Effects of Altered Backbone Composition on the Folding Kinetics and Mechanism of an Ultrafast-Folding Protein

Jacqueline R. Santhouse, Jeremy M. G. Leung, Lillian T. Chong*, W. Seth Horne*

Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15260.

SUPPORTING INFORMATION

Methods

Protein synthesis and purification. The synthesis and purification of **β³-H2** and **β³-H3** has been described previously.[1] The synthesis of **β³/βᶜʸᶜ-H2** was achieved by the same methods, employing automated Fmoc solid phase methods on NovaPEG Rink Amide Resin (0.1 mmol scale) using a Biotage Alstra synthesizer. Couplings were performed with Fmoc-protected amino acid in DMF (4 equiv. relative to resin, 0.2 M), HCTU in NMP (3.9 equiv., 0.2 M), and DIEA (6 equiv.) for 45 minutes. Fmoc deprotections were performed using 20% 4-methylpiperidine in DMF twice for 5 minutes each. Resin was washed 3 times with DMF between each step. After synthesis, the resin was washed with CH_2Cl_2 and methanol, dried under vacuum, and the peptide was cleaved from resin by treatment with a solution of trifluoracetic acid (TFA) / ethanedithiol / water / triisopropylsilane (94 / 2.5 / 2.5 / 1 by volume) followed by agitation for 3.5 hours. The resulting mixture was filtered and excess TFA evaporated under a stream of nitrogen. Protein was precipitated by addition of cold ether, centrifuged, and the liquid decanted. The resulting pellets were dissolved in solutions of 0.1% TFA in water (solvent A) and 0.1% TFA in acetonitrile (solvent B) and purified via preparative HPLC on a C18 column using gradients between 0.1% TFA in water and 0.1% TFA in acetonitrile. Fractions containing pure protein were combined and lyophilized. Identity of purified material was confirmed by ESI mass spectrometry and purity assessed by analytical HPLC (Figure S2).

NMR structure determination. The structure of **β³/βᶜʸᶜ-H2** was determined following methods described previously for the characterization of **WT**, **β³-H2**, and **β³-H3**.[1] Briefly, a sample was prepared consisting of 1.4 mM protein and 0.2 mM DSS in 9:1 H_2O / D_2O at pH 5 (uncorrected for presence of deuterium). Water-suppressed 1D ^1H and 2D ^1H/^1H NOESY (150 ms mixing time), TOCSY (70 ms mixing time), magnitude COSY, and DQF-COSY spectra were acquired on a Bruker Avance 700 MHz spectrometer at 283 K. Spectra were processed in Topspin and chemical shifts referenced to DSS. Resonances were assigned manually using NMRFAM-SPARKY.[2] Structure determination was performed by simulated annealing using ARIA[3] in conjunction with CNS.[4] ARIA settings were modified from defaults to improve model quality and convergence, as described.[5] H-bond restraints for helices were generated based on manual analysis of medium-range NOEs, and backbone dihedral restraints prepared based on $^3J_{H-H}$ coupling constants determined from well-resolved signals in the 1D ^1H or phase-sensitive COSY.[6] NOE distance restraints were generated by the ARIA program in fully automated fashion, starting from a list of ^1H resonances and an unassigned set of integrated NOESY peaks. The set of ten lowest energy structures resulting from the calculation was taken as the final NMR ensemble. Coordinates and additional experimental data are deposited in the PDB (7URJ) and BMRB (31012).

NMR kinetics measurements. Commercially available urea-d_4 and thiourea-d_4 (98 atom % D) were further isotopically enriched by dissolving in D_2O (99.9 atom % D) followed by lyophilization. This process was repeated a minimum of three times, and the resulting material used to prepare solutions described below. Protein samples were deuterated by dissolving 15-20 mg of purified lyophilized synthetic protein in a solution composed of 3.33:1 urea / thiourea at a total concentration of 8.8 M in D_2O. The solution was allowed to stand for 1 hour, then solvent exchanged by seven consecutive rounds of dilution with pure D_2O to a volume ~12 mL followed by centrifugal concentration (3 kDa membrane cutoff) to a volume ~2 mL. After the last round of concentration, the protein solution was split into two equal volume aliquots and lyophilized. One lyophilized aliquot was dissolved in a solution composed of 20 mM sodium acetate-d_3, 100 mM NaCl, 0.2 mM DSS in D_2O and adjusted to pH 5 ± 0.1 (uncorrected for presence of deuterium);

the other lyophilized aliquot was dissolved in an equal volume of a solution composed of 20 mM sodium acetate-d_3, 100 mM NaCl, 8.8 M 3.33:1 urea / thiourea, 0.2 mM DSS in D_2O and adjusted to pH 5 ± 0.1 (uncorrected for presence of deuterium). The resulting two solutions, identical in composition save the denaturant, were mixed in different ratios to produce samples at a range of denaturant concentrations. 1D ^1H spectra were acquired at 283 K on a Bruker Avance III 700 MHz spectrometer (512 scans, 65,536 data points in the FID, 11,261 Hz sweep width). Residual water signal was suppressed using an excitation sculpted gradient pulse sequence. Each spectrum was processed in TopSpin, manually phased, baseline corrected, calibrated to DSS signal, intensity normalized to DSS signal, and exported in ASCII format. Line shape analysis was performed using Mathematica, following a mathematical model and fitting methods detailed previously.[7-9] Most samples showed sharp signals in the region of interest corresponding to species not engaged in the rapid chemical exchange process. This has been noted before for other BdpA mutants and attributed to *cis* amide isomers for the denatured state; the presence of these species was accounted for in fitting to the chemical exchange model as described.[7,8]

Generation of unfolded state conformations. For simulations of each protein folding process, initial unfolded state conformations were previously generated by two stages of WE simulation.[1] In the first stage, the equilibrated structure of the folded protein was unfolded at room temperature using a WE simulation. In the second stage, a set of 5-8 of the most unfolded conformations (based on RMSD from the folded state and fraction of native contacts in the hydrophobic core) were selected as starting conformations for five independent WE simulations to extensively sample the unfolded state ensemble. The resulting unfolded-state conformations (within regions of -lnP < 4, where P is the probability as a function of the fraction of native contacts and radius of gyration) were clustered based on pairwise "best-fit" C_α RMSD of the three helices and conformation closest to cluster centers were selected as initial unfolded conformations with the corresponding cluster weight for WE simulations of the folding process. Each unfolded conformation was modeled using the AMBER ff15ipq-m force field for protein mimetics[10,11] and immersed in a sufficiently large truncated octahedral box of SPC/E$_b$ explicit water molecules[12] to provide a minimum clearance of 15 Å between the protein and box walls. Counterions were added to neutralize the net charge of the simulation system and achieve salt concentrations consistent with experimental conditions (100 mM NaCl and 20 mM NaOAc). Joung & Cheatham parameters[13] for the SPC/E water model were used for Na^+ and Cl^- ions, and AMBER ff15ipq parameters were used for OAc^- ions.

Prior to running WE simulations, each solvated system was first subjected to energy minimization followed by two stages of solvent equilibration while applying harmonic positional restraints to the heavy atoms of the proteins with a force constant of 1 kcal mol^{-1} Å^{-2}. In the first stage, the restrained system was heated from 0 K to 298 K for 25 ps in an NVT ensemble using a weak Langevin thermostat with a collision frequency of 1 ps^{-1}. In the second stage, the solvent was subjected to 1 ns equilibration in an NPT ensemble using the same thermostat and a Monte Carlo barostat with pressure changes attempted every 0.2 ps. The entire system was then equilibrated without any restraints for 1 ns in the NPT ensemble. For further details, see the section below.

Weighted ensemble simulations of protein folding. All weighted ensemble (WE) simulations were run using the WESTPA 2.0 software package.[14] For each BdpA variant, we carried out five independent WE simulations of the folding process, starting from unfolded conformations obtained from room-temperature WE simulations of the unfolding process, as described above. For each WE simulation, we used a one-dimensional progress coordinate consisting of the heavy-atom RMSD of all three helices from the equilibrated NMR structure after

alignment on helix 3. Helix 3 was chosen for alignment in the RMSD calculation, because this helix remains the most folded among the three helices over the course of a 3.6-us WE unfolding simulation of the WT BdpA despite explicitly using the RMSD of Helix 3 as a progress coordinate.[1] The minimal adaptive binning (MAB) scheme was applied with 15 bins, a target number of 5 trajectories per bin, and a resampling time interval of 100 ps. To maintain non-equilibrium steady state conditions, trajectories reaching the target folded state were "recycled" by initiating new trajectories from an unfolded conformation with the same trajectory weight. Using 24 NVIDIA V100 GPUs in parallel, we generated an aggregate simulation time of 66 μs (13 days) for **WT**, 114 μs (22 days) for **β³-H2**, 176 μs (34 days) for **β³-H3**, and 62 μs (10 days) for **β³/βᶜʸᶜ-H2**.

