Multifunnel Energy Landscapes for Phosphorylated Translation Repressor 4E-BP2 and Its Mutants

Wei Kang [1,2,3], Fan Jiang[1], Yun-Dong Wu[*1,2], and David J. Wales[*3]

1. Laboratory of Computational Chemistry and Drug Design, Laboratory of Chemical Genomics, Peking University Shenzhen Graduate School, Shenzhen, 518055, China.
2. College of Chemistry and Molecular Engineering, Peking University, Beijing, 100871, China.
3. Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW, U.K.

[*]To whom correspondence and requests for materials should be addressed.

Supporting Information

Atom type modifications

For phosphorylated threonine (residue name TPO), in the original parameter files obtained from the work of Case et al.[1], atom types were assigned by following the convention of the AMBER ff99SB force field. To make these parameters compatible with the AMBER 14SB force field, we changed the following two atom types, following the convention of AMBER ff14SB:

Atom name	Old atom type	New atom type	Note
CA	CT	CX	protein C-alpha
CB	CT	3C	sp3 aliphatic C with three heavy atoms

General simulation settings

All molecular dynamics (MD) simulations were performed with PMEMD.CUDA[2] in *Amber*[3] (version 2016) (GPU cores: GeForce GTX 1080). Long-range electrostatic interactions were treated using the particle-mesh Ewald (PME) method[4-5]. A cutoff of 9 Å was used for non-

bonded interactions. A Langevin thermostat with collision frequency γ= 1 ps^{-1} and a Berendsen barostat with pressure relaxation time τ_P= 1 ps were used to maintain constant temperature and constant pressure, respectively. All bonds involving hydrogen were constrained using SHAKE[6]. Hydrogen mass repartitioning (HMR)[7] was performed and a time step of 4 fs was used. The coordinates were recorded every 2 ps.

REMD simulations of model peptides

All model peptides were capped with Ace (N-terminal) and Nme (C-terminal). Each dipeptide was solvated in a truncated octahedral box and counterions were added for charge neutralization. After energy minimization and 4 ns NPT equilibrium at 300 K, REMD simulations were conducted (detailed settings are listed in Table S1)

Calculation of PMFs

PMFs were calculated following the procedure reported by Okur et al.[8] Firstly, distances along a system-specific coordinate (TAApT: Hγ_1 of Thr and P of pThr; RAApT: Cξ of Arg and P of pThr; KAApT: Nξ of Lys and P of pThr; RAAE: Cξ of Arg and Cδ of Glu; KAAE: Nξ of Lys and Cδ of Glu; SAAE: Hγ of Ser and Cδ of Glu) were calculated from the trajectory at 300 K using cpptraj in *Amber*. Histogram analysis was then performed and bin populations were converted to energies based on the assumption of Boltzmann-weighted populations

$$\Delta G = -RTln\frac{N_i}{N_0}$$

where N_0 was the population of the most populated bin and N_i was the population of the *i*th bin. Error bars were calculated from results using only the first or second half of the trajectory.

Trajectory analyses

Native contacts were calculated using *AmberTools*. All pairs of heavy atoms belonging to residues 18-23 or 34-55 (i.e. excluding terminal residues and the long loop) and separated by at least 3 residues were used in the native contact calculations. The first NMR structure was used as a reference and the default distance cutoff 7.0 Å was used. To highlight major changes in the fraction of native contacts (Q), simple moving averages with a sliding window size of 1 ns were

calculated (SI Figure 1). Snapshots extracted from MD trajectories were also shown along the Q plots.

Hydrogen bond analysis was conducted with *cpptraj* with 3.0 Å as the donor-acceptor distance cutoff and 135° as the angle cutoff (similar results were obtained with a different angle cutoff 150°). Secondary structure analysis was conducted with the DSSP algorithm[9] implemented in *cpptraj*.

