

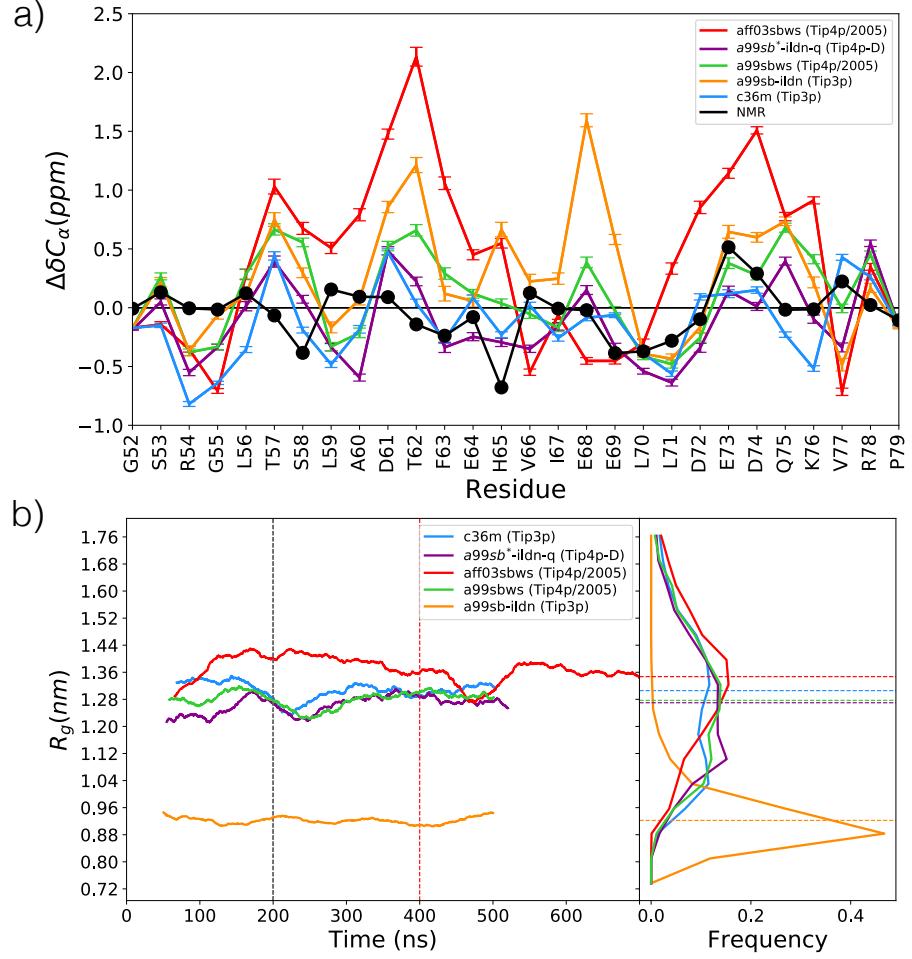
Sequence specificity despite intrinsic disorder: how a disease-associated Val/Met polymorphism rearranges tertiary interactions in a long disordered protein

Ruchi Lohia¹, Reza Salari¹, Grace Brannigan^{1,2*},

1 Center for Computational and Integrative Biology, Rutgers University, Camden, NJ, USA

2 Department of Physics, Rutgers University, Camden, NJ, USA

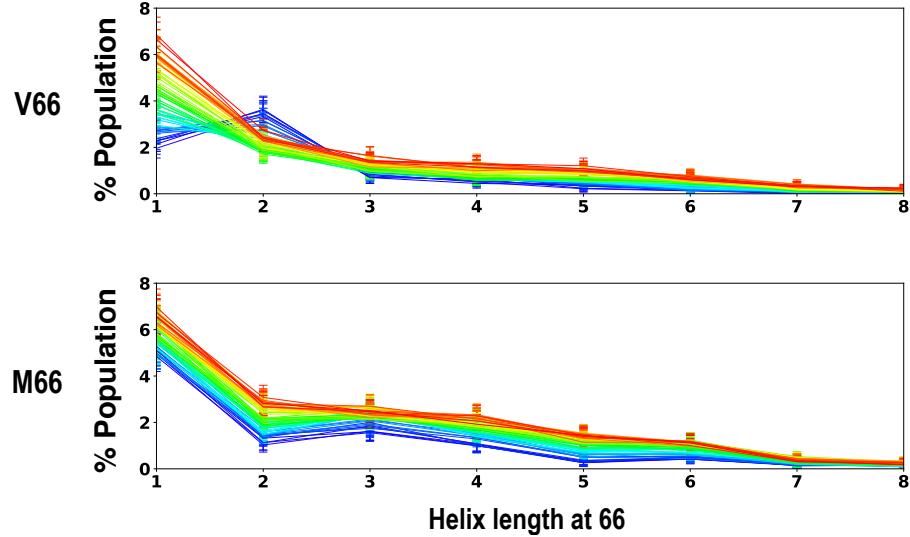
* grace.brannigan@rutgers.edu(GB)



S1 Fig. Force field comparison. We ran T-REMD simulations of a 30 residue fragment of the V66 prodomain with several commonly used force field and water model combinations. (a) Comparison of $\Delta\delta C_\alpha$ secondary chemical shifts at 280K from MD ensembles for a99sb*-ildn-q [1, 2] with Tip4p-D [3], c36m [4], a99sbws [1, 5], a03sbws [5, 6], a99sb-ildn with Tip3p [7], calculated using SPARTA+ [8] and NMR from Ref. 9. (b) R_g vs the simulation time, using a 100 ns moving window on left and R_g distribution for each force field on right. Tip3p and a03sbws generates most collapsed and expanded R_g distribution respectively. The equilibration time and $\langle R_g \rangle$ is shown with vertical and horizontal dashed lines for each force field.

Table S1. Summary of force field comparison simulations.