Dynamics were propagated using the AMBER 20 software package.[15] Given that the WE strategy is rigorous with stochastic dynamics,[16] a weak stochastic thermostat was used (i.e. Langevin thermostat with a collision frequency of 1 ps⁻¹) to maintain a constant temperature of 25 °C. To maintain a constant pressure of 1 atm, a Monte Carlo barostat was applied with pressure changes attempted every 0.2 ps. To enable a 2-fs timestep, all bonds to hydrogens were restrained to their equilibrium values using the SHAKE algorithm.[17] Short-range nonbonded interactions were calculated using a cutoff of 10 Å and the particle mesh Ewald method[18] was applied to treat long-range electrostatics.

Reweighting simulations for a steady state. To further accelerate convergence of the WE simulations to a non-equilibrium steady state, we applied a history-augmented Markov state model (haMSM) iterative-reweighting approach,[19] in which trajectories were iteratively reweighted and restarted to converge the haMSM, i.e. estimate folding rate-constants that were reasonably converged to a steady state (Figure S12-S19). The first cycle of reweighting was applied after 450 WE iterations when all five WE simulations had generated folding events. Subsequent cycles of reweighting were applied after the generation of an additional 300 WE iterations by at least three of the WE simulations. The reweighting procedure involved first discretizing the trajectories according to a set of "microbins" ("microstates" in MSM terminology) by "stratifying" a set of fixed bins along the progress coordinate, generating microbins within each bin and then constructing a history-augmented Markov state model (haMSM). This stratification avoids the generation of microbins that span large regions of the progress coordinate. Microbins were generated based on a metric that is distinct from the progress coordinate, i.e. the pairwise heavy-atom RMSD of all three helices after alignment on helix 3 of an equilibrated structure of the folded state.

For each cycle of the reweighting procedure, all trajectory weights and conformations (sampled every 100 ps) from that WE iteration were used to construct an haMSM, and the resulting steady-state estimates of the stationary population were used to reweight each conformation based on Algorithm 5.3 from Aristoff & Zuckerman,[20] which distributes each microbin population to its constituent conformations proportional to their original WE weights using the following equation:

$$\omega_{new}^i = pSS_p \cdot \omega^i / \Sigma_{j \epsilon p} \omega^j$$

where ω^i is the statistical weight of trajectory i, p is the microbin occupied by trajectory i, and pSS_p is the estimated equilibrium state population of microbin p. A new set of trajectories was then restarted from the reweighted conformations, with velocities and other simulation conditions preserved to maintain the continuity of pathways.

State and helix definitions. For all analysis, a common set of definitions for the folded and unfolded states were defined based on our folding simulations for the entire set of BdpA variants. The folded state was defined as the region with -lnP < 4 where P is the probability as a function of the fraction of tertiary native contacts and radius of gyration (R_g) from the unfolding simulations (Figure S11); this region corresponds to >60% native contacts and an R_g of 11-14 Å. The unfolded state was defined to be within the region with -lnP < 4 and < 20% interhelical native contacts. The three helices were defined as follows: residues 6-17 for helix 1, residues 24-36 for helix 2, and residues 41-54 for helix 3.

Pseudo-committor analysis. To order conformations from multiple folding trajectories along a "kinetic ruler", we calculated the probability of each conformation committing to the folded state before reaching the unfolded state. Given that our simulations were run under non-equilibrium steady state conditions, we refer to these probabilities as "pseudo-committor" values. Pseudo-committor values were calculated from the haMSM using the msm_we package[19] by numerically solving for the stationary distribution with an initial dual-absorbing boundary condition in which the probabilities of the source (unfolded state) and sink states (folded state) for which the probabilities were each set to one and the probabilities of all other microbins were set to zero. The initial distribution was then propagated (multiplied) with the transition matrix until a convergence of < 10^{-5} was reached.

Calculation of the folding rate constant. The folding rate constant k_f for each BdpA variant was calculated as averages of estimates from three haMSMs that were constructed with the same microbin cluster centers, but using different subsets of the data. The first haMSM was constructed from the second half of each restarted set of trajectories. The remaining two haMSMs were based on two cross-validation models that were each built using odd or even subsets of the same restarted set of trajectories, with the parity assigned by evenly splitting each restart into 10 evenly sized blocks based on their iteration count. The k_f value was calculated from steady state fluxes for trajectories entering the folded state using the Hill relation:[21]

$$k_f = \frac{1}{MFPT(unfolded \rightarrow folded)}$$

where the mean first-passage time (MFPT) for transitions from the unfolded state to the folded state is equal to the inverse of the conditional probability flux into the folded state. Uncertainties in the k_f values represent 95% credibility regions, as estimated using a Bayesian bootstrapping approach.[22] Unlike the haMSM generated for restarts, the microbins for these post-simulation haMSMs were generated using C_α-C_α pairwise distance matrices that were dimensionally reduced using a variational approach for Markov processes (VAMP).[23]