Proline cis-trans isomerization

Full-length 4E-BP2 has eleven Pro residues, and four of them are included in the truncated system in this study. Bah et al.[10] pointed out that Pro *cis-trans* isomerization may occur, especially for Pro31 located in the long loop. In this study, we started database construction with five starting states, where all X-Pro peptide bonds (X stands for any residue preceding Pro) adopt the *trans* conformation, and no *cis* isomer was observed in the final database. However, *trans*-to-*cis* transformation was observed in MD simulations at both 450 K (in run C at 816 ns) and 480 K (in run A at 696 ns, and in run B at 447 ns) after the system fully unfolded.

Table S1. REMD simulation settings for model peptides.

Peptide	No. of TIP3P water	Temperature (K)	No. of Replicas	Time (ns)
TAApT	2215	296.2 - 576.1	46	50
RAApT	2282	296.2 - 576.1	46	50
KAApT	2317	296.3 - 567.8	46	50
RAAE	2216	296.2 - 580.4	46	50
KAAE	2295	296.2 - 576.1	46	50

Table S2. Number of minima (MIN) and transition states (TS) in the DPS databases.

Protein	No. of MIN	No. of TS
pWT	41 931	50 319
p(D33K)	46 239	53 776
p(Y54A/L59A)	53 248	64 360

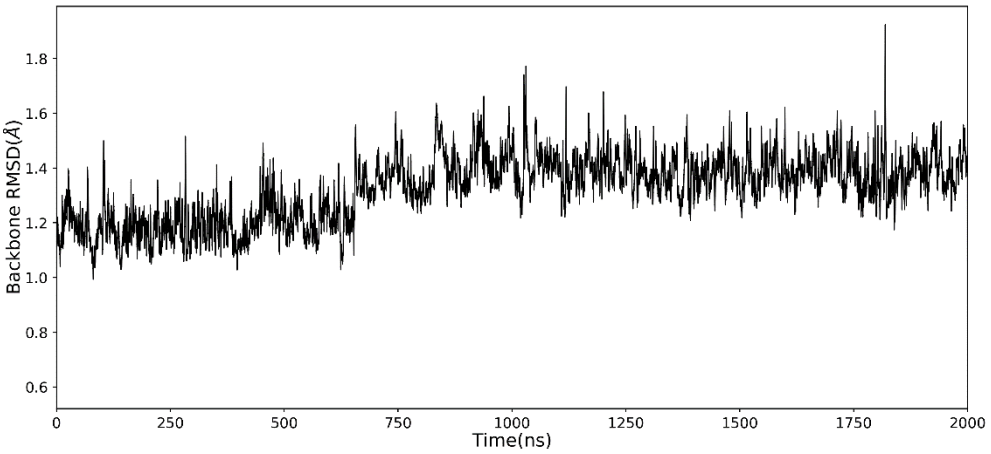
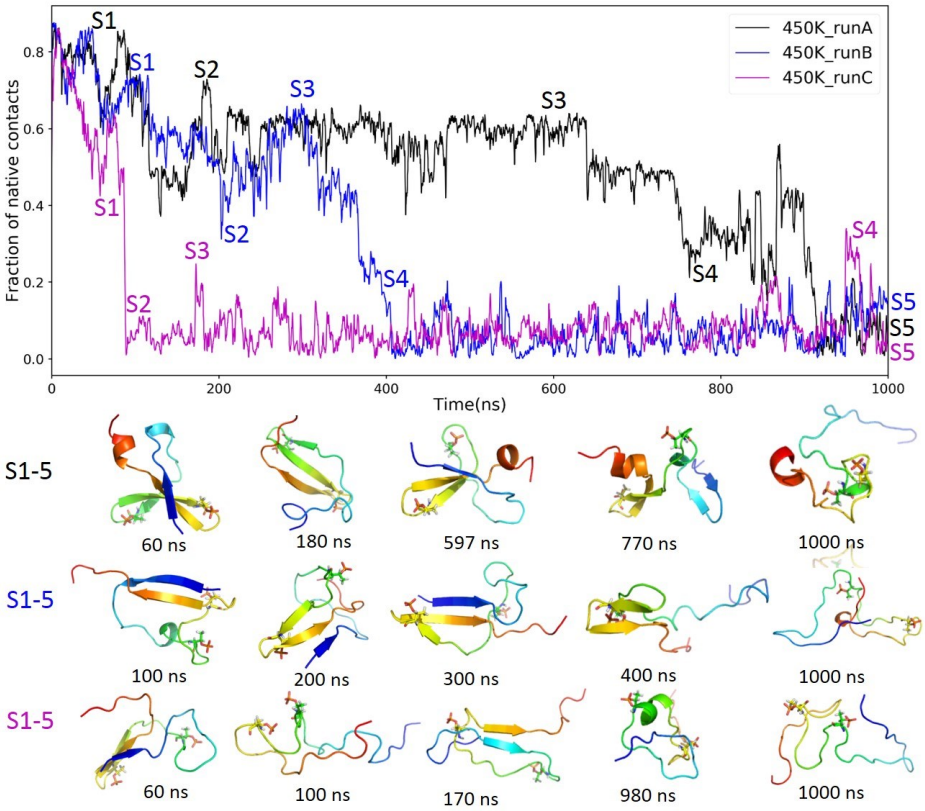


