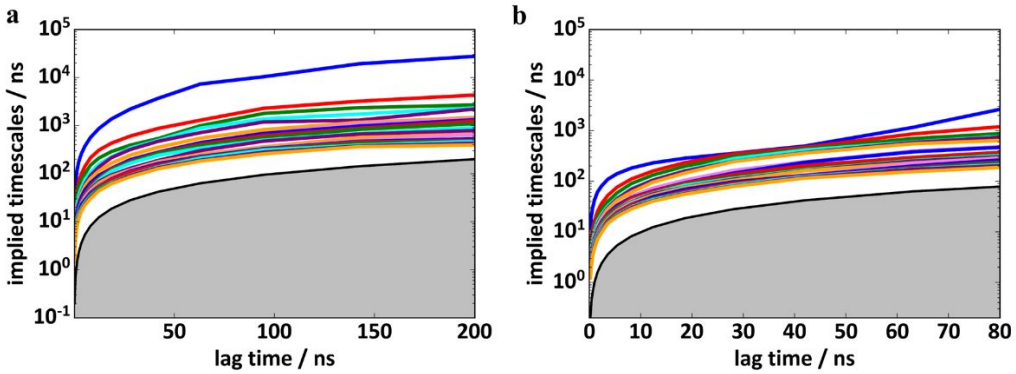

Figure S1. The variations of implied timescales with lag time for apo-GGBP (a) and GGBP/Ca^{2+} (b).

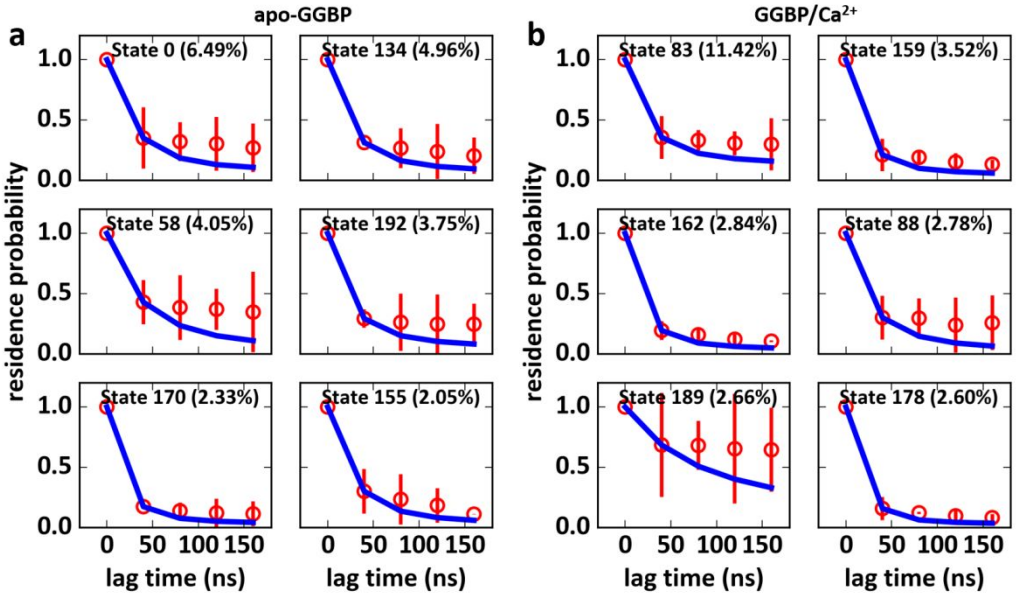

Figure S2. The Chapman-Kolmogorov tests of top six population microstates for apo-GGBP (a) and GGBP/Ca^{2+} (b). The error analyses were performed for the test using the boostrap method.