Probability maps of residue-level contacts. To generate probability maps of pairwise residue contacts for a selected ensemble of conformations, a heavy-atom distance matrix for each protein conformation in the ensemble was first calculated using the cpptraj program of the AMBER 20 software package.[15] Next, the matrix of each conformation was assigned the corresponding statistical weight from the WE simulation to generate a weighted-average probability map of pairwise residue contacts for the ensemble of conformations. A contact was defined as a pair of residues with ≤ 5 Å distance between heavy atoms and referred to as a native contact if the contact was present in a reference folded conformation that is closest to the average structure of the folded state ensemble generated by WE simulation.
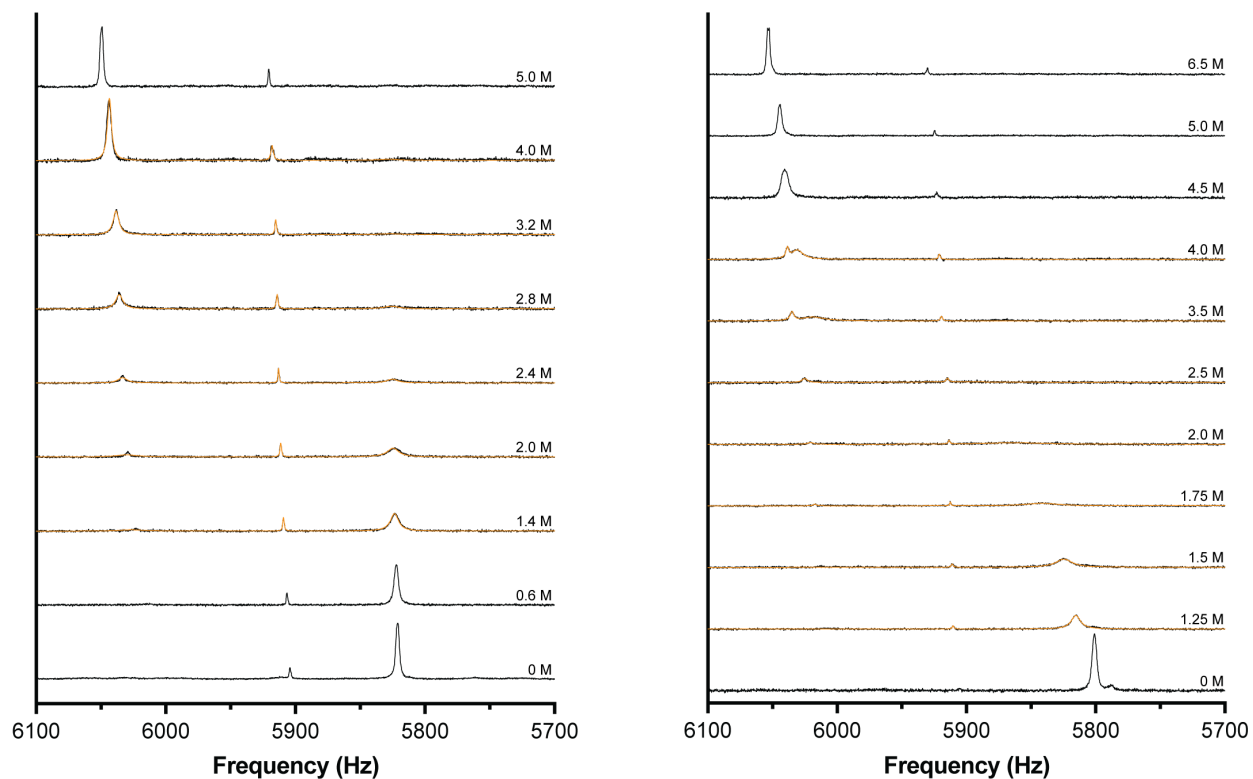
Figure S1. ^{1}HNMR signal for H_{ε} from residue His[18] in **β³-H2** (left) and **β³-H3** (right) at 700 MHz field strength as a function of concentration of a denaturant composed of urea and thiourea in a 3.33 : 1 molar ratio. Orange lines depict the result of lineshape fitting over the unfolding transition region used to extract the individual folding and unfolding rate constants shown in Figure 2.
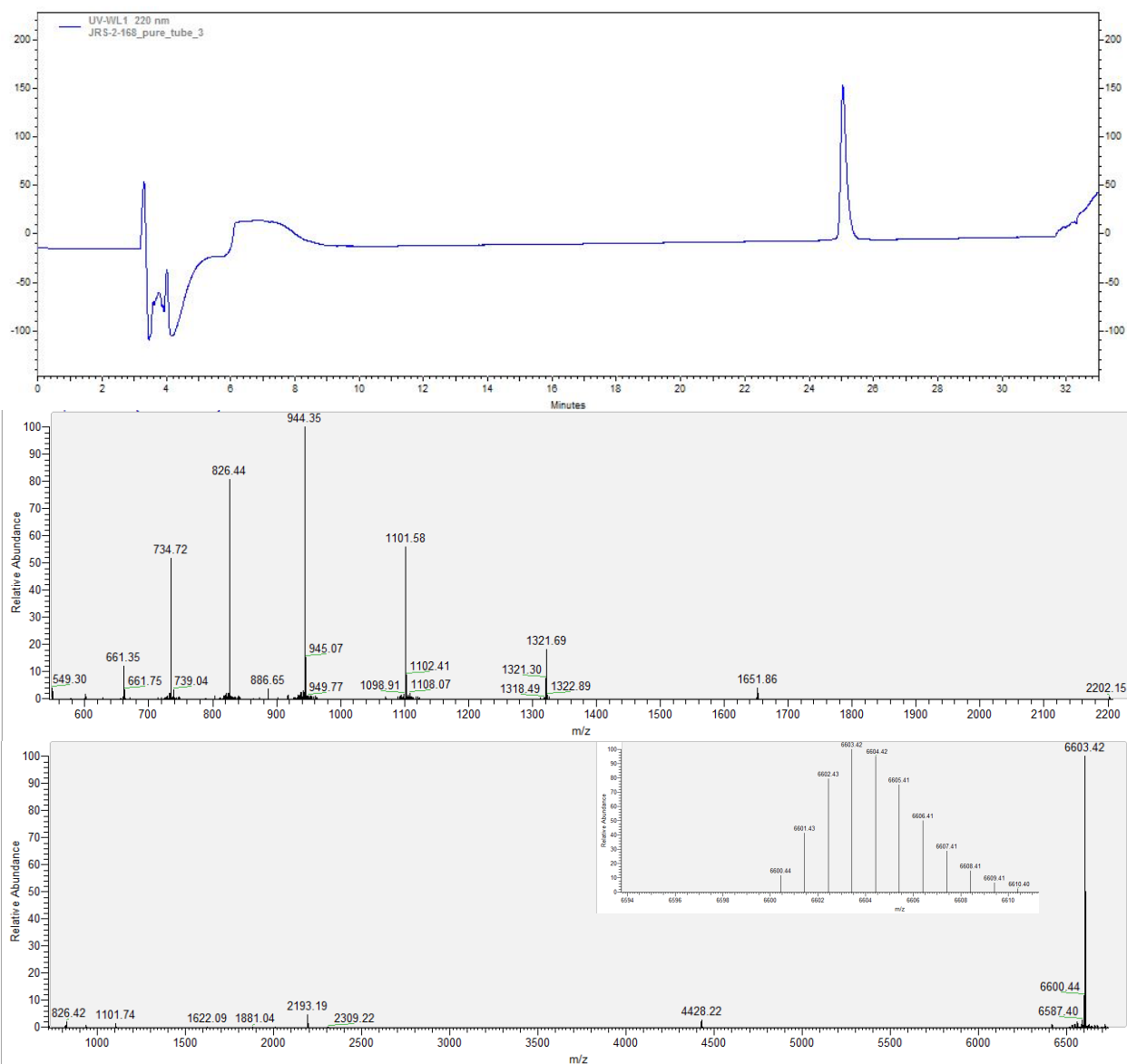
Figure S2. Analytical HPLC (top), raw ESI-MS (middle), and deconvoluted ESI-MS for species [M]$^+$ (bottom) for purified **β³/β^{cyc}-H2** (monoisotopic [M]$^+$ m/z calc. = 6600.4).

Table S1. Statistics from NMR structure calculations for β^3/β^{cyc}-H2.

PDB Accession Code	7URJ
Experimental restraints	
Unambiguous NOEs	772
Intra-residue	419
Sequential ($|i - j| = 1$)	144
Medium-range ($1 < |i - j| < 5$)	103
Long-range ($|i - j| \geq 5$)	106
Ambiguous NOEs	197
Total NOEs	969
H-bonds	56
Dihedrals	2
Violations	
NOE >0.5 Å	32.0 ± 2.5
NOE rmsd (Å)	0.16 ± 0.01
H-bond >0.5 Å	0
Dihedral >5°	0
Ensemble rmsd	
Backbone heavy atoms	0.89 ± 0.31
All heavy atoms	1.10 ± 0.22
Geometry analysis	
rmsd bonds (Å)	0.00320 ± 0.00007
rmsd angles (°)	0.463 ± 0.009
rmsd impropers (°)	1.17 ± 0.07
Ramachandran analysis[a]	
Favored (%)	91.5
Allowed (%)	8.3
Disallowed (%)	0.2

[a] Performed using the MolProbity server;[24] artificial residues excluded.
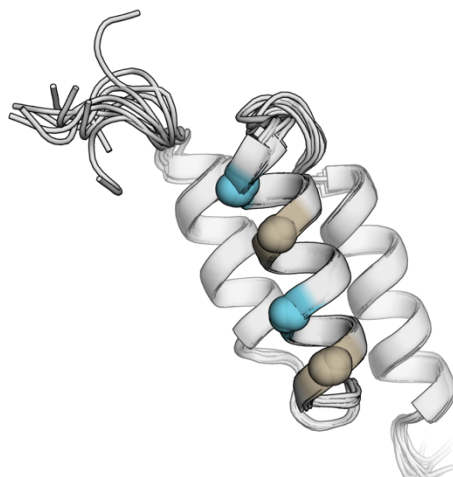
Figure S3. NMR structure ensemble for **β³/βᶜʸᶜ-H2**. The ensemble of 10 models resulting from simulated annealing with NMR-derived restraints is shown in a cartoon representation. Artificial residues are indicated with spheres colored according to the scheme in Figure 1.
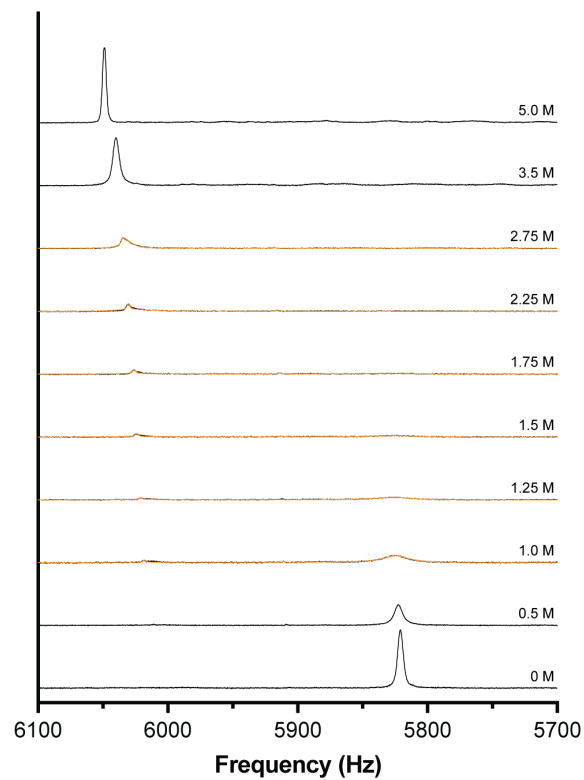
Figure S4. ^{1}HNMR spectra in the frequency region for H_ε from residue His[18] in **β³/β^{cyc}-H2** at 700 MHz field strength as a function of concentration of a denaturant composed of urea and thiourea in a 3.33 : 1 molar ratio. Orange lines depict the result of lineshape fitting over the unfolding transition region used to extract the individual folding and unfolding rate constants shown in Figure 2.