Figure S1. Backbone RMSD to NMR structure of pT37pT46 at 300 K with RSFF2C/TIP3P. Residues 19-56 (indexes in original full sequence) were used in RMSD calculation.



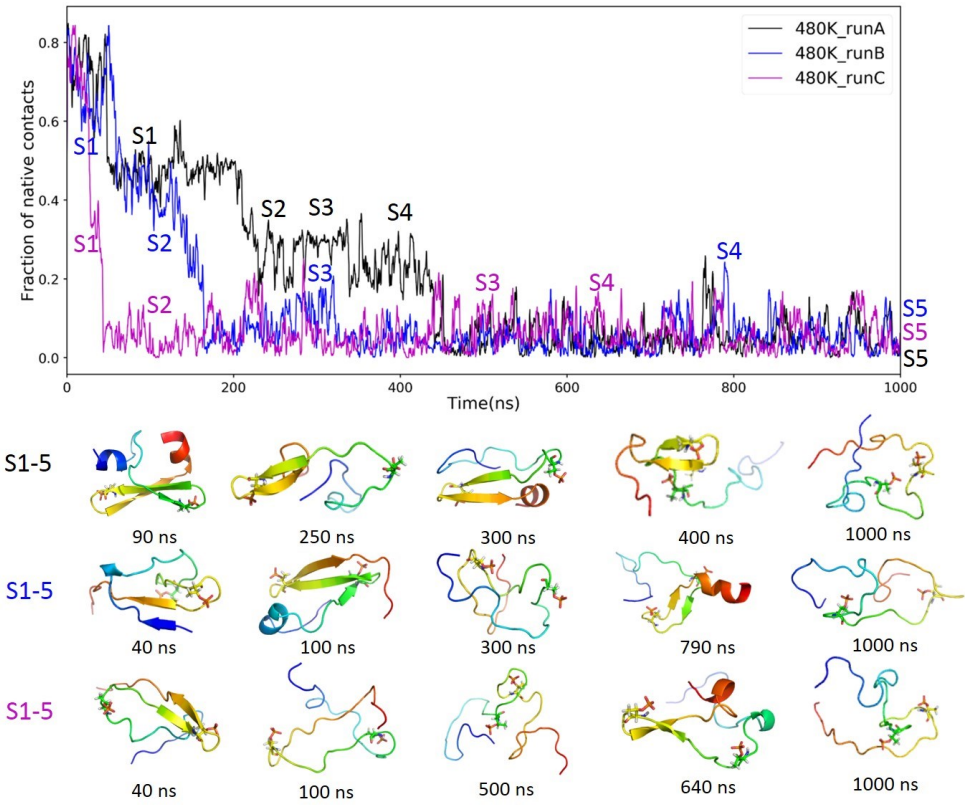


Figure S2. Fraction of native contacts during explicit solvent simulations at 450 K (top) and 480 K (bottom). Some intermediates are shown.

