

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

Ruchi Lohia,<sup>†</sup> Reza Salari,<sup>†</sup> and Grace Brannigan<sup>\*,†,‡</sup>

<sup>†</sup>*Center for Computational and Integrative Biology, Rutgers University, Camden, NJ, USA*

<sup>‡</sup>*Department of Physics, Rutgers University, Camden, NJ, USA*

E-mail: [grace.brannigan@rutgers.edu](mailto:grace.brannigan@rutgers.edu)(GB)

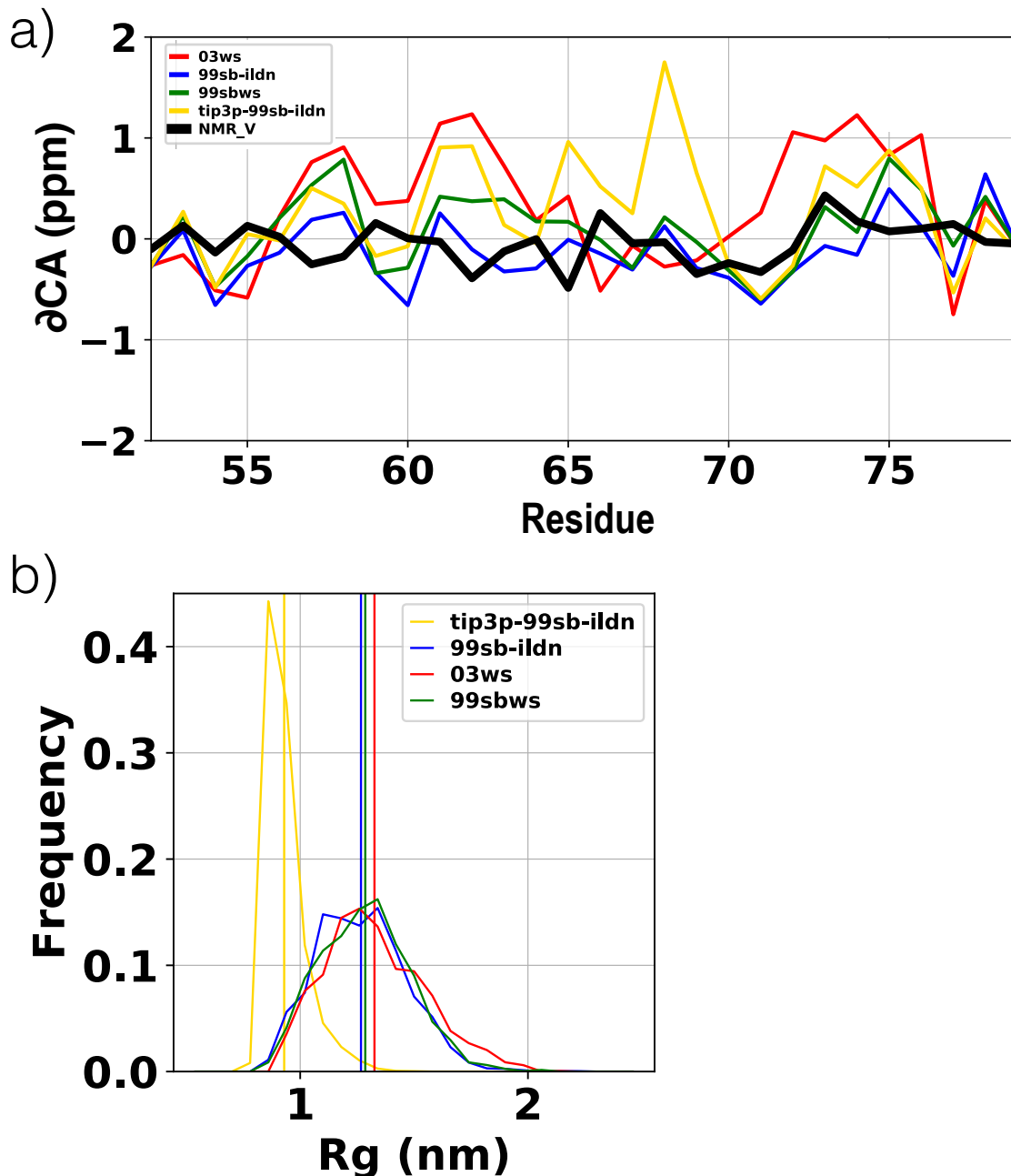


Fig S1: **Force field comparison.** We ran 500ns of T-REMD simulations of a 30 residue fragment of the V66 prodomain with several commonly used force field and water model combinations. The first 200ns of data was discarded as equilibration. (a) Comparison of calculated  $C_\alpha$  chemical shifts from MD ensembles at 280K for Amber99sb\*-ildn-q<sup>1,2</sup> with Tip4p-D<sup>3</sup> (RMSD 0.36 ppm), Amber99sbws<sup>1,4</sup> (RMSD 0.42 ppm), Amberff03sbws<sup>4,5</sup> (RMSD 0.73 ppm), Amber99sb-ildn with Tip3P<sup>6</sup> (RMSD 0.65 ppm), calculated using SPARTA+<sup>7</sup> and NMR  $C_\alpha$  chemical shifts (black line) from Ref. 8 at 280K. (b)  $R_g$  distribution for each force field. Tip3P generates very collapsed ensembles and the remaining three force field generates similar  $R_g$  distribution. The  $\langle R_g \rangle$  is shown with vertical lines.