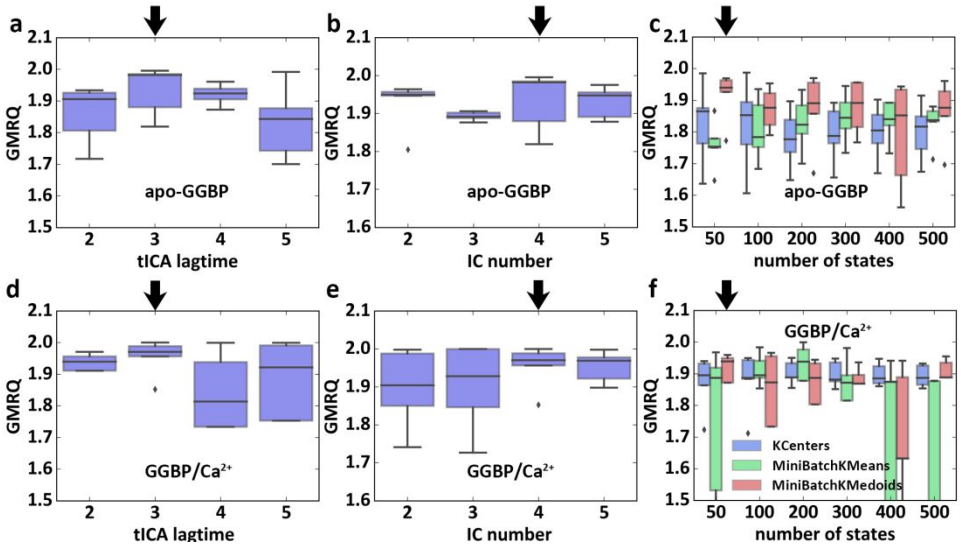

Figure S3. Box plot for GMRQ tests of the parameters for MSM construction, including tICA lag time, IC number and state number. The optimal parameters are denoted with the black arrows (a-c for apo-GGBP and d-f for GGBP/Ca^{2+}).

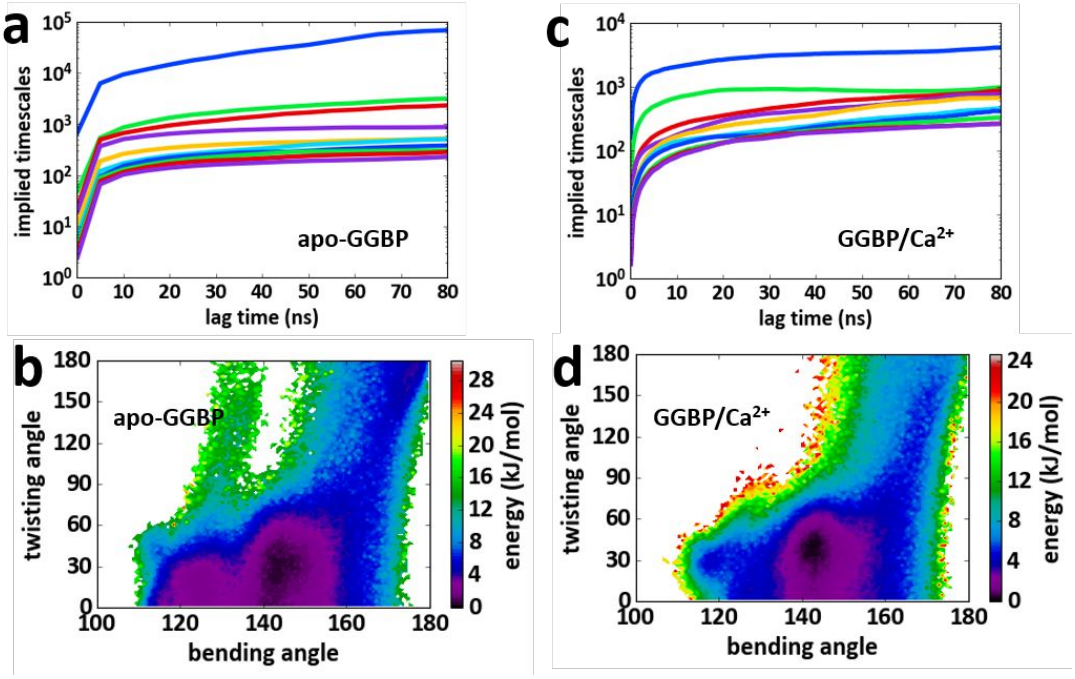

Figure S4. Implied timescale and the calculated free energy landscape of the 50-state MSM constructed by using tICA method for apo-GGBP a-b and GGBP/Ca^{2+} c-d.