Table S2. Folding rate constants determined from simulations at 25 °C.

	WT	β^3-H2	β^3-H3	β^3/β^{cyc}-H2
$k_f \times 10^3$ (s^{-1})	110 ± 90	3.4 ± 2.9	32 ± 6	2.1 ± 2.5
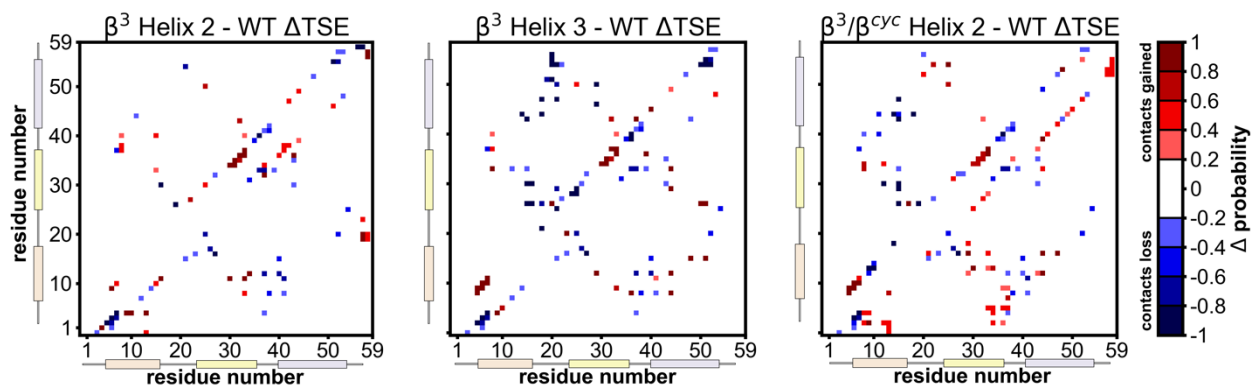
Figure S5. Difference probability maps of residue-level contacts in the transition state ensemble of each BdpA variant relative to that of **WT**. The upper-left triangle shows native contacts present in the reference folded structure while the lower-right triangle shows non-native contacts absent in the reference structure. Residues are considered in contact when a pair of residues contains heavy atoms are within 5 Å. Shades of red indicate the extent to which contacts are present in the variant but not in **WT**, whereas shades of blue indicate the extent to which contacts are present in **WT** and not the variant.
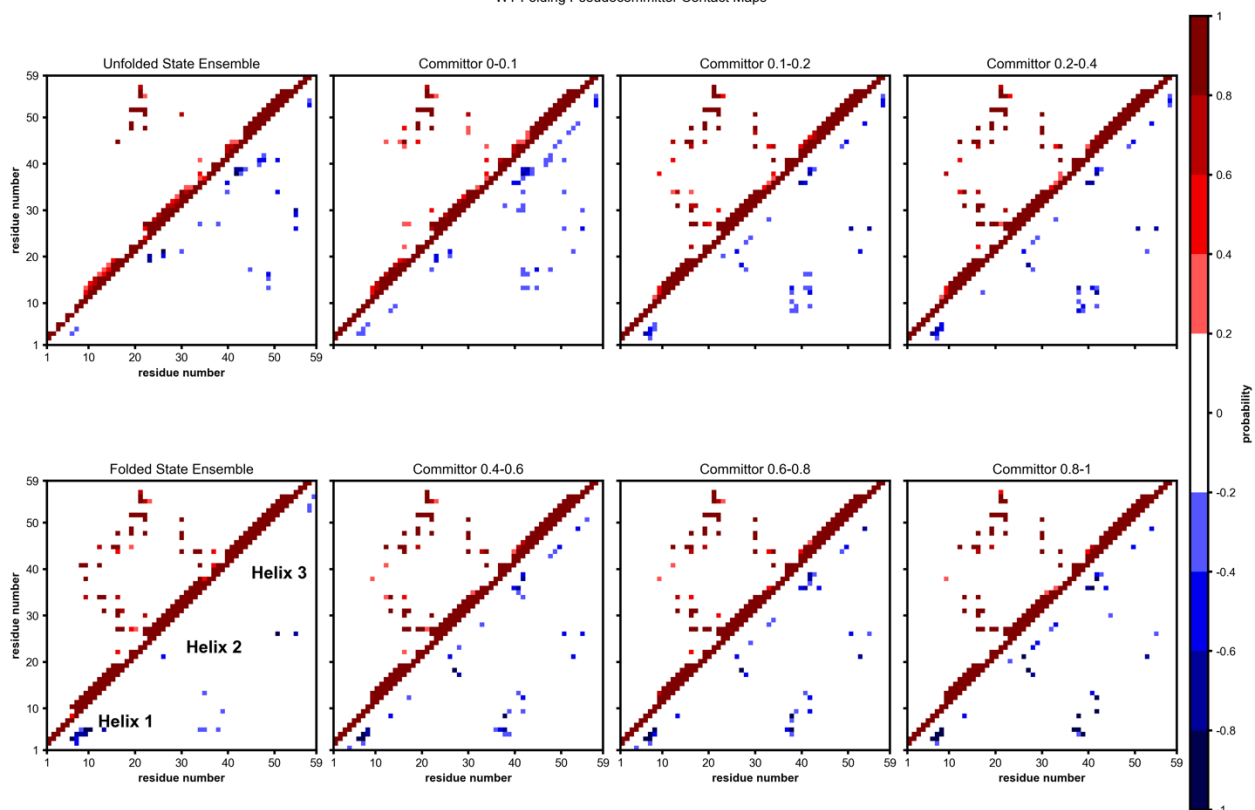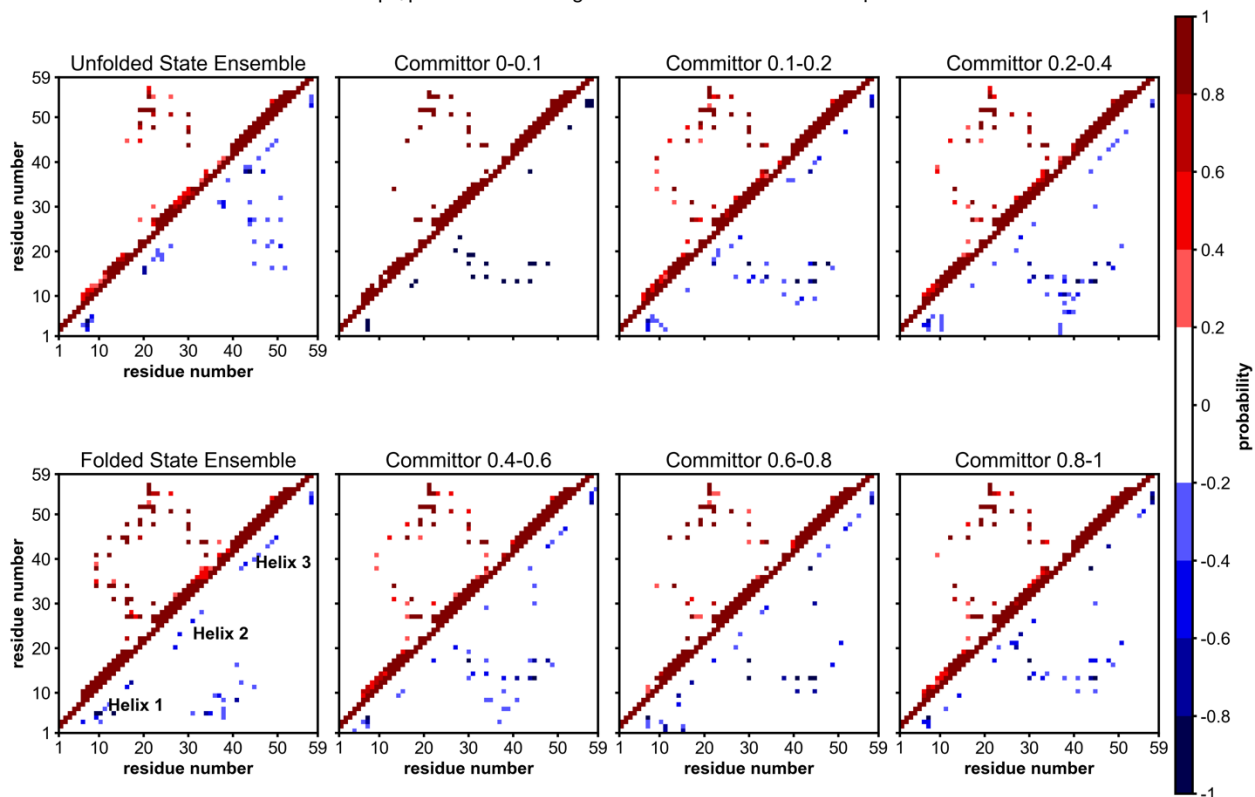