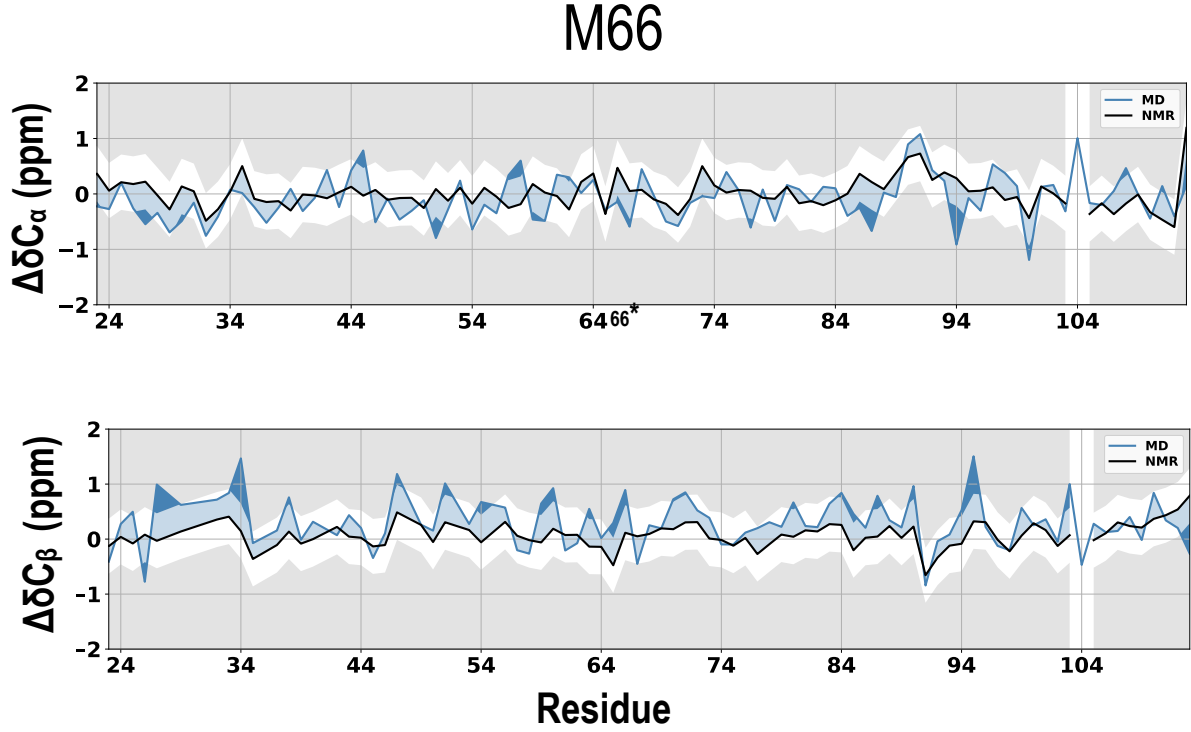


Fig S2: **Comparison of MD and NMR observables.**  $C_\alpha$  (top) and  $C_\beta$  (bottom) chemical shifts from NMR at 280K (black lines)<sup>8</sup> and MD at 300K (blue lines), calculated using SPARTA+<sup>7</sup> for M66 sequence. Chemical shifts outside the range of 0.5 ppm from NMR chemical shifts are shaded grey.

Table S1: Flory exponent ( $\nu$ ) and prefactor (A) for each proregion domain, calculated from MD trajectories. Errors represent fit error of  $\langle R_{|i-j|} \rangle$  vs  $An^\nu$ , weighted by each point's standard deviation.

domain	A (V66)	$\nu$ (V66)	A (M66)	$\nu$ (M66)
p1	$0.53 \pm 0.16$	$0.58 \pm 0.24$	$0.52 \pm 0.15$	$0.60 \pm 0.23$
h1a	$0.53 \pm 0.16$	$0.59 \pm 0.25$	$0.53 \pm 0.16$	$0.60 \pm 0.24$
h1b	$0.61 \pm 0.24$	$0.44 \pm 0.40$	$0.60 \pm 0.26$	$0.44 \pm 0.41$
p2	$0.56 \pm 0.22$	$0.50 \pm 0.34$	$0.59 \pm 0.18$	$0.51 \pm 0.29$
h2a	$0.56 \pm 0.17$	$0.50 \pm 0.22$	$0.57 \pm 0.17$	$0.48 \pm 0.23$
h2b	$0.64 \pm 0.14$	$0.51 \pm 0.18$	$0.60 \pm 0.15$	$0.56 \pm 0.20$
p3	$0.49 \pm 0.09$	$0.63 \pm 0.10$	$0.50 \pm 0.09$	$0.62 \pm 0.10$
h3b	$0.56 \pm 0.32$	$0.55 \pm 0.62$	$0.55 \pm 0.28$	$0.60 \pm 0.57$
h3c	$0.52 \pm 0.13$	$0.75 \pm 0.29$	$0.52 \pm 0.14$	$0.74 \pm 0.31$
h3d	$0.58 \pm 0.18$	$0.52 \pm 0.27$	$0.59 \pm 0.19$	$0.49 \pm 0.28$