Force field	$\Delta\delta C_\alpha$	$\langle R_g \rangle$	equilibration time	no of replica
aff03sbws (Tip4p/2005)	0.855	1.347 ± 0.007	400 ns	36
a99sb*-ildn-q (Tip4p-D)	0.355	1.270 ± 0.007	200 ns	36
a99sbws (Tip4p/2005)	0.425	1.277 ± 0.007	200 ns	36
c36m (Tip3p)	0.350	1.306 ± 0.007	200 ns	30
a99sb-ildn (Tip3p)	0.617	0.922 ± 0.003	200 ns	32



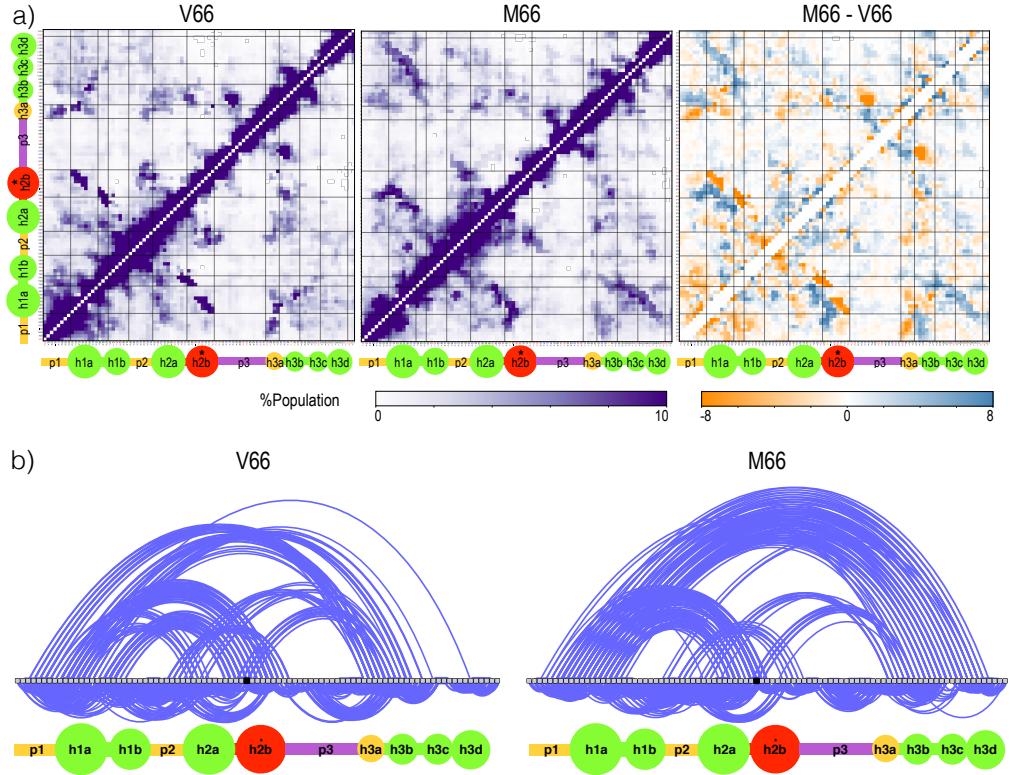
S2 Fig. Effects of temperature and Val66Met mutation on helix propensity around residue 66. Frequency of helix of a given length at residue 66 in V66 (top) and M66 (bottom) in the temperature range of 300K to 385 K. With the increase in temperature the color transitions from cooler (blue) to hotter (red). It is entropically unfavorable for V66 and its neighboring residue to be simultaneously in the helical region of the Ramachandran map, as indicated by the decreasing helical propensity with increasing temperature. For longer helices, the trend will depend more on the additional side-chains in the helix, and the trend with temperature is reversed, but it remains weaker than the analogous trend for the M66 sequence. Errors represent standard error of a Bernoulli trial with n number of samples, where n is the product of total number unique replicas forming the helix of given length at residue 66 at a given temperature and average number of roundtrips per replica, 17.

S0.1 Heterogeneous behavior of individual domains

Disordered proteins can be well-described by Flory scaling theory $\langle R_{|i-j|} \rangle = A|i-j|^\nu$, where $\langle R_{|i-j|} \rangle$ is the ensemble-averaged internal distance, $|i-j|$ is residue separation along the chain, and ν is the Flory scaling coefficient [10]. Larger values of ν correspond to swollen coils, while smaller values correspond to compact globules [11]. In particular, when $\nu=0.6$ (“good solvent”) the protein maximizes its interaction with solvent, and for $\nu=0.33$ (“poor solvent”), the protein maximizes self-interactions. The special intermediate case of $\nu=0.5$ is called a “theta solvent” [10]. Most IDPs that obey this scaling behavior have $\nu>0.5$ [11–14].