Figure S6. Probability maps of residue-level contacts for **WT** for the folded ensemble, unfolded ensemble, and ensembles with various 0.2-intervals of pseudocommittor values. The region above and left of the diagonal shows probabilities in shades of red for contacts present in the reference folded structure (i.e., "native") while the region below and right of the diagonal shows probabilities in shades of blue for contacts absent in the reference structure (i.e., "non-native"). Residues are considered in contact when the residue pair contains heavy atoms within 5 Å.
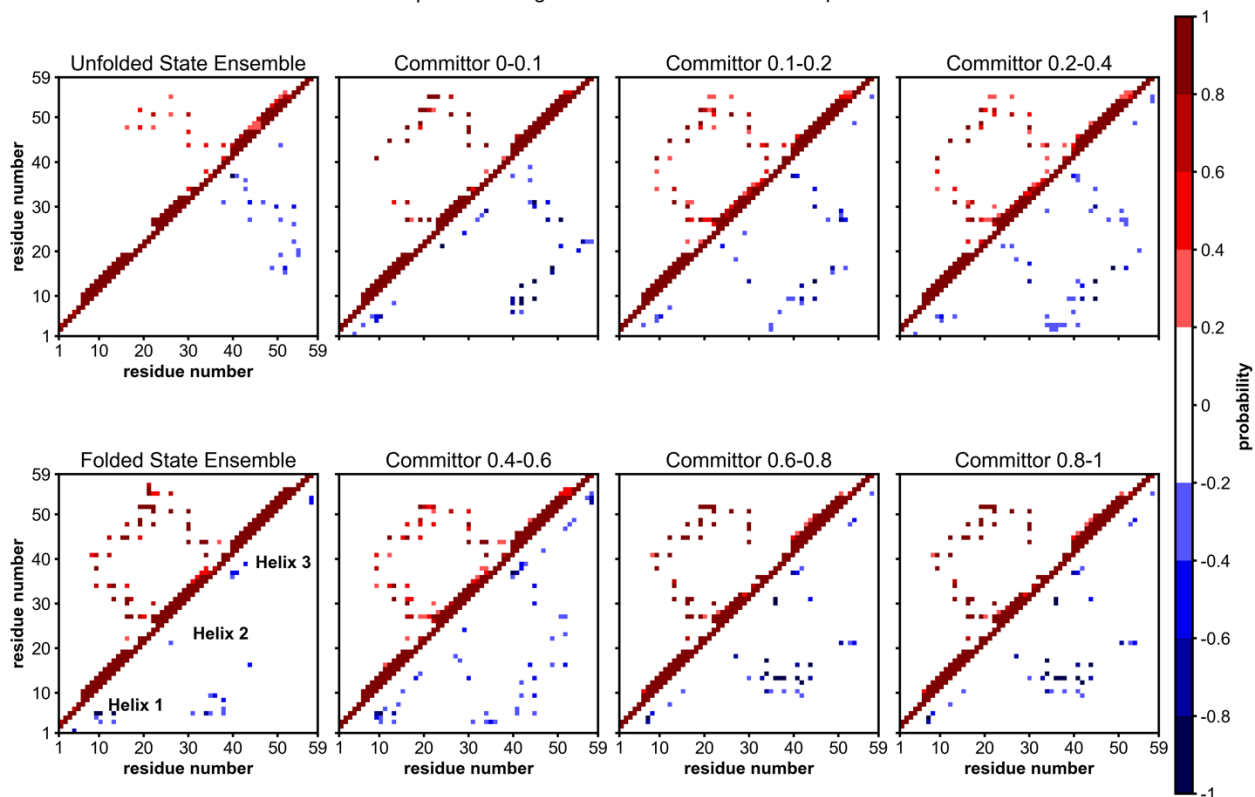
Figure S7. Probability maps of residue-level contacts for **β³-H3** for the folded ensemble, unfolded ensemble, and ensembles with various 0.2-intervals of pseudocommittor values. All other details are as reported for Figure S6.
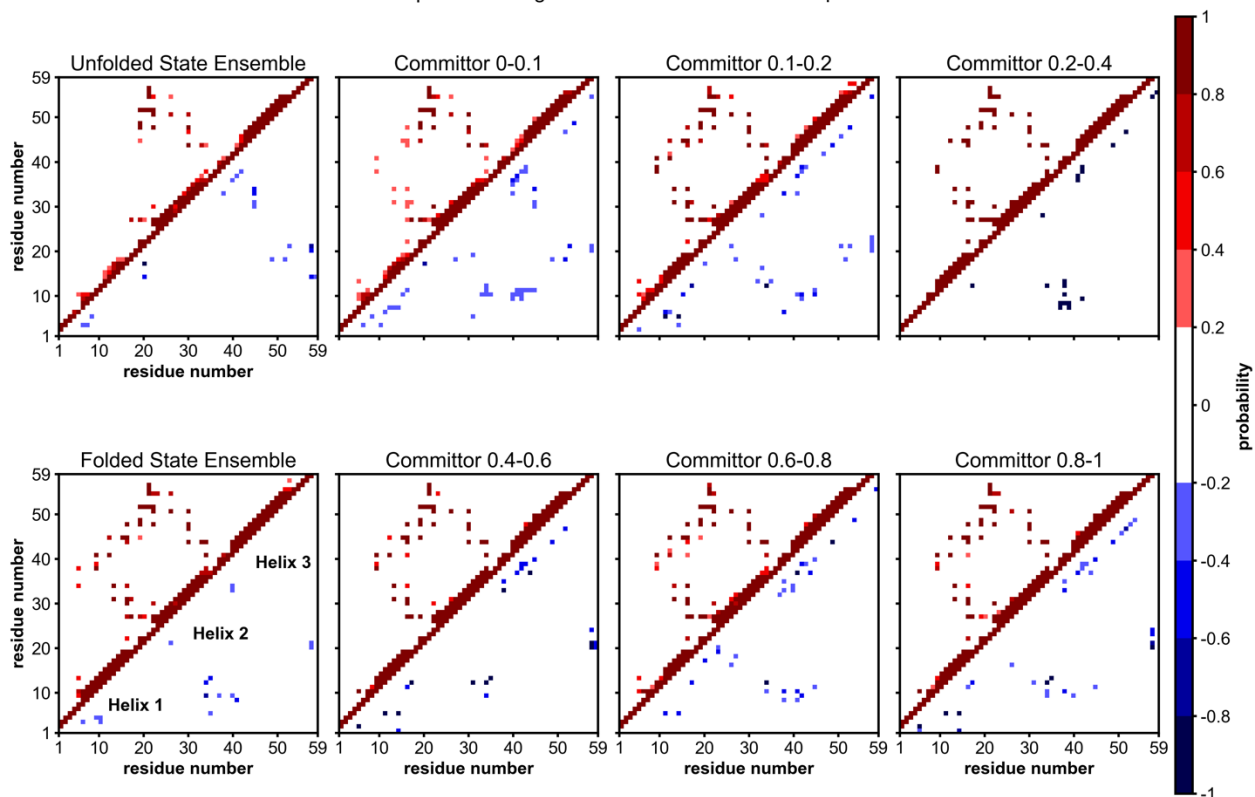
Figure S8. Probability maps of residue-level contacts for **β³-H2** BdpA for the folded ensemble, unfolded ensemble, and ensembles with various 0.2-intervals of pseudocommittor values. All other details are as reported for Figure S6.