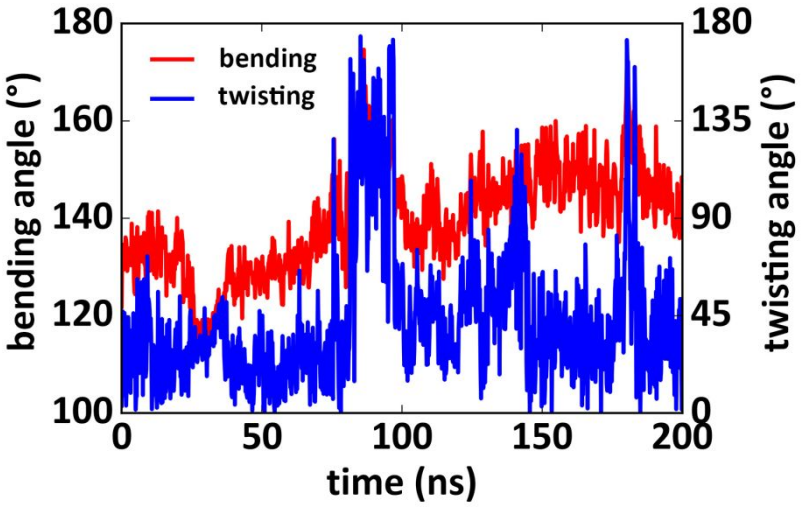

Figure S5. The time variations of the bending (red) and twisting (blue) angles along a selected trajectory. The values of the two angles fluctuate wildly (bending angle: 120˚~180˚, twisting angle: 0˚~180˚).

