Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a β-sheet miniprotein

SUPPLEMENTARY MATERIAL

Stefanie Muff and Amedeo Caflisch

Department of Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland

(Dated: April 10, 2007)

Keywords: Complex network, non-native interactions, transition state, multiple folding pathways

I. THE CSN OF W10V

Fig. S1 shows the analogous of Fig. 1 (main text) for W10V. The probability that the weight-ratio of a denatured-state node visited by either peptide is larger than two, five or ten (i.e., that the node of one of the two peptides has been visited more often than two, five or ten times than in the other peptide) is 67%, 33% and 24%, respectively. These differences in weight are statistically significant because the same analysis on two subsets of Beta3s trajectories yields Beta3s(0-10 μs)/Beta3s(10-20 μs) weight-ratios larger than two, five or ten with a probability of only 39%, 18% and 9%, respectively. Analogously, the W10V(0-10 μs)/W10V(10-20 μs) weight-ratios larger than two, five or ten have a probability of only 39%, 10% and 8%, respectively.

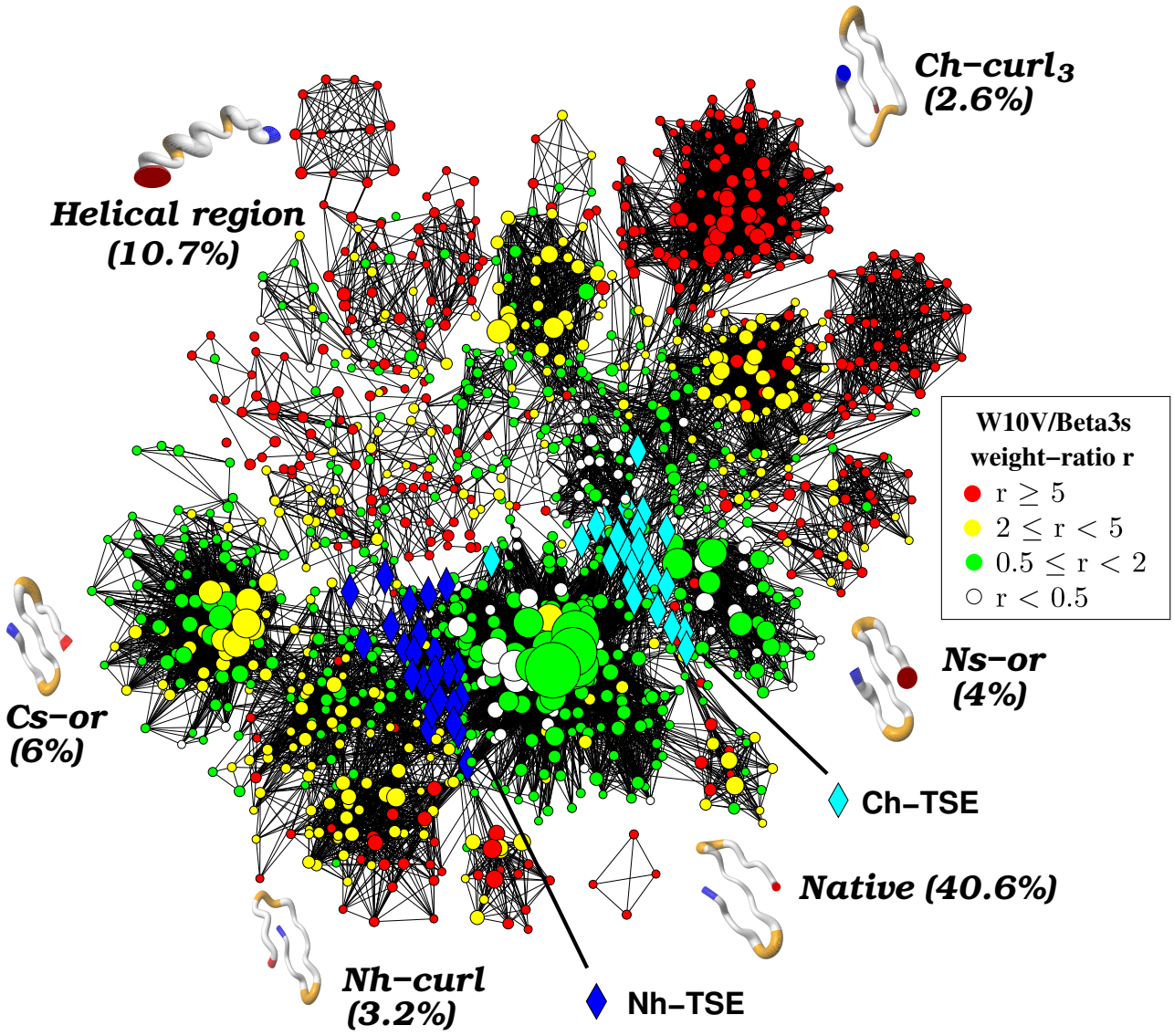

FIG. S1: The CSN of W10V. Each node (i.e., conformation) of the network represents a secondary structure string. The surface of each node is proportional to its statistical weight and only the 1192 nodes with at least 40 snapshots in W10V are shown to avoid overcrowding. Nodes are colored according to the W10V/Beta3s weight-ratio. Conformations in the most populated basins are shown by flexible tubes of variable diameter with N-terminus in blue, C-terminus in red and residues Gly6, Ser7, Gly14 and Ser15, which are at the two turns in the folded structure, in orange. The helical conformation shown on top left is the most populated helical string (`--IIHHHHHHHHHHHHHH-`) in the W10V network. Blue and cyan diamonds emphasize TSE nodes with N-terminal and C-terminal hairpin formed, respectively. This Figure was made using visone (www.visone.de) and MOLMOL[1].