Figure S9. Probability maps of residue-level contacts for **β³/βᶜʸᶜ-H2** BdpA for the folded ensemble, unfolded ensemble, and ensembles with various 0.2-intervals of pseudocommittor values. All other details are as reported for Figure S6.
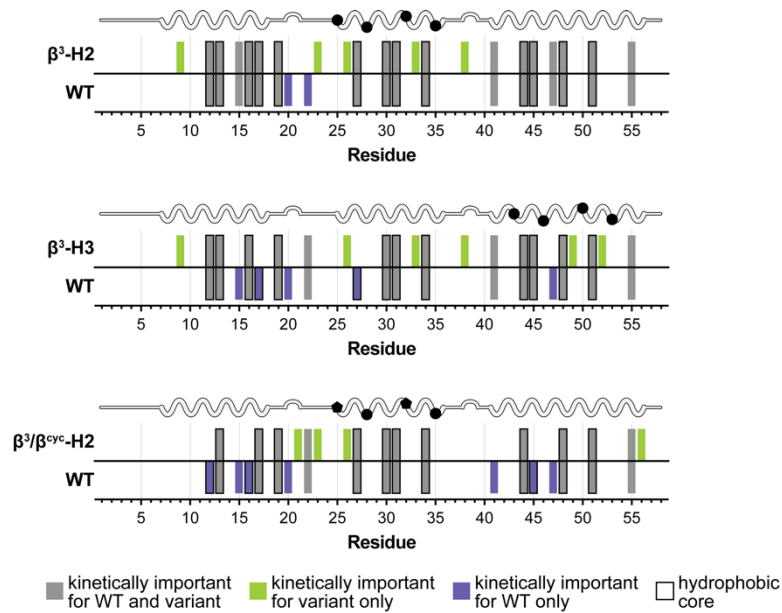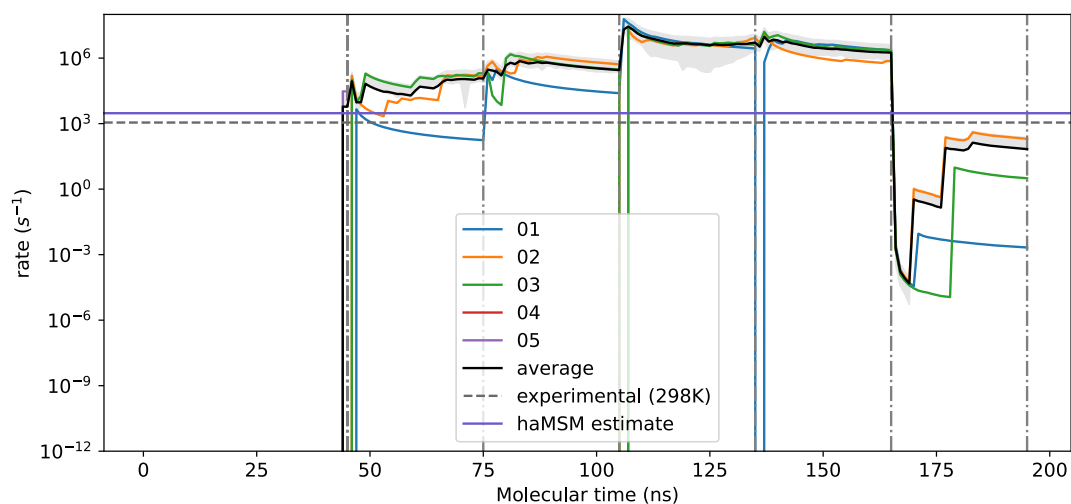
Figure S10. Pairwise comparison of kinetically important residues observed in the simulated folding of each heterogeneous-backbone BdpA variant. Bars above the horizontal line show a residue kinetically important to the indicated variant and bars below the horizontal line a residue kinetically important to **WT**. Bars are colored based on whether the residue is shared between variant and **WT** or unique to one of the two. Outlined bars depict hydrophobic core residues.
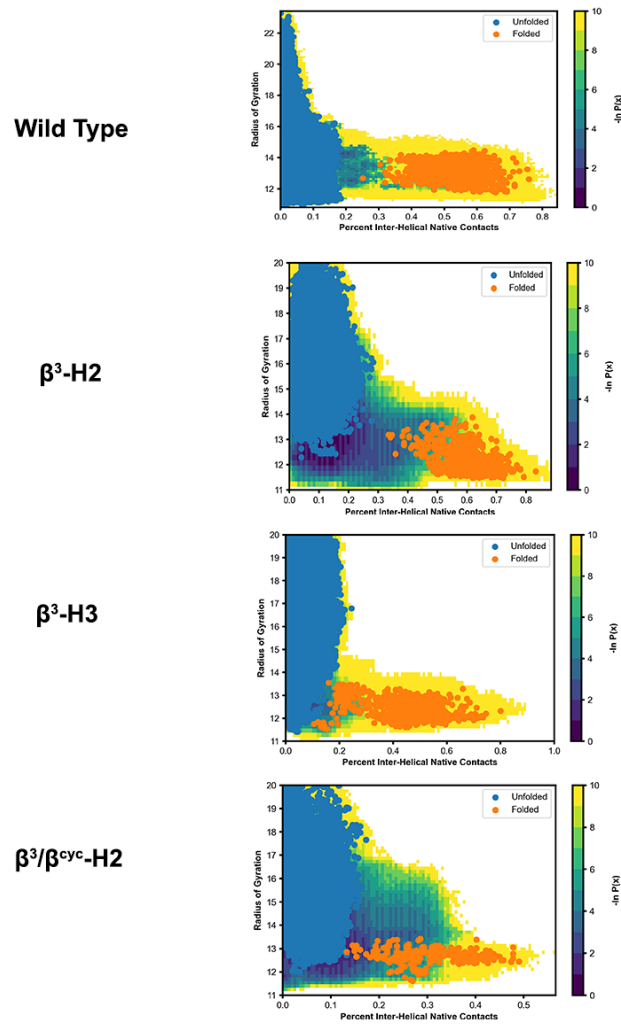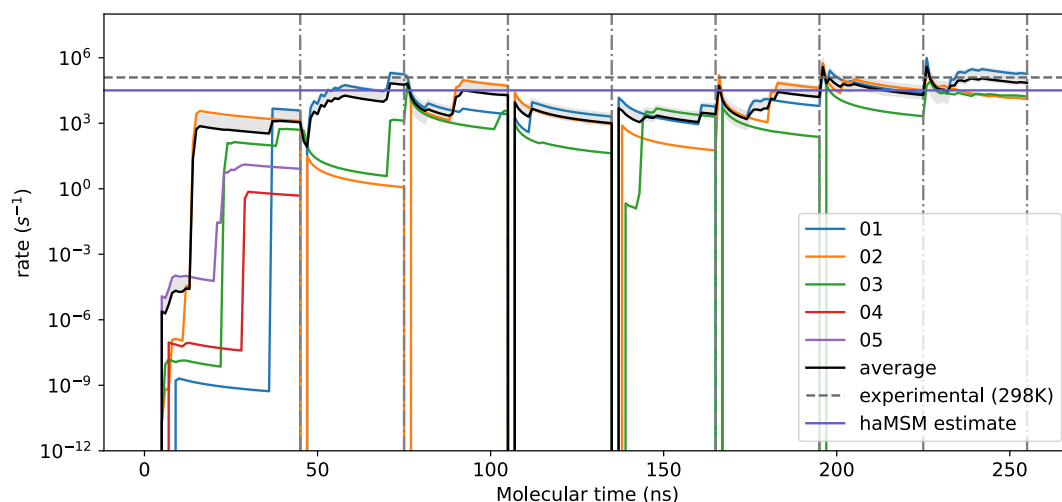
Figure S11. Definitions of the unfolded and folded states based on probability distributions as a function of the radius of gyration and percent inter-helical native contacts.
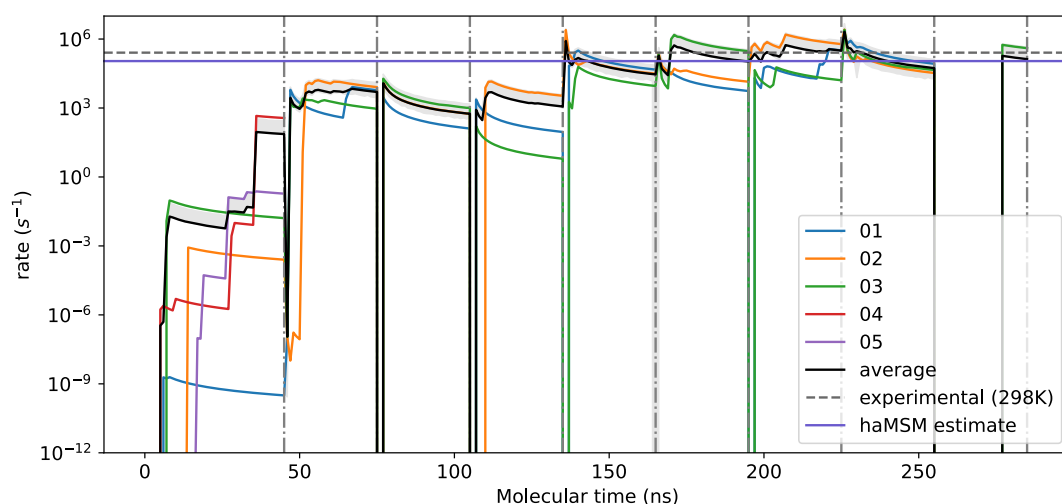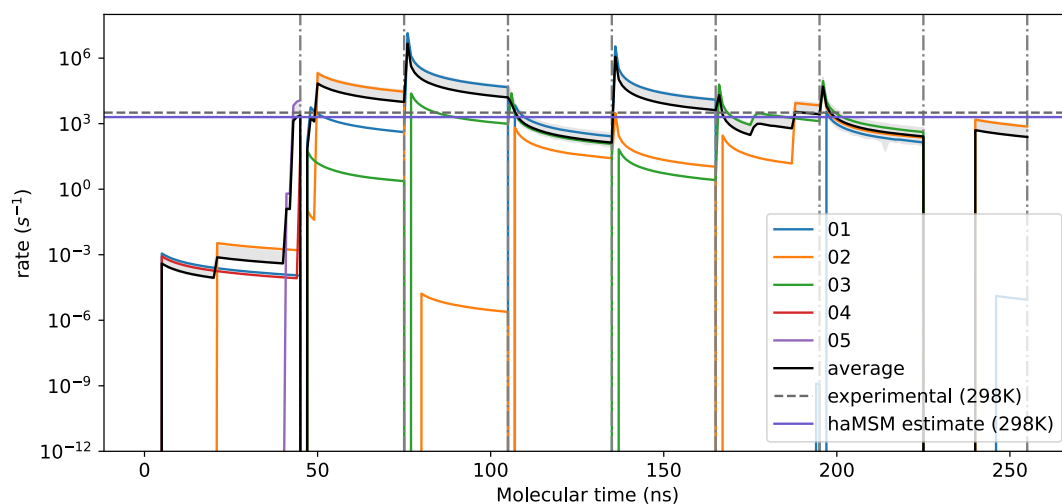
Figure S12. WT BdpA folding rate constants as a function of molecular time, as estimated directly from cumulative fluxes. Dotted vertical lines indicate a new set of simulations restarted from reweighted structures from an haMSM built the previous restart. The data from the first 50 iterations from each restart (as part of the "ramp up") were removed for clarity. The black curve and shaded regions represent the average across individual trials ± 95% credibility region from Bayesian bootstrapping. The grey horizontal dotted line is the experimental rate constant as reported in **Table 1**. The purple line is the final haMSM estimate as reported in **Table S2** (see Methods for more details).