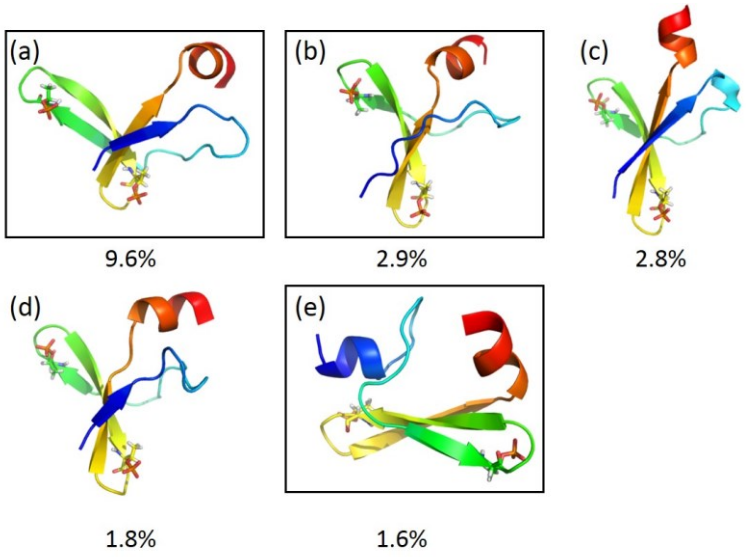


Figure S3. Representative structures for the five most populated states obtained from a clustering analysis for high temperature MD snapshots containing a C-terminal α helix. Since structures (a), (c), and (d) share the same secondary structure feature, namely four β strands and a C-terminal α helix, we chose the three structures in boxes as starting points for basin-hopping runs

addressing structural diversity.