II. KINETIC GROUPING

During the 20 μs simulation time 120 and 105 folding events (i.e., visits to the native node) were observed for Beta3s and the W10V mutant, respectively, thus providing sufficient statistical sampling for the kinetic analysis. Unfolding events were defined as absence from the most populated (i.e., native) node longer than 10 ns.

To perform the kinetic all-against-all grouping, only significantly populated nodes with a statistical weight of at least 4×10^{-5}, i.e., 40 snapshots or more, have been employed (1430 for Beta3s, 1192 for W10V), which does not influence the kinetics since the original trajectory with all snapshots is considered for the p_{commit} calculations. 56% (59%) of the total weight in Beta3s (W10V) lies in nodes above the cutoff. This relatively low values are consistent with the weight distribution of nodes[2] which implies that most of the strings are very rare, occurring only once or twice in the simulation (also known as the *zero-frequency problem*[3,4]), and are not pronounced attractors. In fact, Beta3s and W10V spend 26%, respectively 24% of the time in nodes of weight one or two. Nodes with less than 40 snapshots are assigned to the basins identified by the heavy-node kinetic grouping in a post-processing step. Each "light" node is grouped to the basin to which the p_{commit} $\geq$0.5 criterion is fulfilled. If several candidates are possible, the most populated basin is chosen. The conformation space network (CSN) colored according to the kinetic grouping is illustrated in Fig. S2 and the most populated basins are listed in Table S-I.

A. Most populated strings in the native basin

The native basin includes 7569 and 5829 strings for Beta3s and W10V, respectively. The most populated are listed in Table S-II. Note that there is no correlation between the "geometrical" distance (i.e., number of different bits) and the kinetic distance from the native string. In fact, the strings in Table S-II have a geometrical distance of 4 to 8 and relax to the native string within 0.5 ns, whereas Ns-or (i.e., the string `-EEEESTTEEEEESSEEEE-`) has a geometrical distance of one and relaxes in 138 ns.

		Beta3s				W10V				
		Weight (%)		τ_f (ns)		Weight (%)		τ_f (ns)		
Conformation	Name	Node	Basin	Node	Basin	Node	Basin	Node	Basin	color
-EEEESSEEEEEESSEEEE-	native	5.59	36.42	–	–	8.76	40.63	–	–	green
Larger weight in Beta3s										
-EEEESTTEEEEESSEEEE-	Ns-or	1.17	7.44	138	109	0.82	3.96	92	90	red
---SSGGG---EESSEETT-	Ch-curl$_1$	0.13	3.55	98	90	0	0	–	–	white ovals
---SSGGG-EESSTTTTEE-	Ch-curl$_2$	0.12	2.38	285	257	0	0	–	–	white circles
-----SS--EEEESSEEEE-		0.03	2.20	53	75	0.04	0.87	72	85	orange
-HHHHHHHHHHHS------	Helix$_{1-13}$	0.06	2.06	137	122	0.01	0.50	124	151	white squares
--EESSSEEEEEESSEEEE-		0.10	1.94	87	84	0.04	0.63	148	134	cyan
---SSGGG-EEESSSEEEE-		0.09	1.17	200	198	0	0	–	–	white rectangles
---SSSS--EEESTT-EEE-		0.06	0.94	316	263	0	0	–	–	white diamonds
Larger weight in W10V										
-EEEESSEEEESSSEEEE-	Cs-or	0.26	3.56	63	70	0.65	6.02	69	75	olive
-EEEESSEEEE---TT--B-	Nh-curl	0.04	0.49	59	58	0.13	3.23	69	69	blue
----STT---EEESSEEEE-		0.12	0.81	139	113	0.29	2.70	108	121	violet
--BSS-SSSEEE-STTEEE-	Ch-curl$_3$	0	0	–	–	0.12	2.58	104	105	white diamonds
--SSSSS--EEEESSEEEE-		0.03	0.81	103	97	0.09	2.09	111	100	yellow
-BSSSS---EEEESSEEEE-		0	0	–	–	0.02	0.28	61	53	white circles

TABLE S-I: Results of the kinetic grouping. Statistical weight of the native basin and the most populated free energy basins in the denatured state as identified by the kinetic grouping. The mean folding time (τ_f) to the native node are average values for snapshots in a node or basin. Conformations with names are shown by flexible tubes of variable diameter in Fig. S1 for W10V and in the main text for Beta3s: Ns-or, N-terminal strand out of register; Cs-or, C-terminal strand out of register; Nh-curl, curl-like conformation with structured N-terminal hairpin; Ch-curl, curl-like conformation with structured C-terminal hairpin. The colors indicated in the last column are those used in Fig. S2.