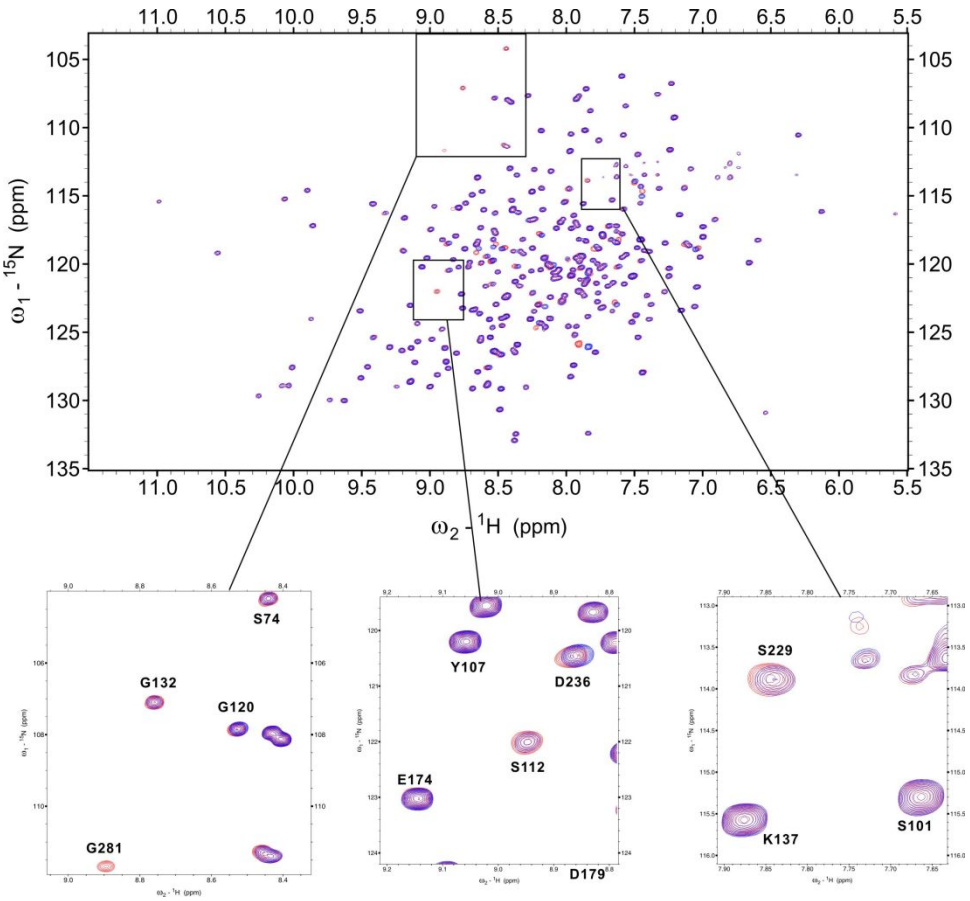

Figure S6. The ^{1}H-^{15}N HSQC spectra of apo-GGBP (blue) and GGBP/Ca^{2+}(red) with the closed-up views of some areas.

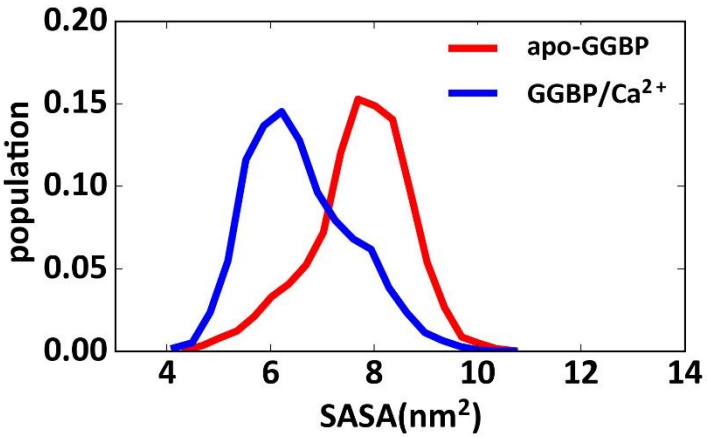

Figure S7. SASA distributions of the Ca^{2+} binding pocket in apo-GGBP (red) and GGBP/Ca^{2+} (blue).