Figure S13. β³-H2 BdpA folding rate constants as a function of molecular time, as estimated directly from cumulative fluxes. Dotted vertical lines indicate a new set of simulations restarted from reweighted structures from an haMSM built the previous restart. The data from the first 50 iterations from each restart (as part of the "ramp up") were removed for clarity. The black curve and shaded regions represent the average across individual trials ± 95% credibility region from Bayesian bootstrapping. The grey horizontal dotted line is the experimental rate constant as reported in **Table 1**. The purple line is the final haMSM estimate as reported in **Table S2** (see Methods for more details).
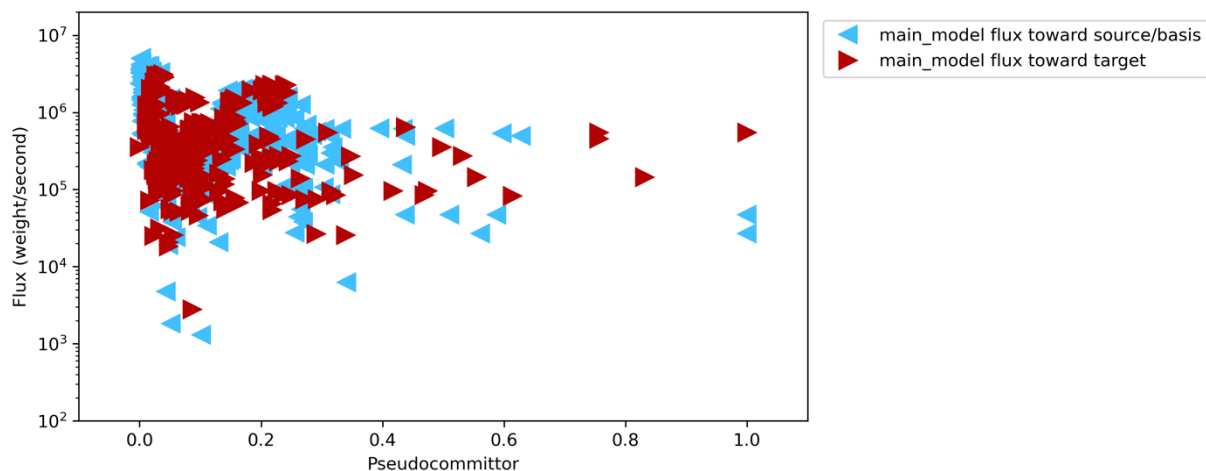
19

Figure S14. β³-H3 BdpA folding rate constants as a function of molecular time, as estimated directly from cumulative fluxes. Dotted vertical lines indicate a new set of simulations restarted from reweighted structures from an haMSM built the previous restart. The data from the first 50 iterations from each restart (as part of the "ramp up") were removed for clarity. The black curve and shaded regions represent the average across individual trials ± 95% credibility region from Bayesian bootstrapping. The grey horizontal dotted line is the experimental rate constant as reported in **Table 1**. The purple line is the final haMSM estimate as reported in **Table S2** (see Methods for more details).


Figure S15. β³/β^cyc-H2 BdpA folding rate constants as a function of molecular time, as estimated directly from cumulative fluxes. Dotted vertical lines indicate a new set of simulations restarted from reweighted structures from an haMSM built the previous restart. The data from the first 50 iterations from each restart (as part of the "ramp up") were removed for clarity. The black curve and shaded regions represent the average across individual trials ± 95% credibility region from Bayesian bootstrapping. The grey horizontal dotted line is the experimental rate constant as reported in **Table 1**. The purple line is the final haMSM estimate as reported in **Table S2** (see Methods for more details).
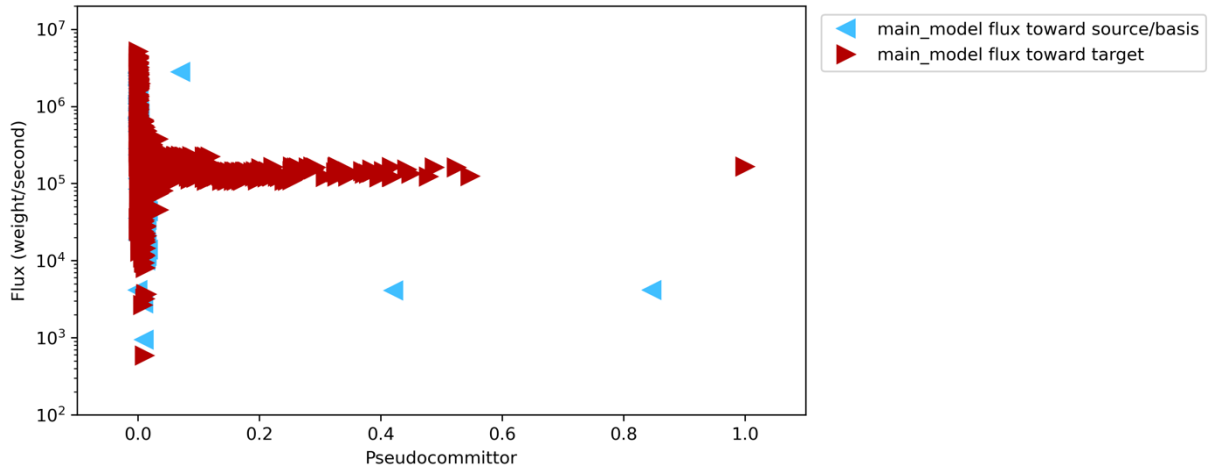
20