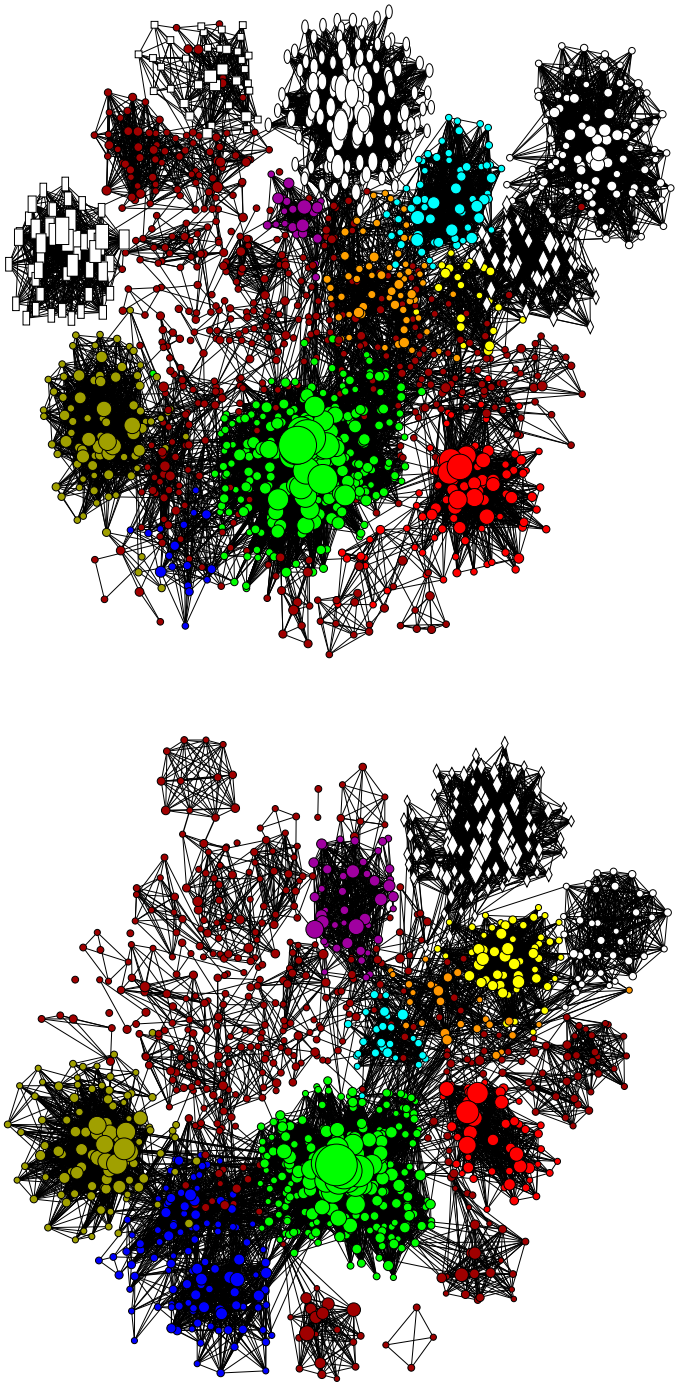

FIG. S2: The CSN of Beta3s (top) and W10V (bottom), colored according to the basins identified by the kinetic grouping. The coloring scheme is chosen such that basins with the corresponding most populated node have the same color in both peptides (last column in Table S-I). White is chosen if no significantly populated region exists in the conformational space of the other peptide. Nodes belonging to less populated basins and entropic regions are in brown.

Beta3s	%	τ_f (ns)	W10V	%	τ_f (ns)
-EEEESSEEEEEESSEEEE-	5.6	0	-EEEESSEEEEEESSEEEE-	8.8	0
-EEE-STTEEEEESSEEEE-	4.7	0.3	-EEE-STTEEEEESSEEEE-	5.2	0.3
-EEEESSEEEEE-STTEEE-	2.8	0.4	-EEEESSEEEEE-STTEEE-	4.9	0.2
-EEE-STTEEEE-STTEEE-	2.1	0.4	-EEE-STTEEEE-STTEEE-	2.9	0.3
-EEEESSEEEEEESSEEE--	1.9	0.4	-EEEESSEEEEESSTTEEE-	1.7	0.2
-EEESSTTEEEEESSEEEE-	1.4	0.3	-EEESSTTEEEEESSEEEE-	1.5	0.3
-EEE-TTTEEEEESSEEEE-	1.3	0.5	-EEE-STTEEEESSTTEEE-	1.0	0.3
-EEEESSEEEEE-STTEE--	1.0	0.4	-EEESSTTEEEE-STTEEE-	0.9	0.2
-EEEESSEEEESSTTEEE-	0.9	0.5	-EEEESSEEEEEESSEEE--	0.9	0.5
-EEE-STTEEEEESSEEEE-	0.8	0.4	-EEE-TTTEEEEESSEEEE-	0.9	0.4

TABLE S-II: The ten most populated strings in the native basin with their relative populations and mean folding times (τ_f) for both peptides. Deviations from the native string are colored in red. Node-p_{fold} values[5] are larger than 0.98 for all nodes.