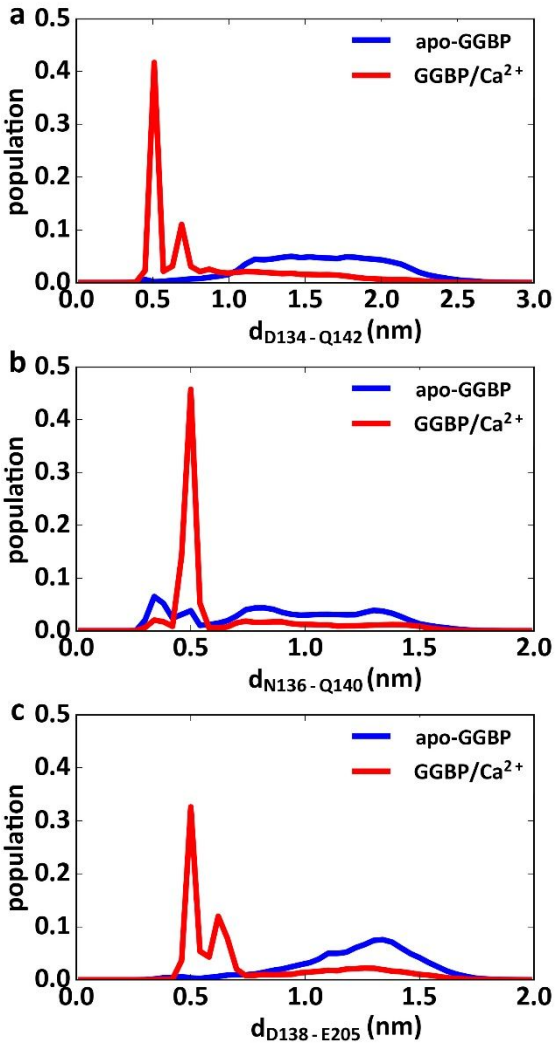

Figure S8. Distance distributions of three atom pairs inside the Ca^{2+}-binding pocket of apo-GGBP (blue) and GGBP/Ca^{2+} (red) systems. The atom pairs are: Asp134-OD1:Gln142-OE1, Asn136-OD1:Gln140-O and Asp138-OD2:Glu205-OE1.

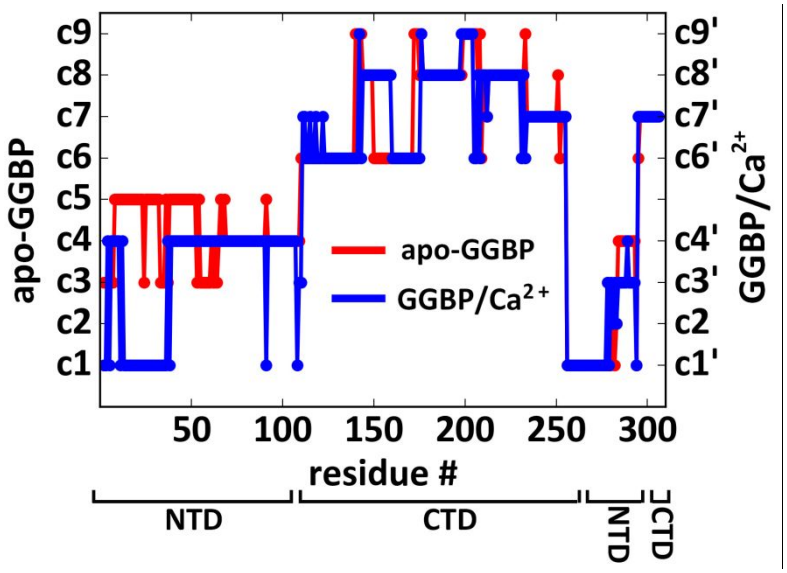

Figure S9. The assignments of communities for all the residues in apo-GGBP (blue) and GGBP/Ca^{2+} (red) systems.

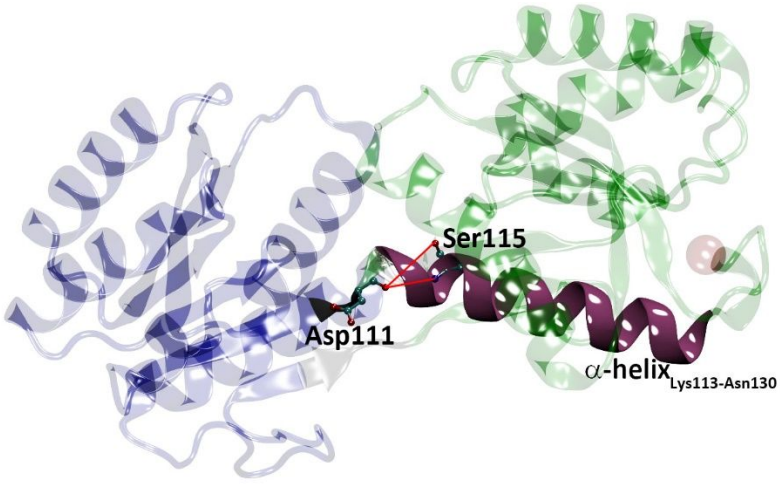

Figure S10. Hydrogen bonds between Ser115 and Asp111. The first α-helix (Lys113-Asn130) of C-terminal domain is colored in mauve.