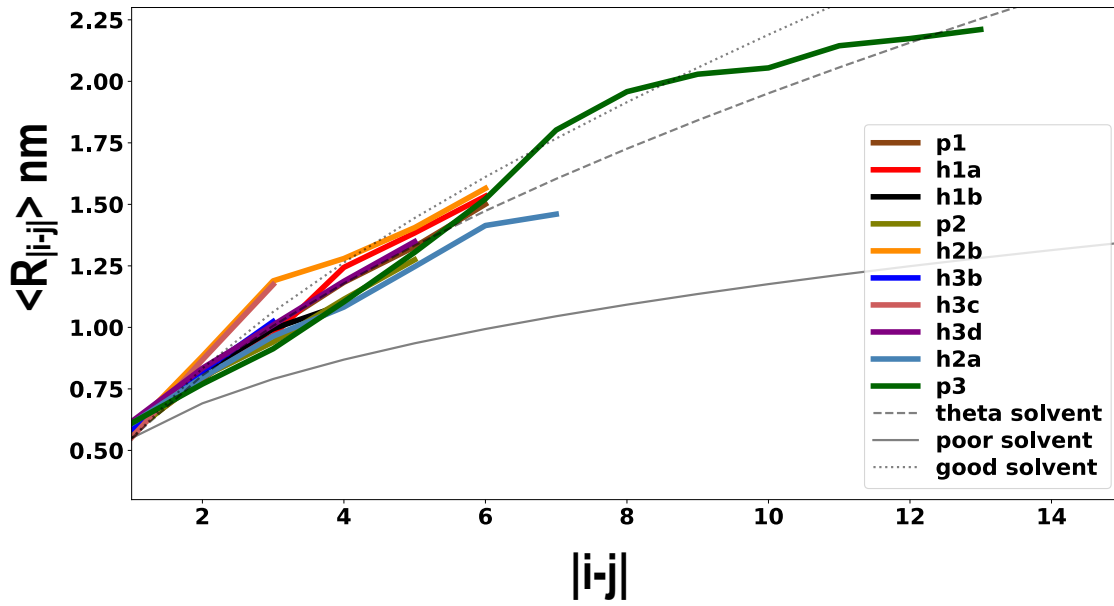


Fig S3: **Scaling behavior of each identified domain.** Ensemble averaged interchain distance profiles for each domain in V66 sequence. Theoretical polymer scaling limits are shown with grey lines (prefactor  $A = 0.55$  nm), while Flory exponents for each curve are given in Table S1.