III. KINETIC PARTITIONING OF THE DENATURED STATE

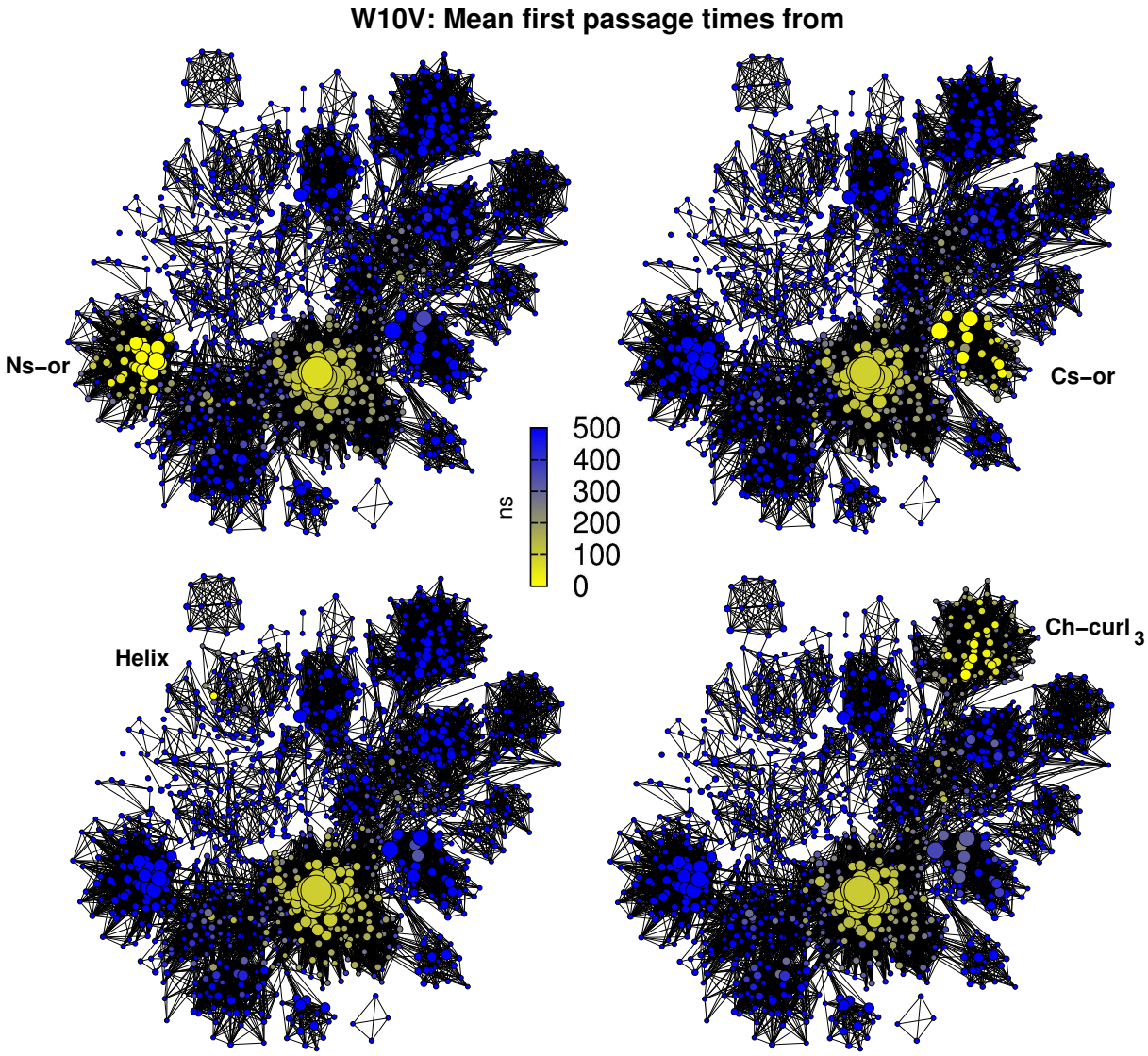

FIG. S3: The denatured state is kinetically partitioned. Mean first passage times from the most populated node of individual free energy basins in the unfolded state to all other nodes of the CSN of W10V are shown. Nodes within the basin of the starting node are visited relatively fast (yellow), indicating rapid intrabasin transitions and supporting the kinetic grouping analysis (Fig. S2 bottom). Equilibration between different unfolded basins (blue) is slower than reaching the folded state (olive) which shows that the denatured state is kinetically partitioned, i.e., no fast equilibration takes place between basins in the denatured state. In other words, the native state is a hub[2].