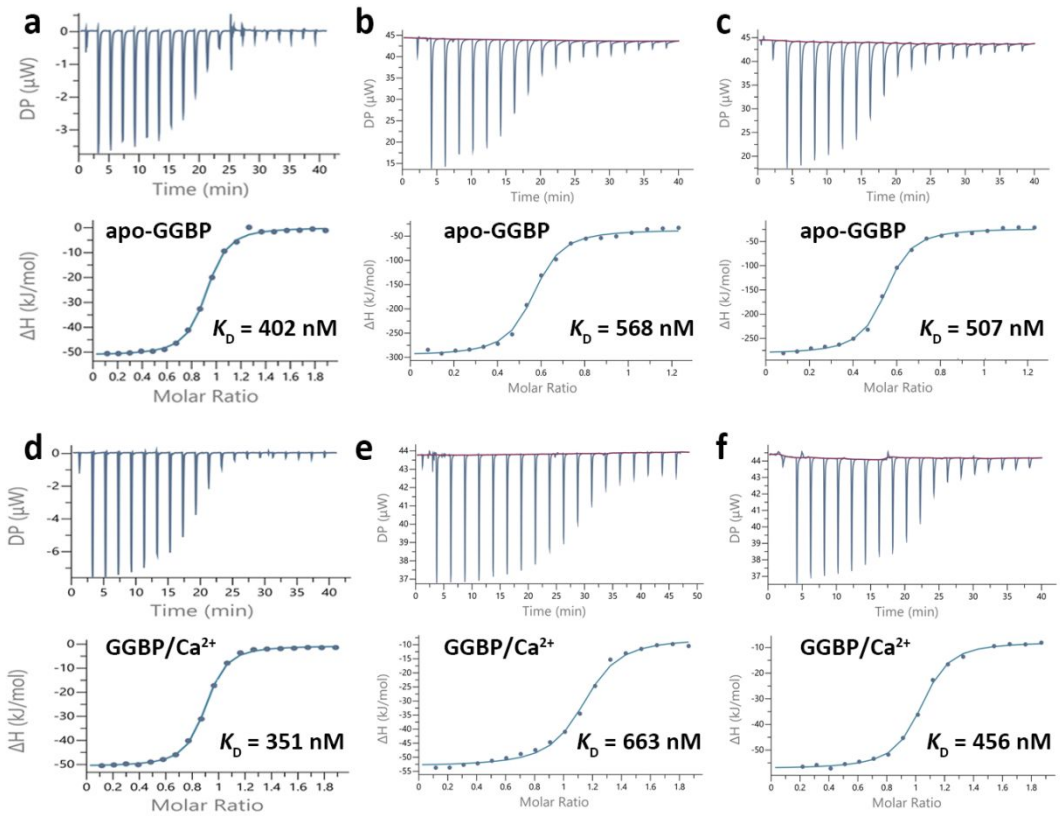

Figure S11. Isothermal calorimetric titration (ITC) experiments that measure the glucose binding affinities of apo-GGBP (a-c) and GGBP/Ca^{2+} (d-f).