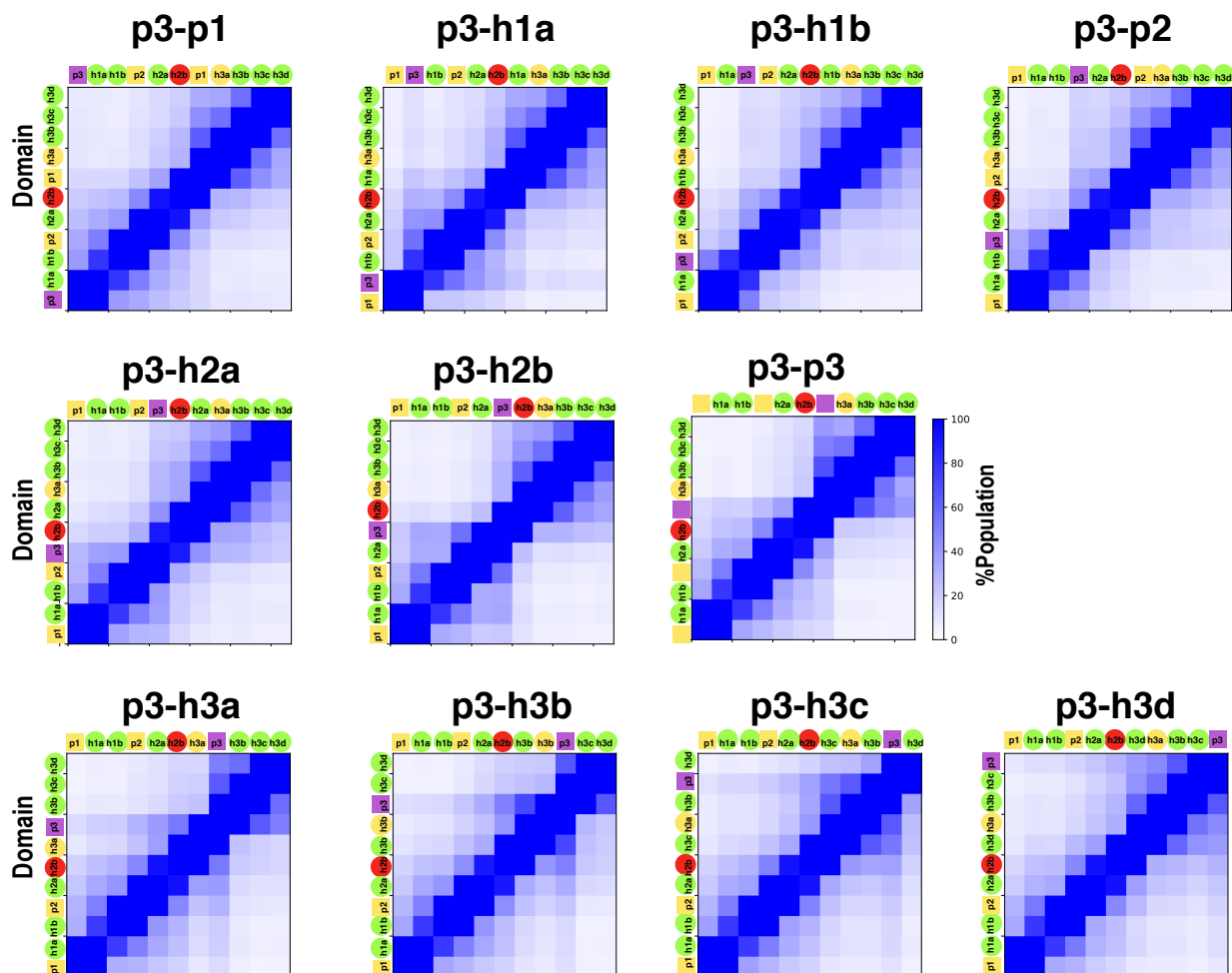


Fig S4: **Effect of perturbing monomer properties on freely-jointed, self-avoiding heteropolymer** Contact probability maps from MC simulations, analogous to those in Figure 3 of the main text, in which the monomer with properties of linker domain p3 is swapped with every other monomer on the chain, with the new location represented by the purple square in the graph annotation. As the p3 bead is shifted along the chain, p3 and p1 consistently bound a white “forbidden” region that has little interaction with the rest of the protein.