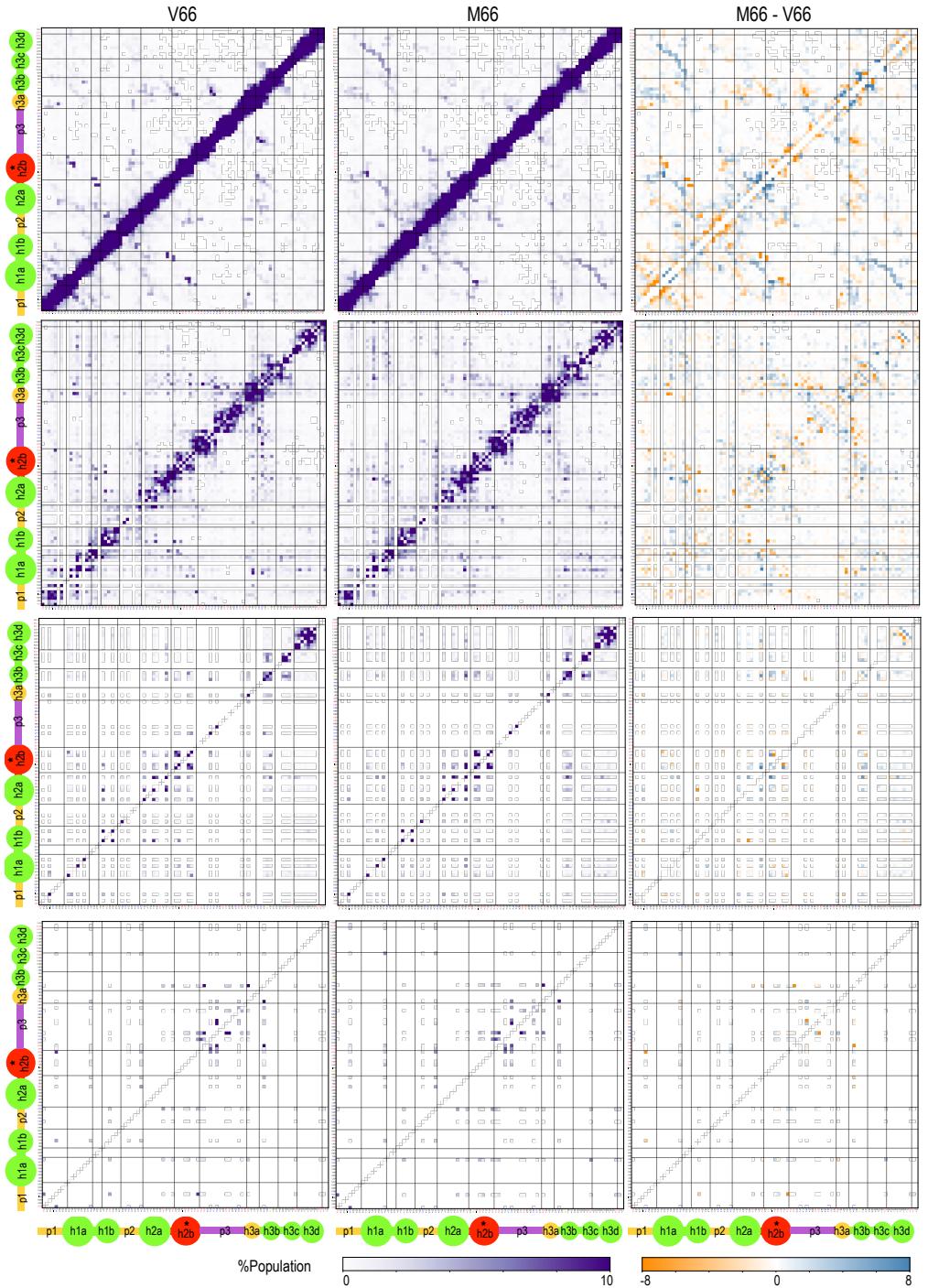
As shown in Fig S3 Fig the prodomain as a whole is not well fit by a single power law: for separations of 15 or fewer residues the prodomain falls in the “theta solvent” regime, while for separations of 20 or more residues it falls in the “poor solvent” regime. Each identified individual domain does obey a power law, and we calculated A and ν for each domain as if it was isolated from rest of the protein (Fig S3 Fig). The highest observed value of ν was in h2b and h3c domain. This is in agreement with strong polyelectrolyte nature of h2b and high content of Proline residue (20%) in h3c.

Method We calculated the average distance between the first atom (N) and last atom (O) for all residue pairs of a given sequence as a function of sequence separation $|i-j|$ using *g_traj*. Errors before fitting were calculated as the standard error in the mean, where $n = 1088$ is the product of total number of replicas simulated (64) and average number of roundtrips per replica (17). ν was calculated by linear fit of

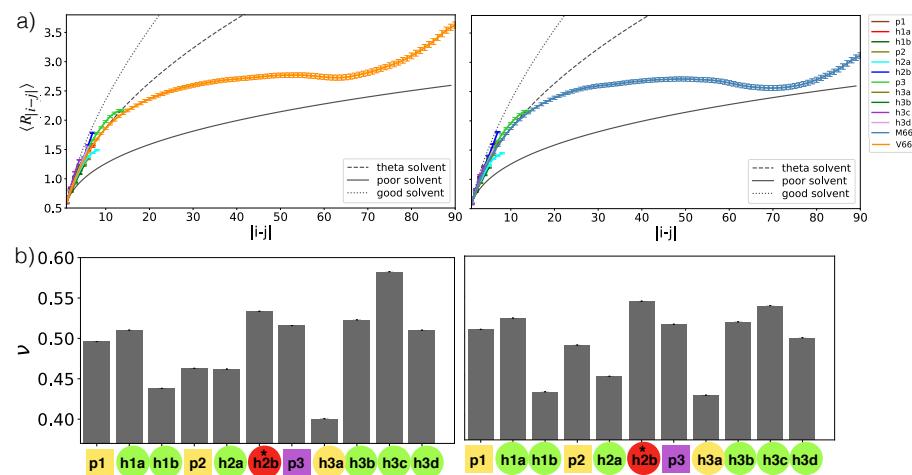


S3 Fig. Scaling behavior of each identified domain. Ensemble averaged interchain distance profiles for the entire V66 and M66 prodomain and each blob in the sequence. Theoretical polymer scaling limits are shown with grey lines (prefactor A = 0.59 nm) (top). Flory exponents for each blob (bottom).