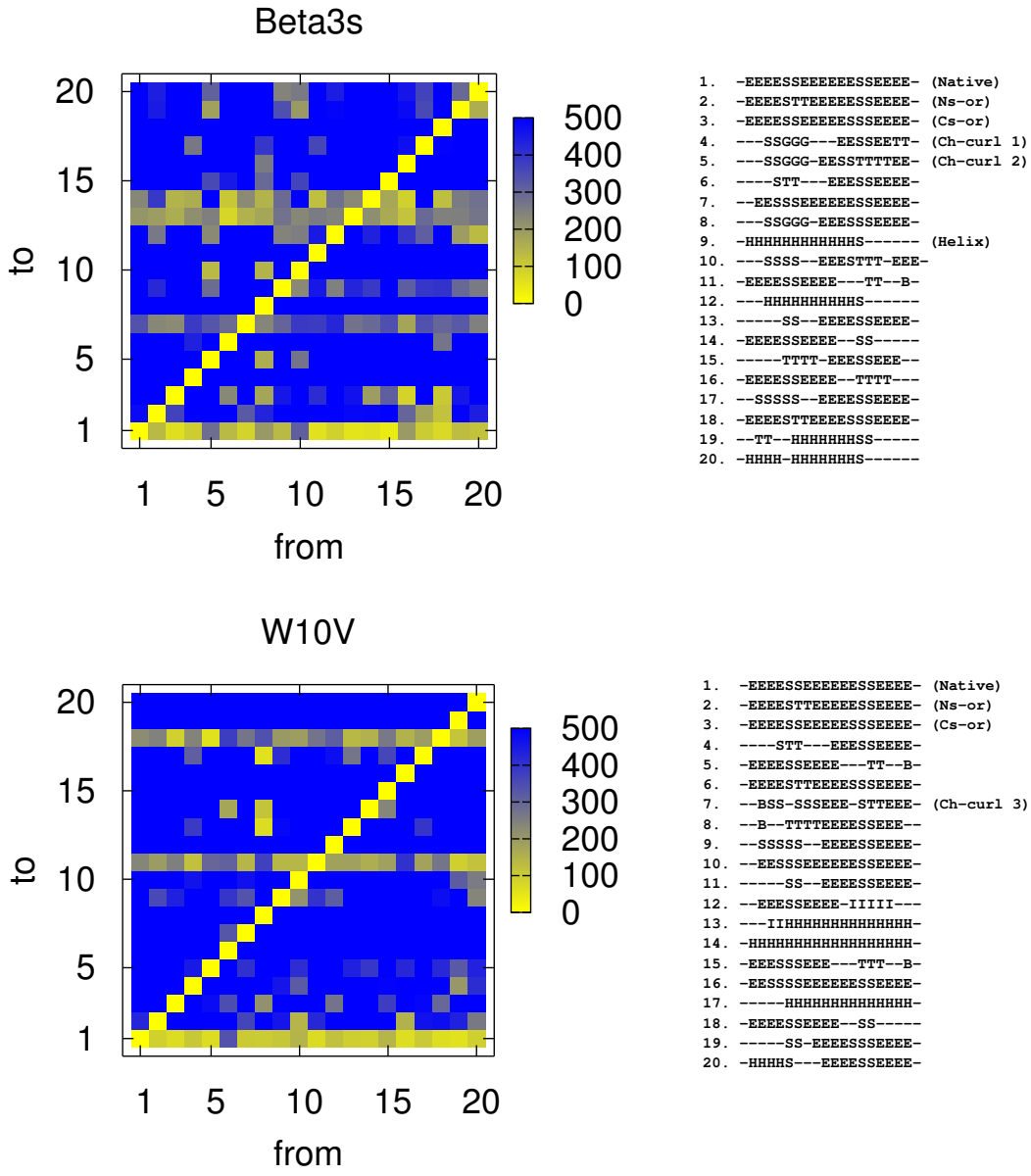

FIG. S4: Mean first passage times between the most populated nodes of the 20 most populated free energy basins as identified by kinetic grouping for Beta3s (top) and W10V (bottom). Transitions to the folded state (basin 1) are generally faster than transitions to other basins, indicating that the denatured state is partitioned by high barriers. Execptions are basin 13 and 14 (11 and 18) of Beta3s (W10V) that are involved in may folding events and turn out to lie on-pathway. Note that these basins are the same in both peptides and have either of the two hairpins fully formed, while the other hairpin is unstructured (-) except for the turns (SS at positions 6-7 or 14-15).

IV. NATIVE STRUCTURE IN THE DENATURED STATE

See Fig. S5.

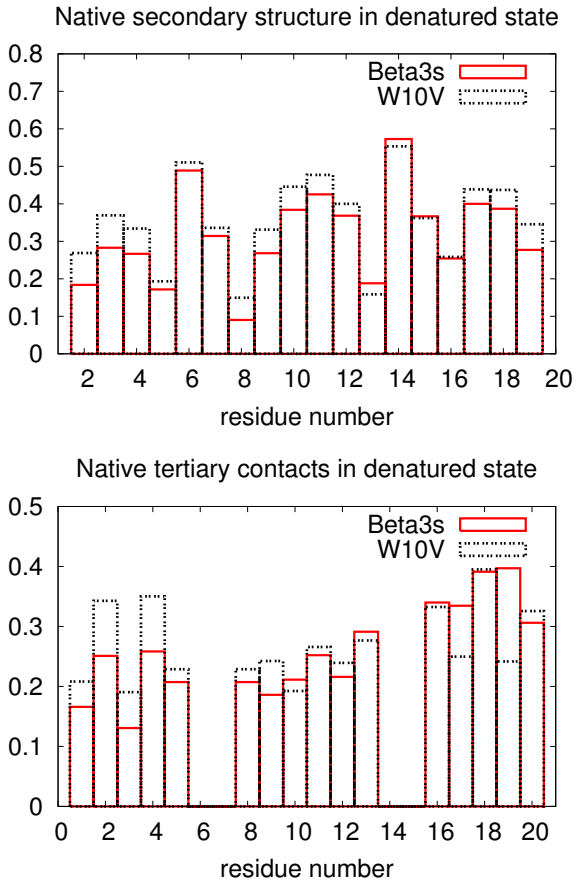

FIG. S5: The N-terminal hairpin (segment 2-11) has 19% higher content of native secondary structure (top) and 20% more native tertiary contacts (bottom) in the denatured state of the W10V mutant than the wild type peptide. Note the different y-axes.