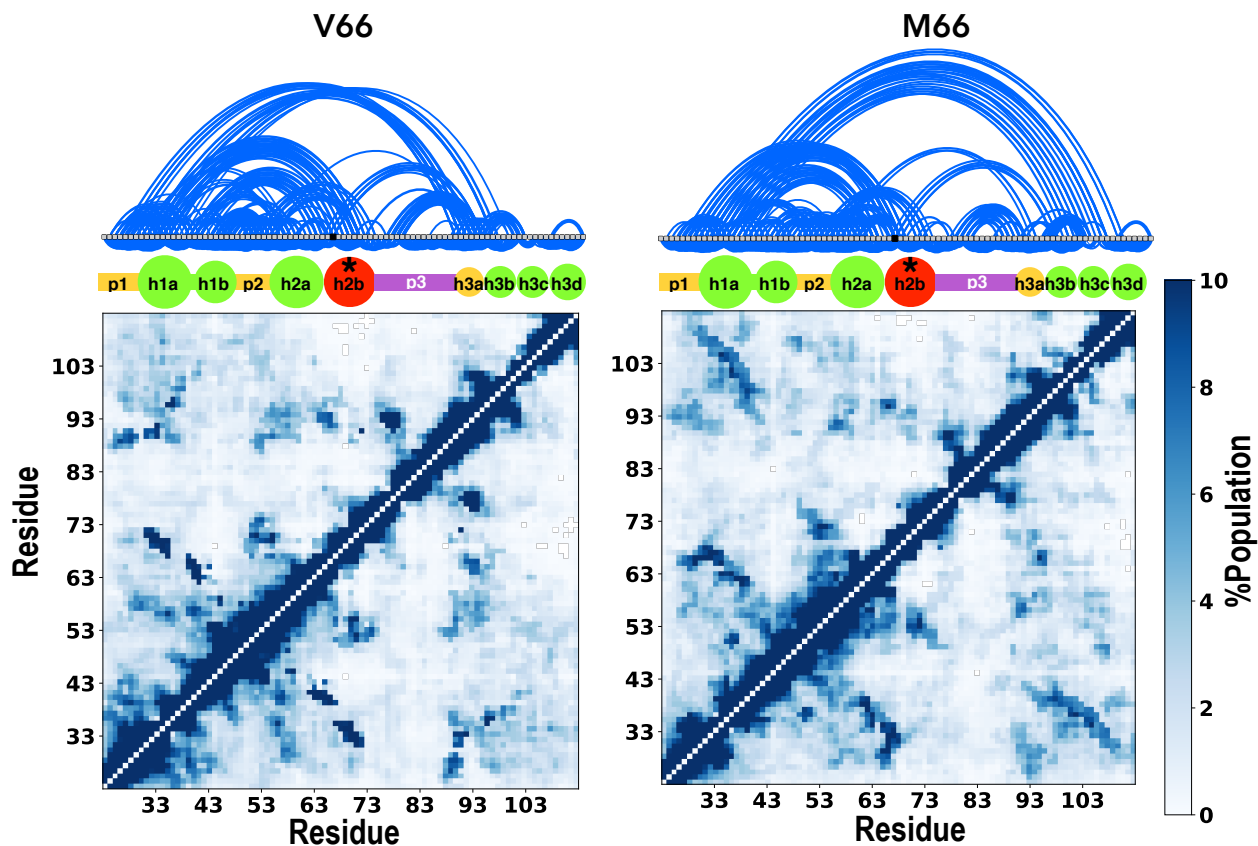


Fig S5: **Residue level contacts for the entire prodomain.** Contact probability between every residue pair for V66 (left) and M66 (right). A linear network of transient tertiary contacts is also shown (top) for each form. Two residue pairs are in contact if the distance between  $C_{\alpha}$ - $C_{\alpha}$  atoms between the two residues are 0.8nm or less. The contact networks were build using Cytoscape<sup>9</sup> 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.