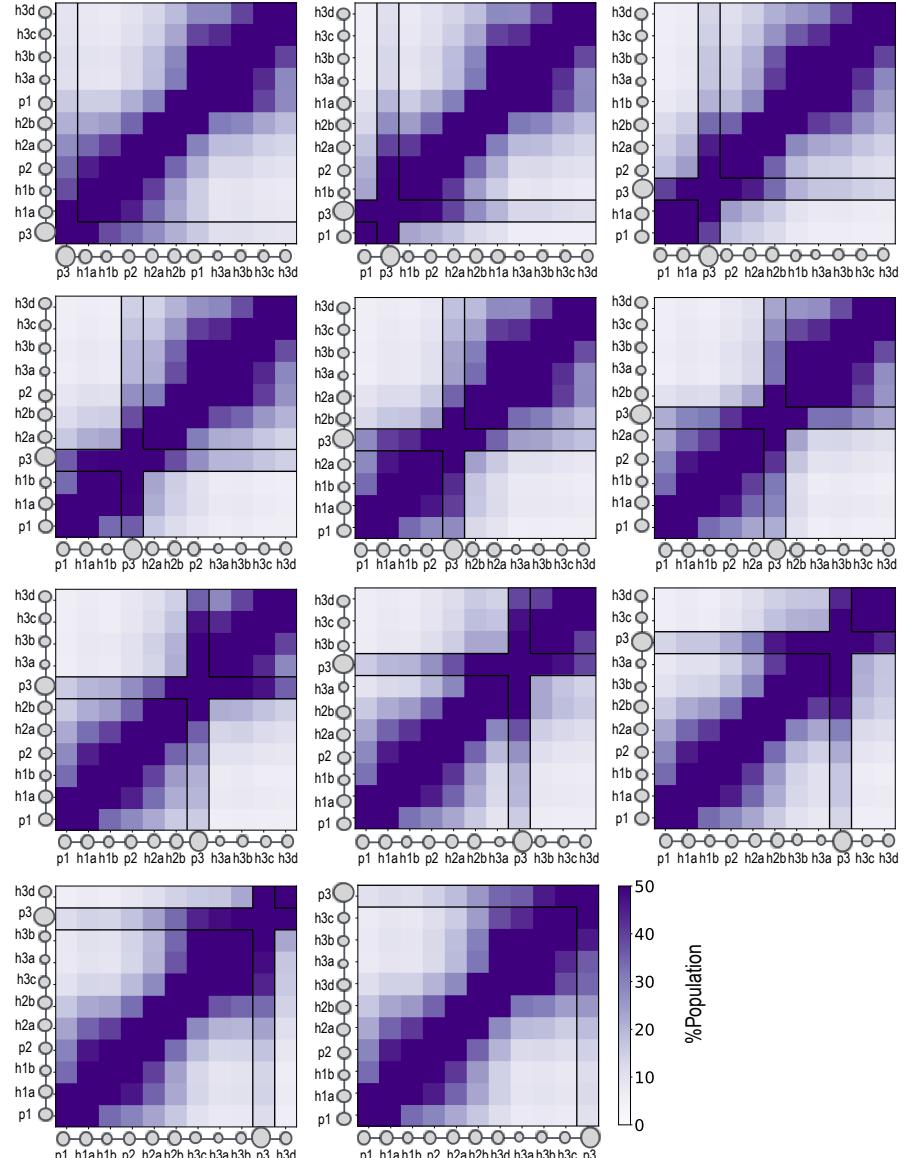
$\ln(\langle R_{|i-j|} \rangle)$ vs $\ln(|i-j|)$ weighted by each point's pre fit error with fixed A of 0.59nm. To exclude the short-range backbone rigidity, distances with $|i-j| < 3$ were not fit.



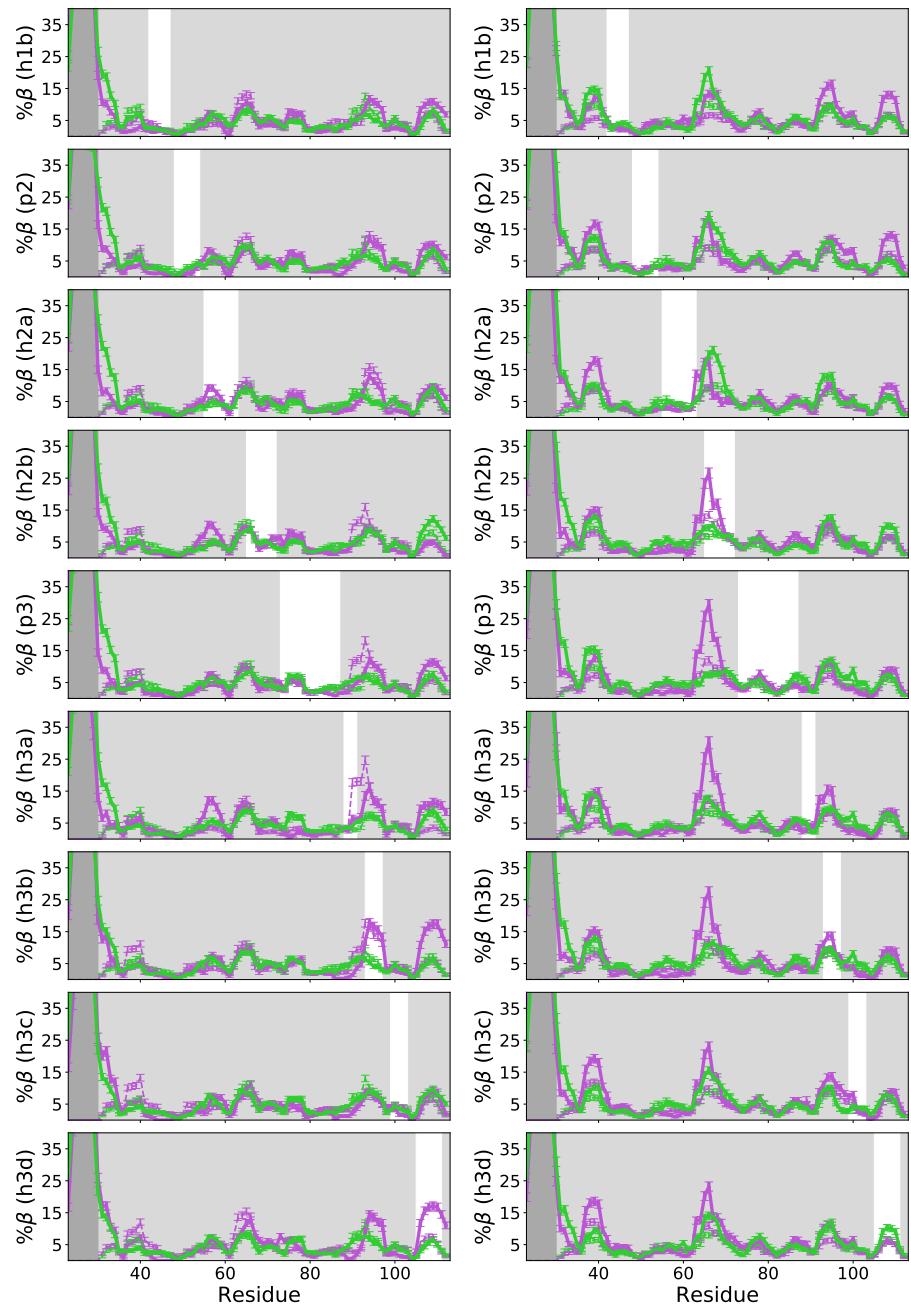
S4 Fig. Effect of perturbing monomer properties on freely-jointed, self-avoiding heteropolymer Contact probability maps from MC simulations, analogous to those in Figure 5a of the main text, in which the blob p3 is swapped with every other blob in the chain, with the new location represented by the purple square in the graph annotation. As the p3 blob is shifted along the chain, p3 and p1 consistently bound a white “forbidden” region that has little interaction with the rest of the protein.