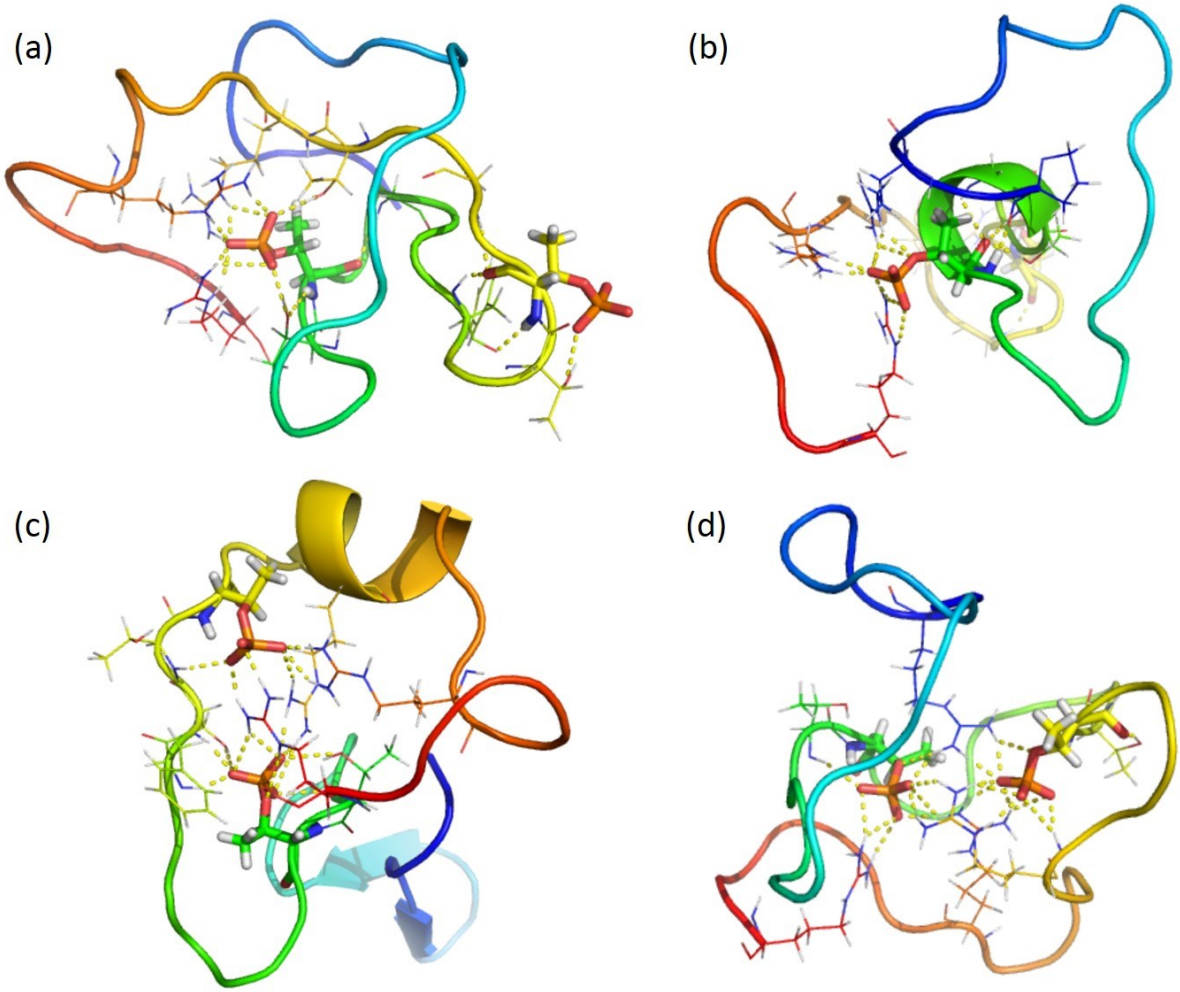


Figure S4. Structures with unphysical hydrogen bond networks involving pThr extracted from high temperature MD simulation trajectories. pThr are shown as sticks and other residues involved in hydrogen-bonding are shown as lines.

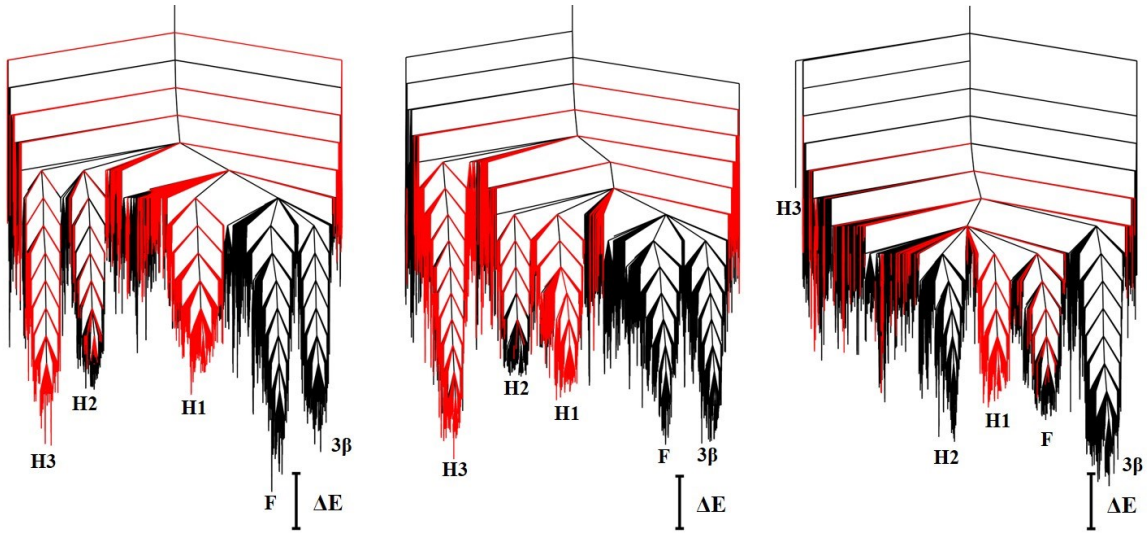


Figure S5. Potential energy disconnectivity graphs (spacing for superbasin analysis ΔE = 10 kcal/mol) for pWT (left), p(D33K) (middle) and p(Y54A/L59A) (right). Minima with the binding motif forming an α helix are highlighted in red.