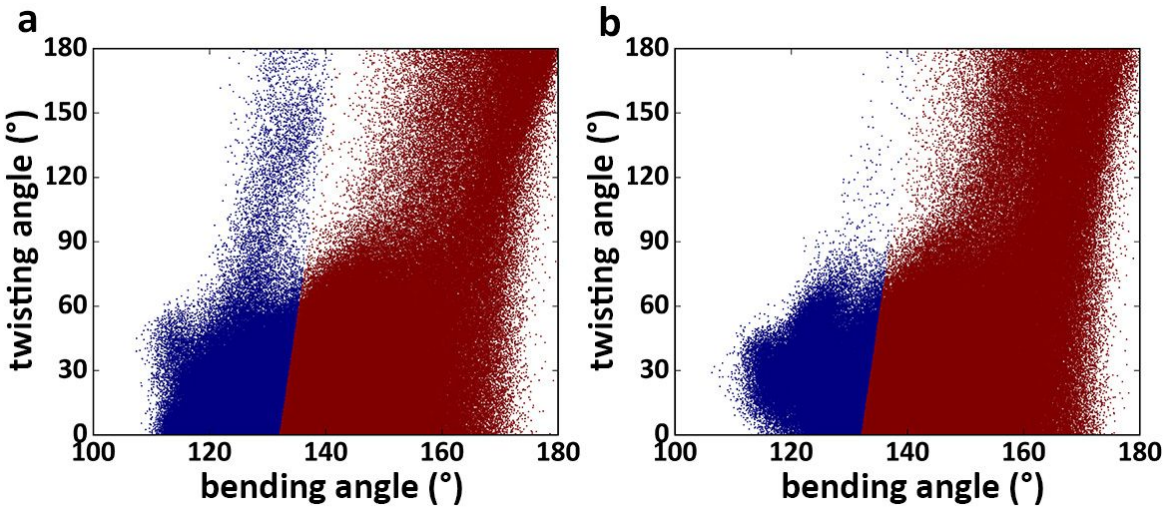

Figure S12. The two-state MSM for apo-GGBP (a) and for GGBP/Ca^{2+} (b).

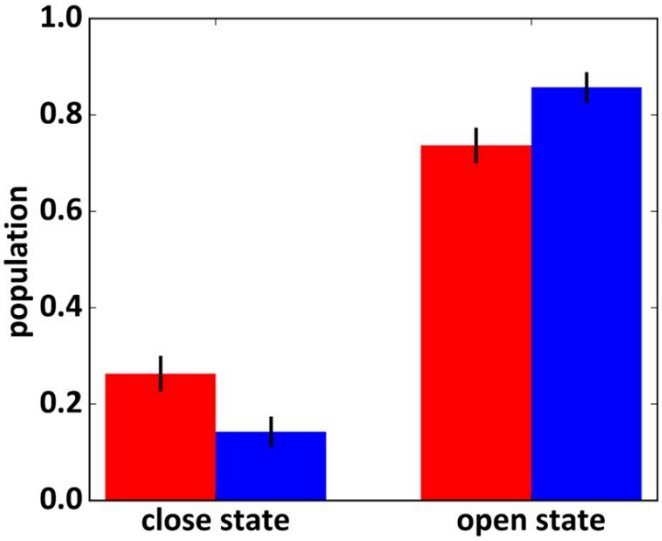

Figure S13. Populations of the closed and open states in the two-state MSM for apo-GGBP (red) and GGBP/Ca^{2+} (blue).

Table S1. Details of the simulation systems

Simulation system	#ions	#water	#atom	Simulation time (μs)
apo-GGBP	0.1M (46 Na^+ 38 Cl^-)	23045	73851	49.34
GGBP/Ca^{2+}	0.1M (44 Na^+ 38 Cl^- 1 Ca^{2+})	23045	73850	48.82

Table S2. The variations of hydrogen bonds upon Ca^{2+}-binding.

Donor	Acceptor	Frequency* (%)		
		apo-GGBP	GGBP/Ca^{2+}	Δ^{**}
Gln142-NE2	Asp138-OD1	1.6771	39.7751	38.098
Gln142-N	Glu205-OE1	12.5377	32.8018	20.2641
Gln142-NE2	Asp138-OD2	1.7526	24.3407	22.588
Lys137-N	Asp134-OD1	10.9122	37.2498	26.3376
Gln142-N	Glu205-OE2	11.3963	33.8059	22.4095
Gln142-NE2	Glu174-OE2	11.3403	31.5245	20.1842
Lys137-N	Asp134-OD2	10.5792	31.6666	21.0874
Gln140-NE2	Asp138-OD1	5.9991	34.2535	28.2544
Gln140-N	Asp138-OD2	1.2704	30.5177	29.2473
Gln142-NE2	Glu174-OE1	9.9975	32.6351	22.6376
Leu135-N	Glu205-OE2	5.0085	27.3315	22.3229
Ser115-OG	Asp111-O	35.2002	14.9931	-20.2071
Ala237-N	Val254-O	57.7135	37.2926	-20.4208