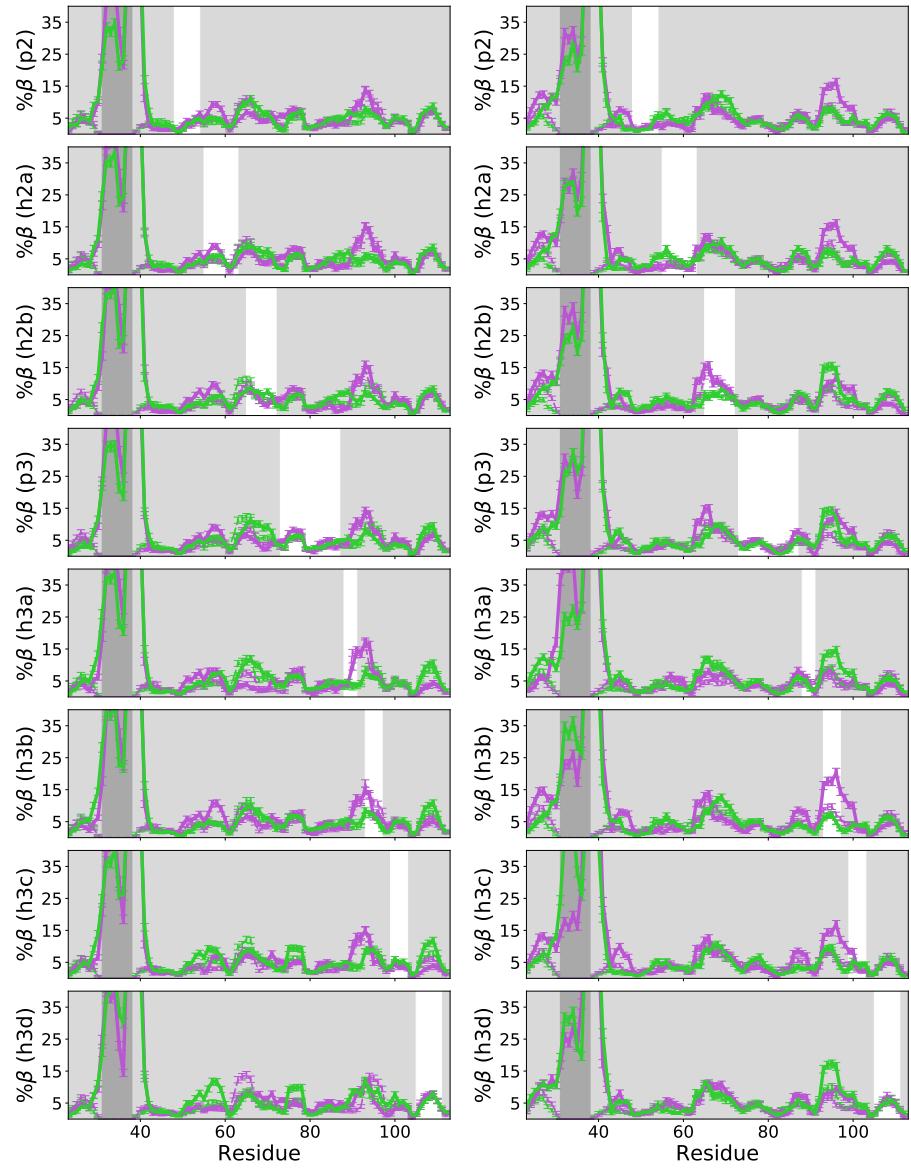
S5 Fig. β -pairing of each blob pair β propensities at each residue in V66 sequence (top) and M66 sequence (bottom) for four clusters. Frames were first clustered by whether the X-Y contact was formed (purple) or broken (green), and then by whether β structure was present in X (solid) or absent (dashed). X represents p1 and is annotated at the top panel and Y represents other blobs identified in the sequence and is annotated on the left for each panel. Errors represent standard error of a Bernoulli trial with n number of samples, where n is the product of total number of unique replicas in a given cluster and average number of roundtrips per replica (17).