V. DEPENDENCE OF THE KINETIC GROUPING ON τ_{commit}

A. The native basin

The isolation of the native basin is robust with respect to the commitment time which has been chosen as 1.6 ns (main text). Using 5 ns as τ_{commit} increases the population of the native basin only slightly from 36.4% to 38.5% in Beta3s and from 40.6% to 42.1% in W10V.

B. The denatured state

The value of τ_{commit} for the isolation of basins in the denatured state of Beta3s and its mutant has been set to 1 ns uniformly for all basins in order to calculate an all-against-all p_{commit}-matrix. The justification for this choice is that the relaxation times in important enthalpic basins lie within the order of magnitude of 1 ns, but transition from outside are two to three orders of magnitudes slower. The free energy profiles for the bottoms of the most populated enthalpic basins in Fig. S6 indeed show that they have only slightly different characteristic intra-basin relaxation times. The effect upon changing τ_{commit} from 0.5 ns to 5 ns has been investigated (Fig. S7). The analysis is robust for small time variations: no relevant changes in the isolation of basins is noticeable between 0.5 ns and 2 ns, which is the range where the most relevant basins have their maxima in Fig. S6. This means that the most relevant basins in the Beta3s unfolded state can be (at least approximatively) extracted using a constant commitment time. The same holds for W10V. Increasing τ_{commit} further, however, results in an almost trivial splitting of the CSN, where only the helical region is separated from the rest of the denatured state (red and white regions in Fig. S7, bottom right).

As mentioned in the Methods section of the main text, increasing values of τ_{commit} allow one to analyze different levels of ruggedness of the free energy surface. In fact, the Beta3s Cs-or and Nh-curl basins "merge" at τ_{commit}=2 ns (olive region in the bottom left part of Fig. S7), whereas Ns-or and Ch-curl$_{1,2}$ remain separated. This observation is consistent with the similar values of τ_f for Cs-or (70 ns) and Nh-curl (58 ns) and the different τ_f values of Ns-or (109 ns) and Ch-curl$_{1,2}$ (90 ns, 257 ns, Table S-I).

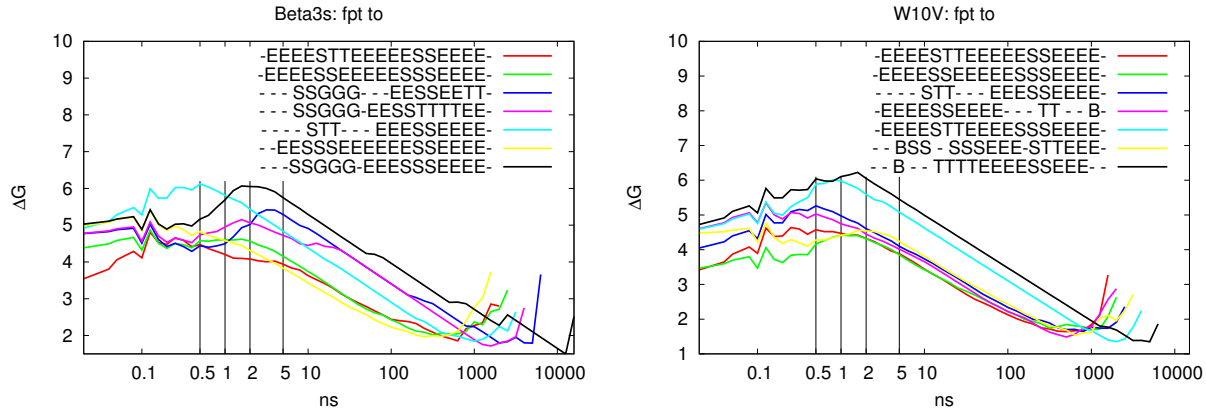

FIG. S6: Distribution of first passage times $P(fpt)$ to the most populated nodes of the largest basins in Beta3s (left) and W10V (right). Values are calculated as $\Delta G = -k_B T \ln(P(fpt))$ and plotted in kcal/mol using logarithmic binning without normalization of the bin size.