Figure S16. Flux vs. pseudocommitor graph for **WT** BdpA's final haMSM. A property of a committor is that its flux profile should be constant.
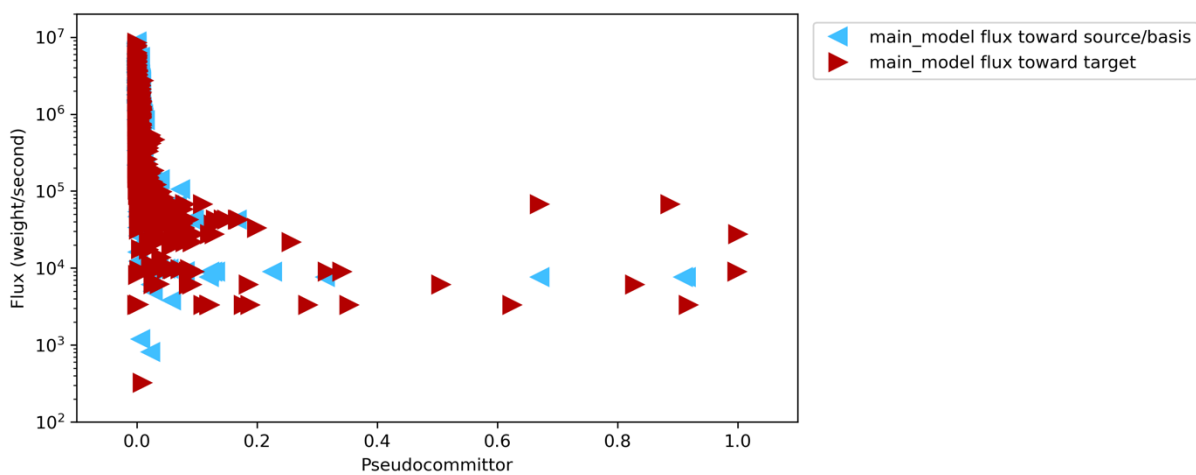
Figure S17. Flux vs. pseudocommitor graph for **β³ -H2** BdpA's final haMSM. A property of a committor is that its flux profile should be constant.
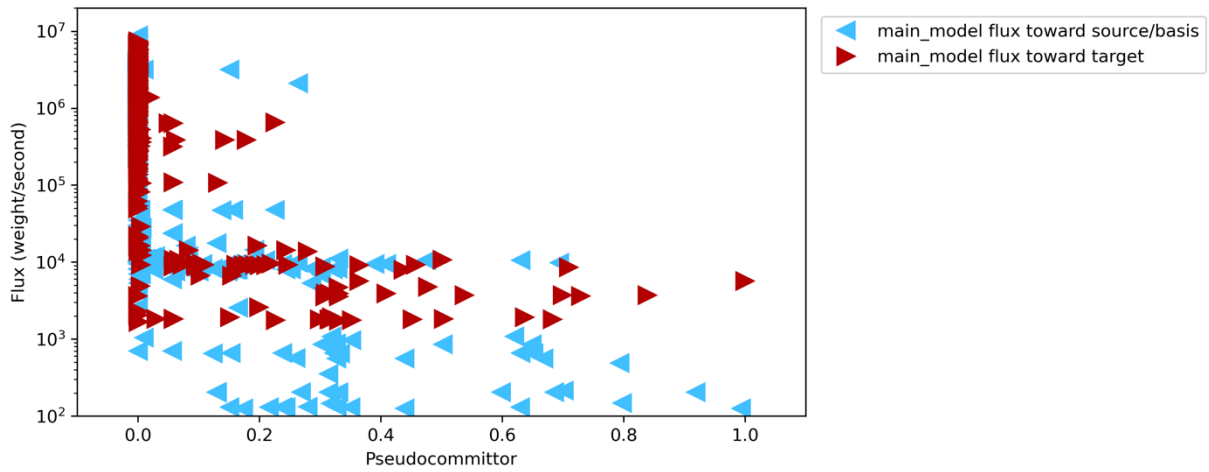
Figure S18. Flux vs. pseudocommitor graph for **β³-H3** BdpA's final haMSM. A property of a committor is that its flux profile should be constant.

Figure S19. Flux vs. pseudocommitor graph for **β³/βᶜʸᶜ-H2** BdpA's final haMSM. A property of a committor is that its flux profile should be constant.

References

1   J. R. Santhouse, J. M. G. Leung, L. T. Chong and W. S. Horne, *Chem. Sci.*, 2022, **13**, 11798–11806.

2   W. Lee, M. Tonelli and J. L. Markley, *Bioinformatics*, 2015, **31**, 1325-1327.

3   W. Rieping, M. Habeck, B. Bardiaux, A. Bernard, T. E. Malliavin and M. Nilges, *Bioinformatics*, 2007, **23**, 381-382.

4   A. T. Brunger, *Nat. Protoc.*, 2007, **2**, 2728.

5   F. Mareuil, T. E. Malliavin, M. Nilges and B. Bardiaux, *J. Biomol. NMR*, 2015, **62**, 425-438.

6   Y. Kim and J. H. Prestegard, *J. Magn. Reson.*, 1989, **84**, 9-13.

7   P. Arora, T. G. Oas and J. K. Myers, *Protein Sci.*, 2004, **13**, 847-853.

8   J. K. Myers and T. G. Oas, *Nat. Struct. Biol.*, 2001, **8**, 552-558.

9   R. E. Burton, R. S. Busby and T. G. Oas, *J. Biomol. NMR*, 1998, **11**, 355-360.

10 A. T. Bogetti, H. E. Piston, J. M. G. Leung, C. C. Cabalteja, D. T. Yang, A. J. DeGrave, K. T. Debiec, D. S. Cerutti, D. A. Case, W. S. Horne and L. T. Chong, *J Chem Phys*, 2020, **153**, 064101.

11 A. T. Bogetti, H. E. Piston, J. M. G. Leung, C. C. Cabalteja, D. T. Yang, A. J. DeGrave, K. T. Debiec, D. S. Cerutti, D. A. Case, W. S. Horne and L. T. Chong, *J. Chem. Phys.*, 2020, **153**, 064101.

12 K. Takemura and A. Kitao, *J. Phys. Chem. B*, 2012, **116**, 6279-6287.

13 I. S. Joung and T. E. Cheatham, 3rd, *J. Phys. Chem. B*, 2008, **112**, 9020-9041.

14 J. D. Russo, S. Zhang, J. M. G. Leung, A. T. Bogetti, J. P. Thompson, A. J. DeGrave, P. A. Torrillo, A. J. Pratt, K. F. Wong, J. Xia, J. Copperman, J. L. Adelman, M. C. Zwier, D. N. LeBard, D. M. Zuckerman and L. T. Chong, *J. Chem. Theory Comput.*, 2022, **18**, 638-649.

15 D. A. Case, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, I. Cheatham, T. E., V. M. D. Cruzeiro, T. A. Darden, R. E. Duke, G. Giambasu, M. K. Gilson, H. Gohlke, A. W. Goetz, R. Harris, S. Izadi, S. A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K. M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F. Pan, S. Pantano, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R. C. Walker, J. Wang, L. Wilson, R. M. Wolf, X. Wu, Y. Xiong, Y. Xue, D. M. York and P. A. Kollman, University of California, San Francisco, 2020.

16 B. W. Zhang, D. Jasnow and D. M. Zuckerman, *J. Chem. Phys.*, 2010, **132**, 054107.

17 J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327-341.

18 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577-8593.

19 J. D. Russo, J. Copperman and D. M. Zuckerman, *arXiv*, June 16, 2020, 2006.09451, Ver 1. https://arxiv.org/abs/2006.09451 (accessed March 30, 2023).

20 D. Aristoff and D. M. Zuckerman, *Multiscale Model. Simul.*, 2020, **18**, 646-673.

21 T. L. Hill, *Free Energy Transduction and Biochemical Cycle Kinetics*, Dover Publications, Mineola, NY, 1989.

22 B. Mostofian and D. M. Zuckerman, *J. Chem. Theory Comput.*, 2019, **15**, 3499-3509.

23 H. Wu and F. Noé, *J. Nonlinear Sci.*, 2020, **30**, 23-66.

24 C. J. Williams, J. J. Headd, N. W. Moriarty, M. G. Prisant, L. L. Videau, L. N. Deis, V. Verma, D. A. Keedy, B. J. Hintze, V. B. Chen, S. Jain, S. M. Lewis, W. B. Arendall III, J.

Snoeyink, P. D. Adams, S. C. Lovell, J. S. Richardson and D. C. Richardson, *Protein Sci.*, 2018, **27**, 293-315.