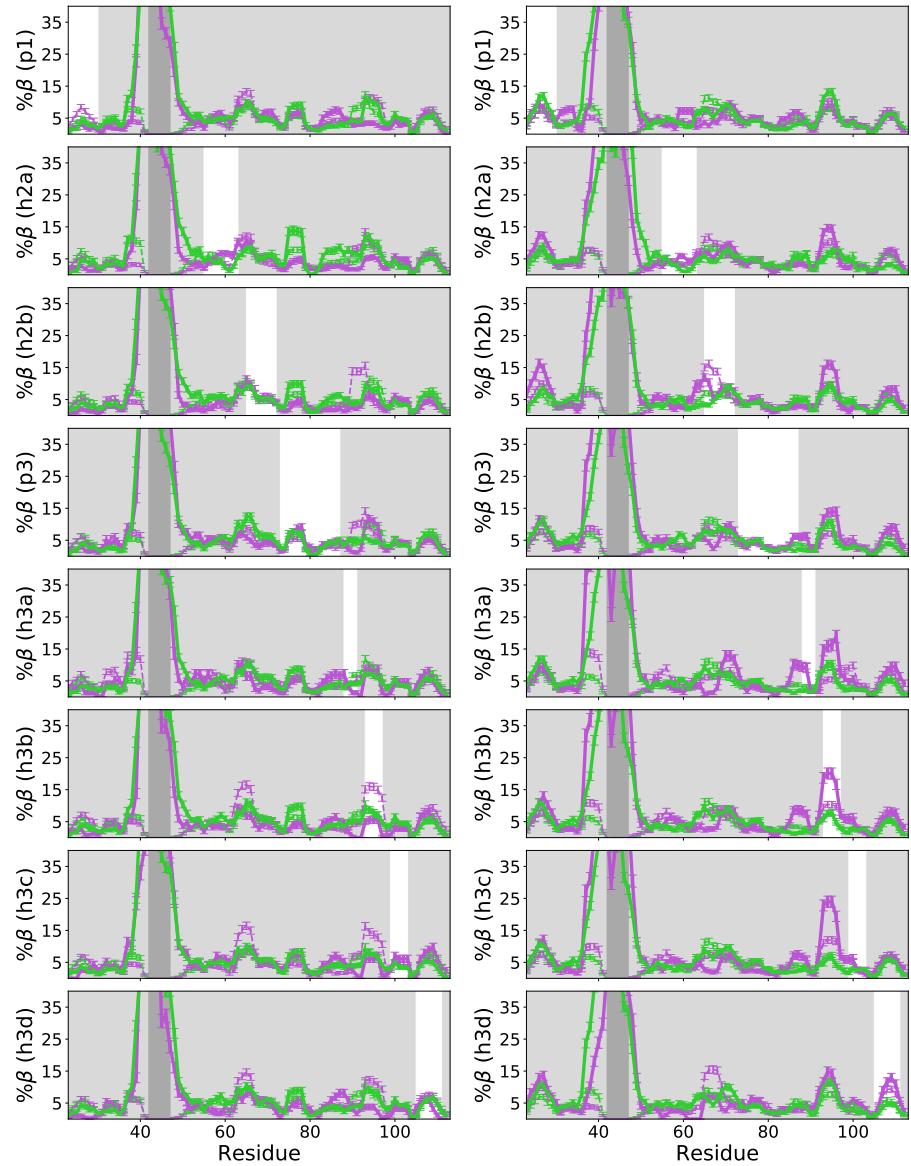
S6 Fig. β -pairing of each blob pair. Same as Fig S5, where X represents h1a (left) or h1b(right).



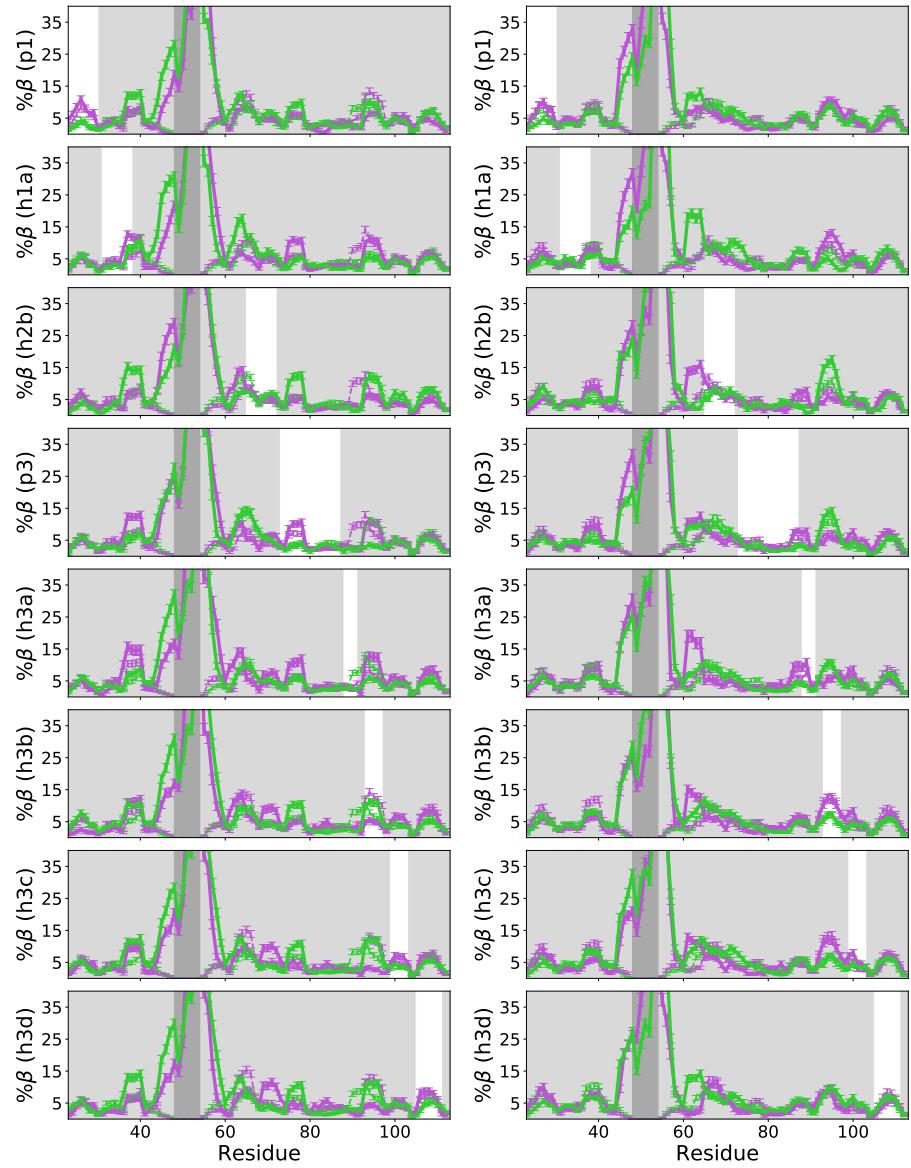
S7 Fig. β -pairing of each blob pair. Same as Fig S5, where X represents h2a (left) or h2b(right).