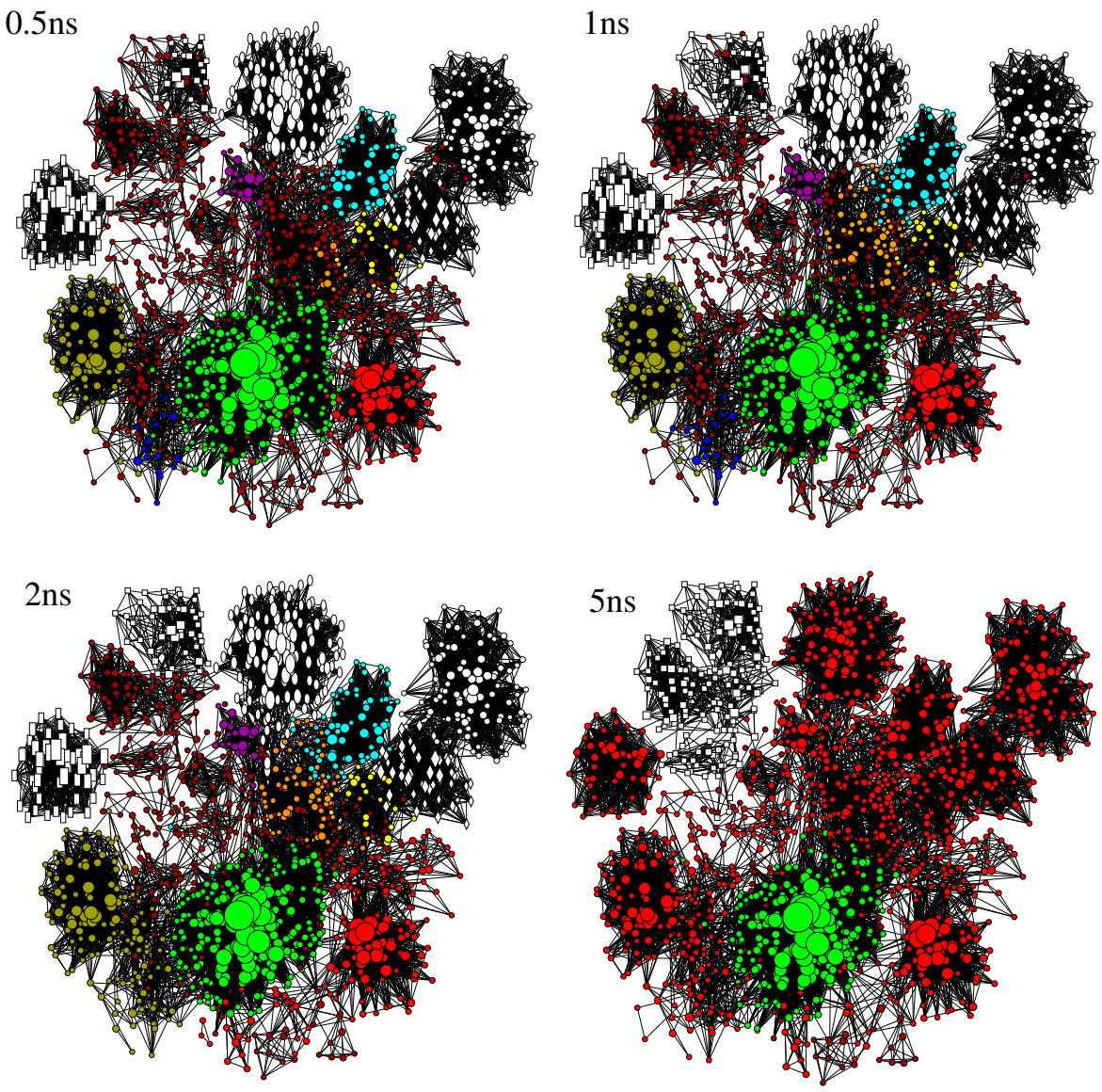

FIG. S7: Robustness of the kinetic grouping in the denatured state of Beta3s: slightly shorter (0.5 ns) and slightly longer (2 ns) τ_{commit} do not change the isolation of important basins compared to the value used in the main text (1 ns, see Table S-I for coloring scheme). The use of considerably longer times (5 ns), however, results in an almost trivial splitting of the CSN into one large region (red), where only the helical basin (white) is separated from the rest.

VI. THE TRANSITION STATE ENSEMBLE (TSE)

A. TSE nodes

See Table S-III.

Beta3s	node-p_{fold}	p_{fold} [a]	$\sigma_{p_{fold}}$ [b]	τ_f (ns)	W10V	node-p_{fold}	τ_f (ns)
-----GGG--EEESSEEE--	0.52	0.50	0.11	45	-----STT-EEEESSEEEE-	0.53	47
-EEEESSEEEES-GGG--B-	0.43	0.51	0.35	45	-EEEESSEEEE--TTT----	0.49	36
-----STT--EE-STTEE--	0.57	0.75	0.09	38	-EEEESSEEEE--HHHHHH-	0.46	33
----SSTT-EEEESSEEEE-	0.42	0.34	0.11	56	-EEEESSEEEE-SSTT----	0.53	38
-EEEESSEEEE-SSSS-EE-	0.59	0.54	0.23	48	-EEEESSEEEE--SSS----	0.45	37
-EEE-SSGGGEEESSEEEE-	0.41	0.34	0.09	9	--EEESSEEEESSSSEEE--	0.42	37
----BSSB--EESSTTEE--	0.51	0.68	0.16	51	-EEE-STTEEE--SSS-EE-	0.44	46
----STTT-EEEESSEEEE-	0.50	0.48	0.15	61	----SSTT-EEEESSEEEE-	0.44	72
-EEE-STTEEE--SSS----	0.41	0.26	0.28	42	-EEEESSEEEE-SSTT-TT-	0.51	18
-EE-SSSS-EEEESSEEE--	0.53	0.65	0.31	23	-----GGG-EEEESSEEEE-	0.47	61