Ser115-N	Asp111-O	41.4467	20.1248	-21.3218

*Here we list the hydrogen bonds which $|\Delta| >= 0$.

**Δ = Frequency(GGBP/Ca^{2+}) – Frequency(apo-GGBP).

Reference

1. Unione, L.; Ortega, G.; Mallagaray, A.; Corzana, F.; Pérez-Castells, J.; Canales, A.; Jiménez-Barbero, J.; Millet, O., Unraveling the conformational landscape of ligand binding to glucose/galactose-binding protein by paramagnetic nmr and md simulations. *ACS chem. Biol.* **2016**, 11, 2149-2157.

2. Lange, O. F.; Grubmüller, H., Generalized correlation for biomolecular dynamics. *Proteins Struct. Funct. Bioinformat.* **2006**, 62, 1053-1061.

3. Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E., Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, 4, 435-447.

4. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., Nmrpipe: A multidimensional spectral processing system based on unix pipes. *J. Biomol. NMR* **1995**, 6, 277-293.

5. Garrett, D. S.; Powers, R.; Gronenborn, A. M.; Clore, G. M., A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson.* **2011**, 213, 357-363.

6. Lee, W.; Tonelli, M.; Markley, J. L., Nmrfam-sparky: Enhanced software for biomolecular nmr spectroscopy. *Bioinformatics* **2015**, 31, 1325-1327.

7. Wiseman, T.; Williston, S.; Brandts, J. F.; Lin, L.-N., Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **1989**, 179, 131-137.

8. Freire, E.; Mayorga, O. L.; Straume, M., Isothermal titration calorimetry. *Anal. Chem.* **1990**, 62, 950A-959A.

9. Brown, A., Analysis of cooperativity by isothermal titration calorimetry. *Int. J. Mol. Sci.* **2009**, 10, 3457-3477.

10. Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F., Identification of slow molecular order parameters for markov model construction. *J. Chem. Phys.* **2013**, 139, 07B604_601.

11. Schwantes, C. R.; Pande, V. S., Improvements in markov state model construction reveal many non-native interactions in the folding of ntl9. *J. Chem. Theory Comput.* **2013**, 9, 2000-2009.

12. Gonzalez, T. F., Clustering to minimize the maximum intercluster distance. *Theor. Comput. Sci.* **1985**, 38, 293-306.

13. Beauchamp, K. A.; Bowman, G. R.; Lane, T. J.; Maibaum, L.; Haque, I. S.; Pande, V. S., Msmbuilder2: Modeling conformational dynamics on the picosecond to millisecond scale. *J. Chem. Theory Comput.* **2011**, 7, 3412-3419.

14. de Hoon, M. J. L.; Imoto, S.; Nolan, J.; Miyano, S., Open source clustering software. *Bioinformatics* **2004**, 20, 1453-1454.

15. Sculley, D., In *Proceedings of the 19th international conference on World wide web*; Association for Computing Machinery: Raleigh, North Carolina, USA, 2010, pp 1177–1178.

16. McGibbon, R. T.; Pande, V. S., Variational cross-validation of slow dynamical modes in molecular kinetics. *J. Chem. Phys.* **2015**, 142, 03B621_621.

17. Singhal, N.; Snow, C. D.; Pande, V. S., Using path sampling to build better markovian state models: Predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. *J. Chem. Phys.* **2004**, 121, 415-425.