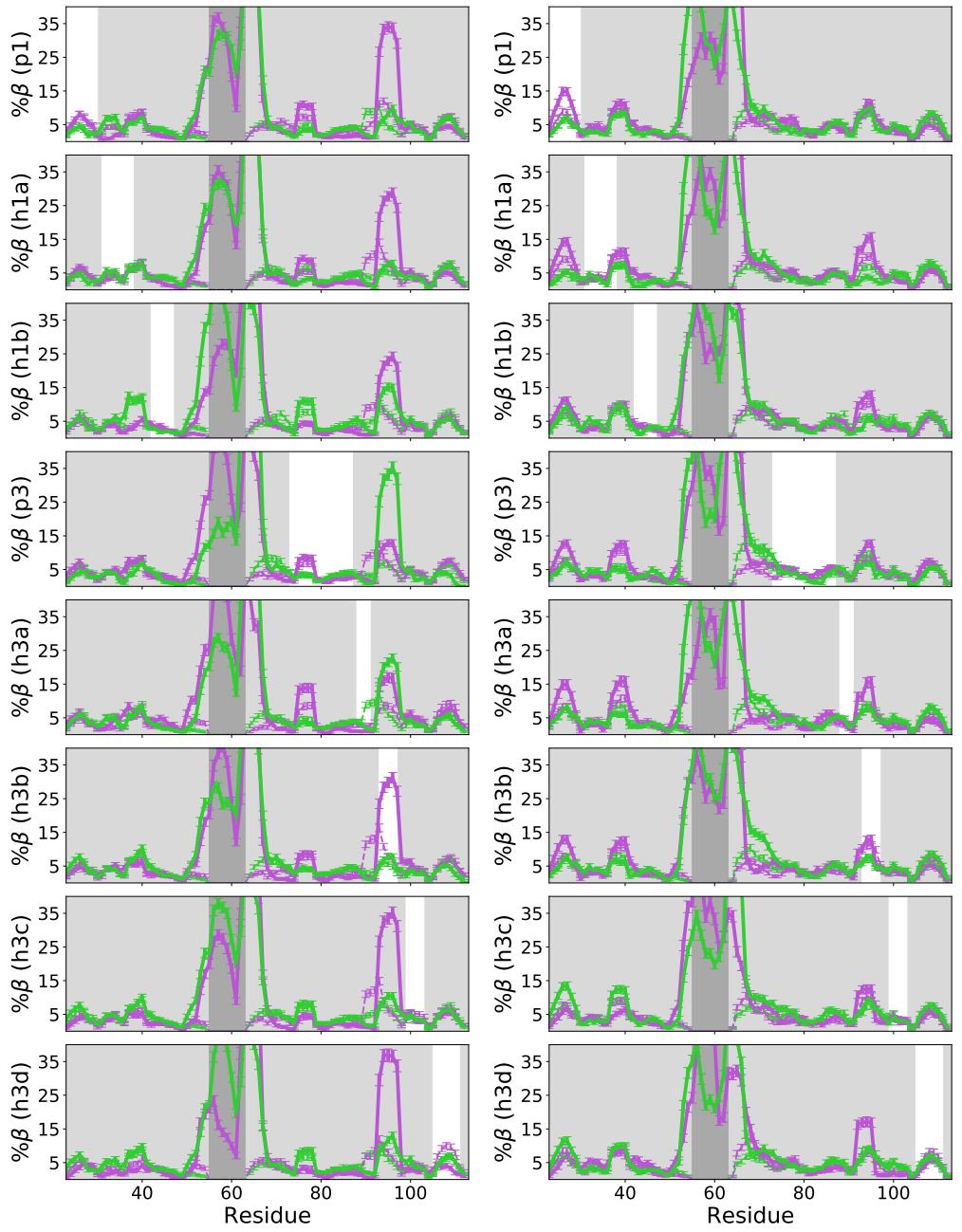
S8 Fig. β -pairing of each blob pair. Same as Fig S5, where X represents h3a (left) or h3b(right).



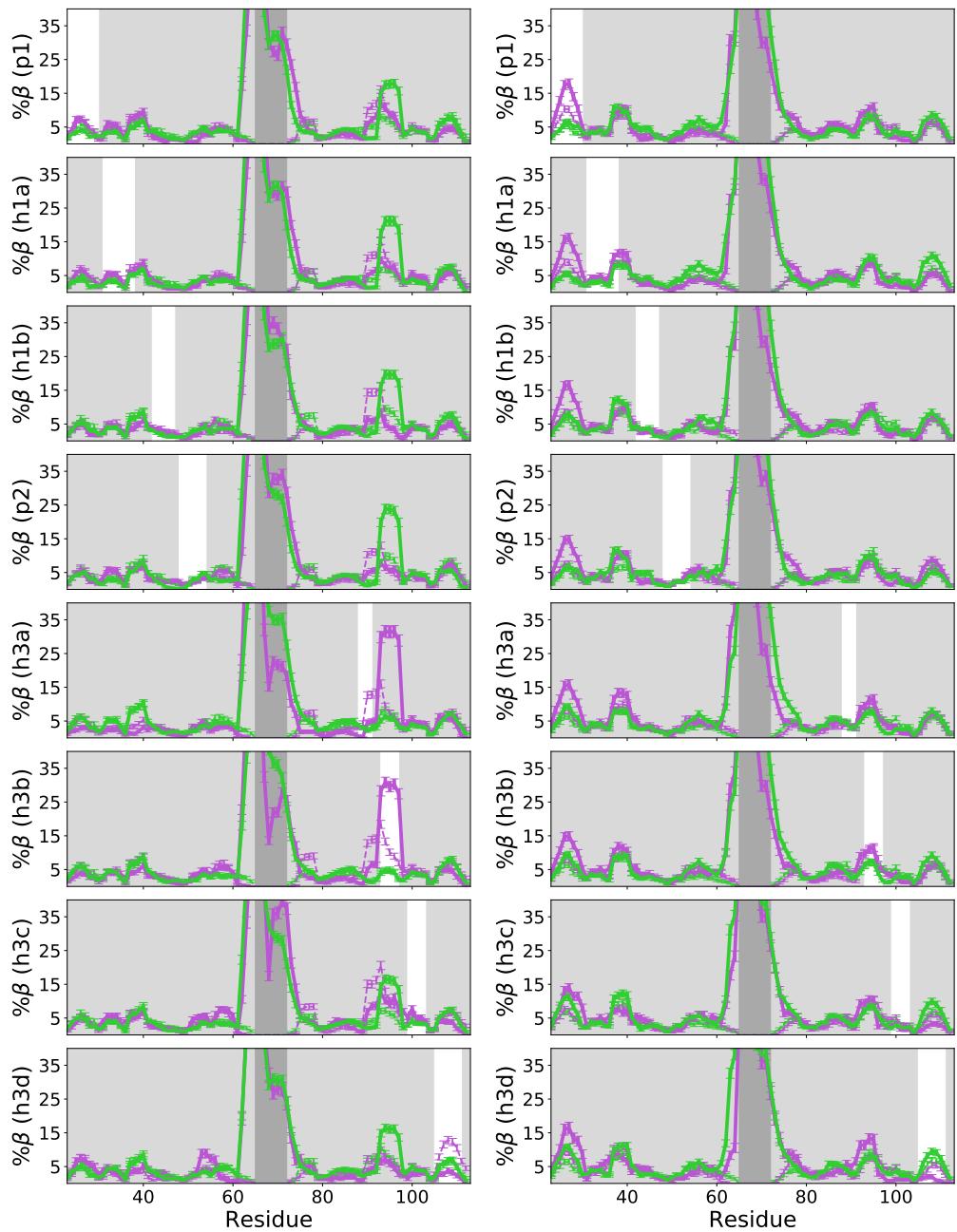
S9 Fig. β -pairing of each blob pair. Same as Fig S5, where X represents h3c (left) or h3d(right).