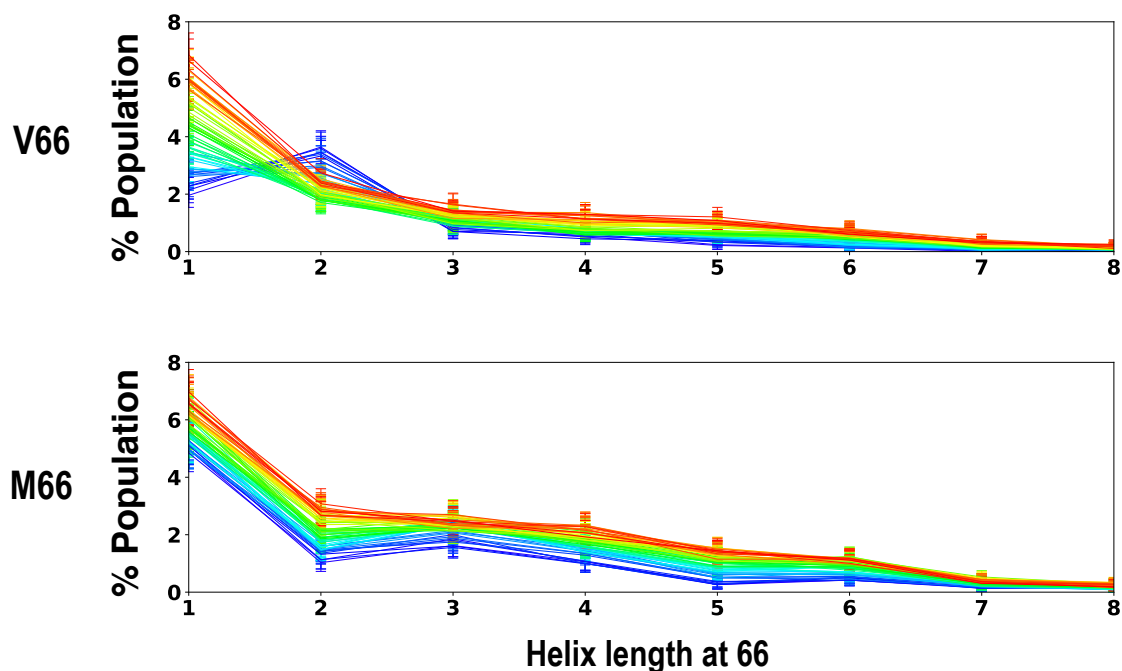


Fig S6: **Effects of temperature and Val66Met mutation on helix propensity around residue 66.** Frequency of helix of a given length at residue 66 in V66 (top) and M66 (bottom) in the temperature range of 300K to 385 K. With the increase in temperature the color transitions from cooler (blue) to hotter (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.

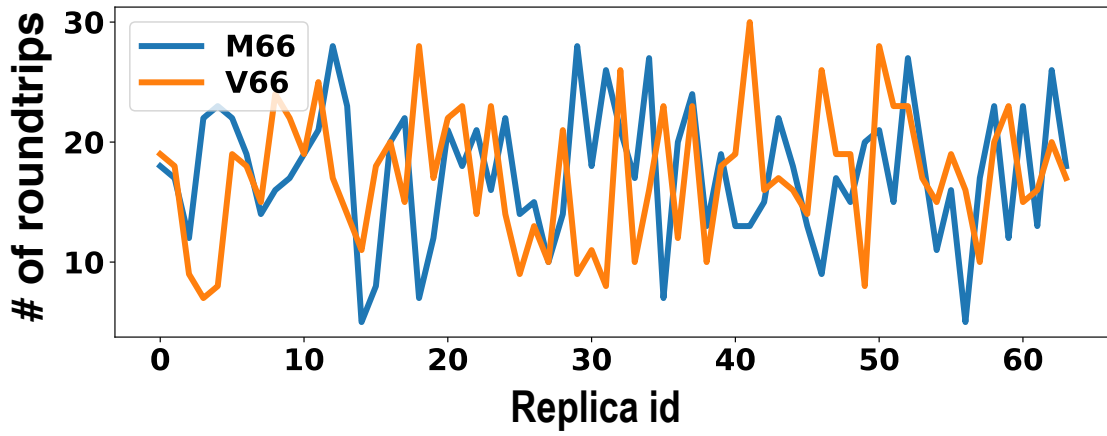


Fig S7: **Number of round trip completed by each replica.** Both V66 and M66 complete a minimum 5 roundtrips for each replica and an average of 17 round-tips per replica over the course of  $76.8 \mu s$  ( $1.2 \mu s \times 64$ ) of simulations.



## References

- (1) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950–8.
- (2) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins Struct. Funct. Bioinforma.* **2006**, *65*, 712–725.
- (3) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. *J. Phys. Chem. B* **2015**, *119*, 5113–5123.
- (4) Best, R. B.; Zheng, W.; Mittal, J. *J. Chem. Theory Comput.* **2014**, *10*, 5113–5124.
- (5) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.
- (6) Jorgensen, W. L. *J. Am. Chem. Soc.* **1981**, *103*, 335–340.
- (7) Shen, Y.; Bax, A. *J. Biomol. NMR* **2010**, *48*, 13–22.
- (8) Anastasia, A.; Deinhardt, K.; Chao, M. V.; Will, N. E.; Irmady, K.; Lee, F. S.; Hempstead, B. L.; Bracken, C. *Nat. Commun.* **2013**, *4*, 2490.
- (9) Ahlstrom, L. S.; Baker, J. L.; Ehrlich, K.; Campbell, Z. T.; Patel, S.; Vorontsov, I. I.; Tama, F.; Miyashita, O. *J. Mol. Graph. Model.* **2013**, *46*, 140–9.