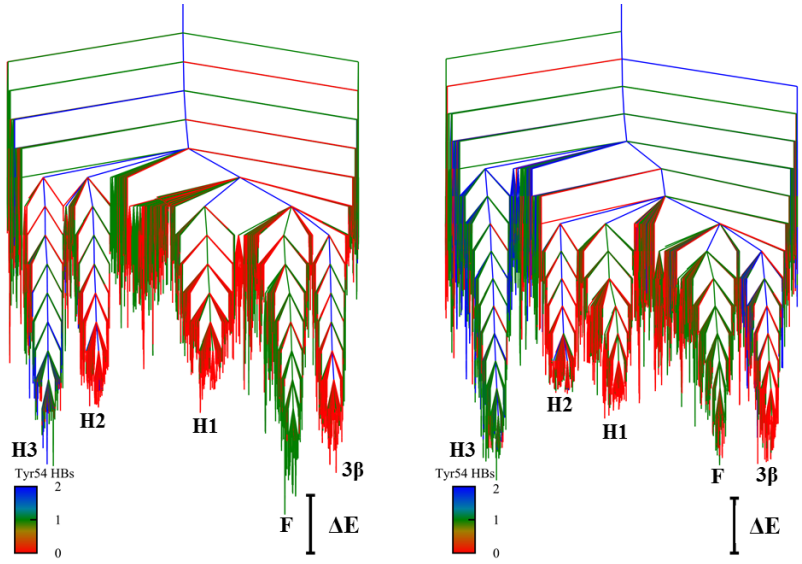


Figure S6. Potential energy disconnectivity graphs (ΔE = 10 kcal/mol) for pWT (left) and p(D33K) (right). Minima are coloured according to the number of hydrogen bonds between the Tyr54 sidechain and other residues.

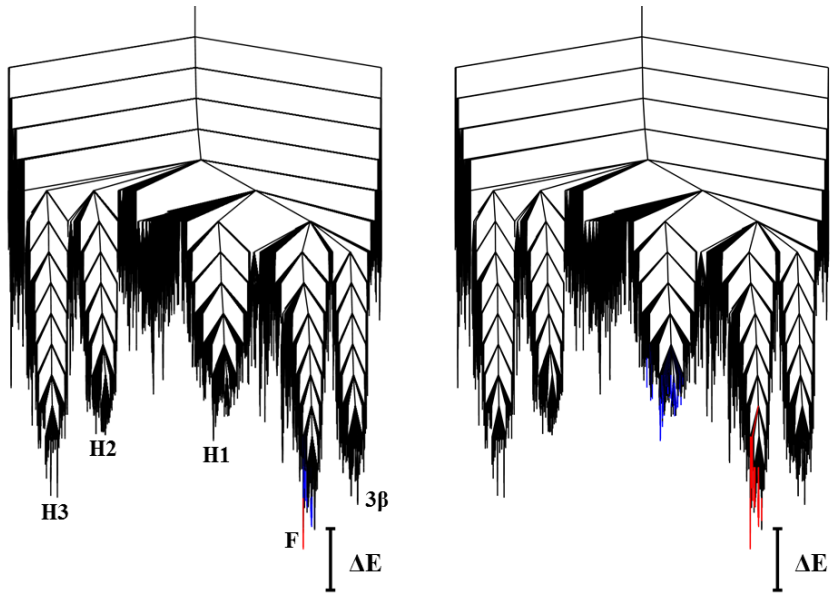


Figure S7. Potential energy disconnectivity graphs (ΔE = 10 kcal/mol) for pWT. Minima are coloured according to their contributions to the heat capacity that account for 99% of the positive (blue) and negative (red) occupation probability gradients[11] at 199 K (left) and 315 K (right). This construction identifies the minima involved in the corresponding transition.