S10 Fig. β -pairing of each blob pair. Same as Fig S5, where X represents p2 (top) or p3 (bottom).



S11 Fig. Residue level contacts for the entire prodomain. Contact probability between every residue pair for V66 (left) and M66 (middle) and M66-V66(right). Two residue pairs are in contact if the distance between C_{α} - C_{α} atoms between the two residues are 0.8nm or less. b) A linear network of transient tertiary contacts shown in a). The contact networks were build using Cytoscape [15] with a linear representation of residues. Each protein residue comprises a node in the network, with interactions between residues represented as edges. The strength of individual interactions can be interpreted by the thickness of the edge line on the network diagram. If the separation between residues forming the contact is more than 3, its edge is drawn above the node; otherwise, the edge is drawn at the bottom of the node. To focus on significant interactions, interactions showing more than 4% persistence were considered in network visualization.



S12 Fig. Residue level contacts for the entire prodomain. Contact probability between every residue pair for V66 (left) and M66 (middle) and M66-V66(right). Two residue pairs are in contact if the distance between backbone-backbone atoms between the two residues are 0.4nm or less (1st row), if the distance between non hydrogen sidechain-siechain atoms between the two residues are 0.4nm or less (2nd row), if the distance between non hydrogen sidechain-siechain atoms between the two hydrophobic residues are 0.4nm or less (3rd row), if the two residue pairs are forming a salt bridge with the distance between the donor and acceptor atoms < 0.32nm (4th row).

References

1. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins.* 2010;78(8):1950–8. doi:10.1002/prot.22711.
2. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct Funct Bioinforma.* 2006;65(3):712–725. doi:10.1002/prot.21123.
3. Piana S, Donchev AG, Robustelli P, Shaw DE. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J Phys Chem B.* 2015;119(16):5113–5123. doi:10.1021/jp508971m.
4. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods.* 2017;14(1):71–73. doi:10.1038/nmeth.4067.
5. Best RB, Zheng W, Mittal J. Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J Chem Theory Comput.* 2014;10(11):5113–5124. doi:10.1021/ct500569b.
6. Best RB, Hummer G. Optimized Molecular Dynamics Force Fields Applied to the Helix–Coil Transition of Polypeptides. *J Phys Chem B.* 2009;113(26):9004–9015. doi:10.1021/jp901540t.
7. Jorgensen WL. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J Am Chem Soc.* 1981;103(2):335–340. doi:10.1021/ja00392a016.
8. Shen Y, Bax A. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J Biomol NMR.* 2010;48(1):13–22. doi:10.1007/s10858-010-9433-9.
9. Anastasia A, Deinhardt K, Chao MV, Will NE, Irmady K, Lee FS, et al. Val66Met polymorphism of BDNF alters prodomain structure to induce neuronal growth cone retraction. *Nat Commun.* 2013;4:2490. doi:10.1038/ncomms3490.
10. Flory PJ. The Configuration of Real Polymer Chains. *J Chem Phys.* 1949;17(3):303–310. doi:10.1063/1.1747243.
11. Das RK, Pappu RV. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc Natl Acad Sci U S A.* 2013;110(33):13392–7. doi:10.1073/pnas.1304749110.
12. Hofmann H, Soranno A, Borgia A, Gast K, Nettels D, Schuler B. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc Natl Acad Sci.* 2012;109(40):16155–16160. doi:10.1073/pnas.1207719109.
13. Zerze GH, Best RB, Mittal J. Sequence- and Temperature-Dependent Properties of Unfolded and Disordered Proteins from Atomistic Simulations. *J Phys Chem B.* 2015;119(46):14622–14630. doi:10.1021/acs.jpcb.5b08619.

14. Meng F, Bellaiche MMJ, Kim JY, Zerze GH, Best RB, Chung HS. Highly Disordered Amyloid- β Monomer Probed by Single-Molecule FRET and MD Simulation. *Biophys J.* 2018;114(4):870–884. doi:10.1016/j.bpj.2017.12.025.
15. Ahlstrom LS, Baker JL, Ehrlich K, Campbell ZT, Patel S, Vorontsov II, et al. Network visualization of conformational sampling during molecular dynamics simulation. *J Mol Graph Model.* 2013;46:140–9. doi:10.1016/j.jmgm.2013.10.003.