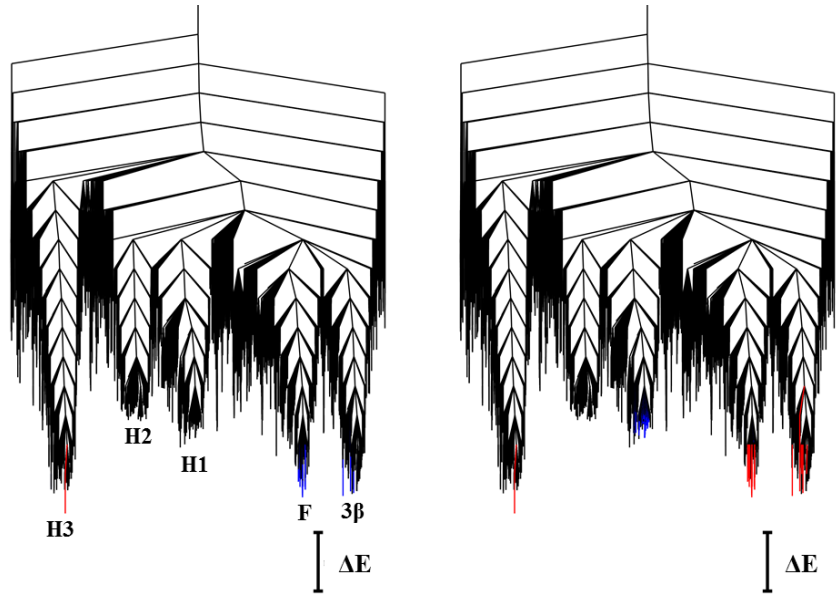


Figure S8. Potential energy disconnectivity graph (ΔE = 10 kcal/mol) for p(D33K). Minima are coloured according to their contributions to the heat capacity that account for 99% of the positive (blue) and negative (red) gradients[11] at 135 K (left) and 216 K (right).

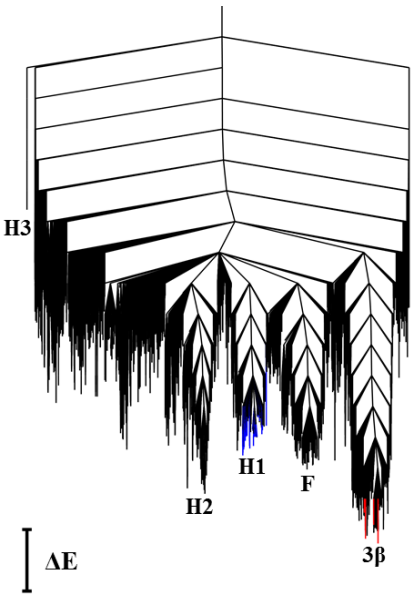


Figure S9. Potential energy disconnectivity graph (ΔE = 10 kcal/mol) for p(Y54A/L59A). Minima are coloured according to their contributions to the heat capacity that account for 99% of the positive (blue) and negative (red) gradients[11] at 241 K.

Reference

(1) Steinbrecher, T.; Latzer, J.; Case, D. A., Revised AMBER Parameters for Bioorganic Phosphates. *J. Chem. Theory Comput.* **2012,** *8*, 4405-4412.

(2) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013,** *9*, 3878-3888.

(3) Case, D. A.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *AMBER 2017*, University of California: San Francisco, 2017.

(4) Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98*, 10089-10092.

(5) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103*, 8577-8593.

(6) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977,** *23*, 327-341.

(7) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E., Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015,** *11*, 1864-1874.

(8) Okur, A.; Wickstrom, L.; Simmerling, C., Evaluation of Salt Bridge Structure and Energetics in Peptides Using Explicit, Implicit, and Hybrid Solvation Models. *J. Chem. Theory Comput.* **2008,** *4*, 488-498.

(9) Kabsch, W.; Sander, C., Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983,** *22*, 2577-2637.

(10) Bah, A.; Vernon, R. M.; Siddiqui, Z.; Krzeminski, M.; Muhandiram, R.; Zhao, C.; Sonenberg, N.; Kay, L. E.; Forman-Kay, J. D., Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **2015,** *519*, 106-109.

(11) Wales, D. J., Decoding heat capacity features from the energy landscape. *Phys. Rev. E* **2017,** *95*, 030105.