TABLE S-III: The ten most populated TSE strings isolated by node-p_{fold} analysis. Green represents native secondary structure. Beta3s is more native in the C-terminal hairpin, while W10V shows native N-terminal hairpin predominance. Node-p_{fold}, p_{fold} and mean folding times (τ_f) are given. [a]p_{fold} was obtained by "shooting" 20 times from 10 individual snapshots to validate the above TSE conformations of Beta3s, and [b]$\sigma_{p_{fold}}$ is the standard deviation over the 10 snapshots. Note that τ_f's are in the order of half of the average folding time which is consistent with 50% unfolding events for trajectories passing through TSE nodes.

B. TSE robustness upon τ_{commit}

See Fig. S8.

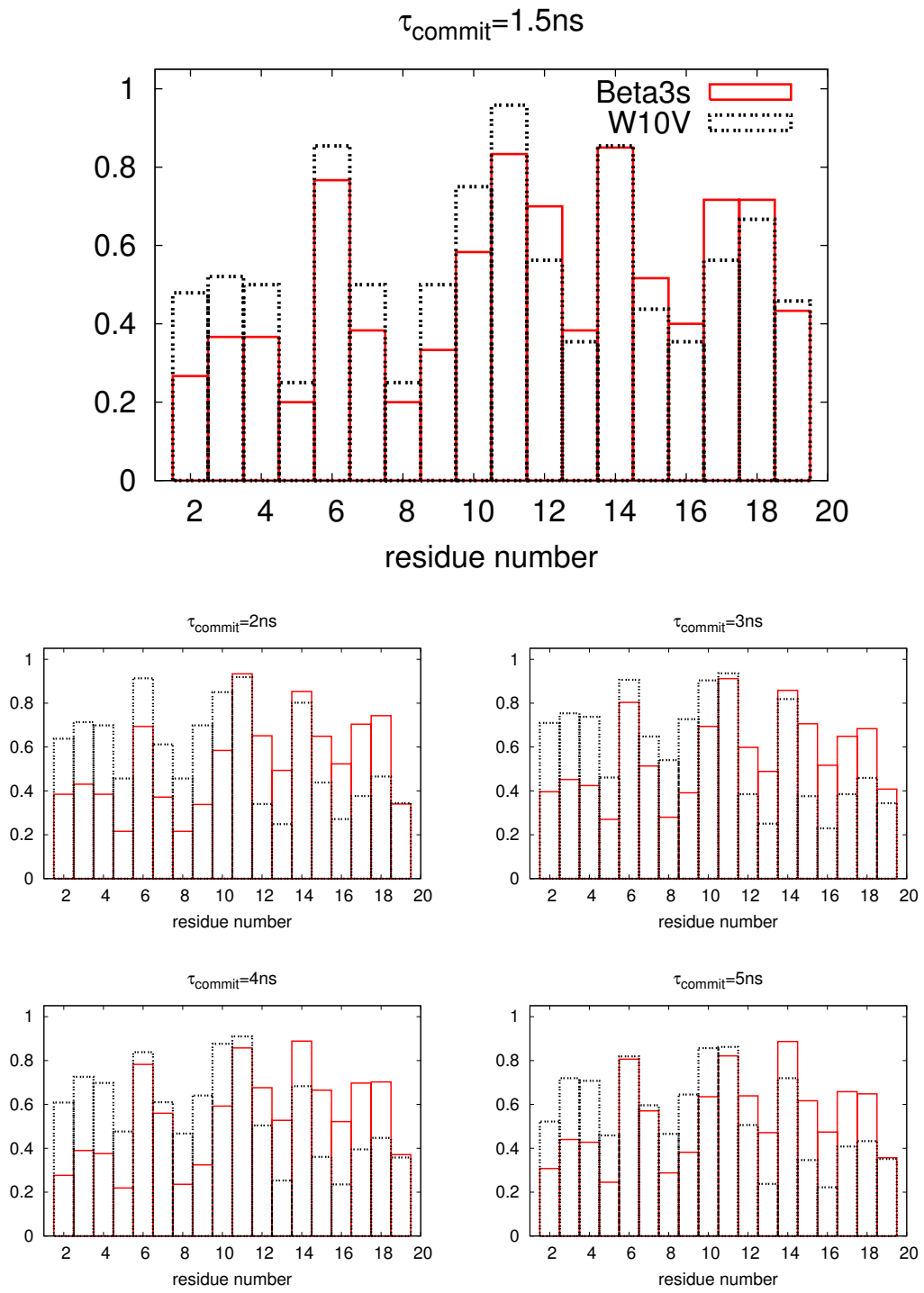

FIG. S8: Robustness of TSE selection upon τ_{commit}: the pathway switch remains evident for all choices between 0.5 ns and 5 ns. Only nodes with weight ≥ 20 have been considered.

C. Energetics of the TSE

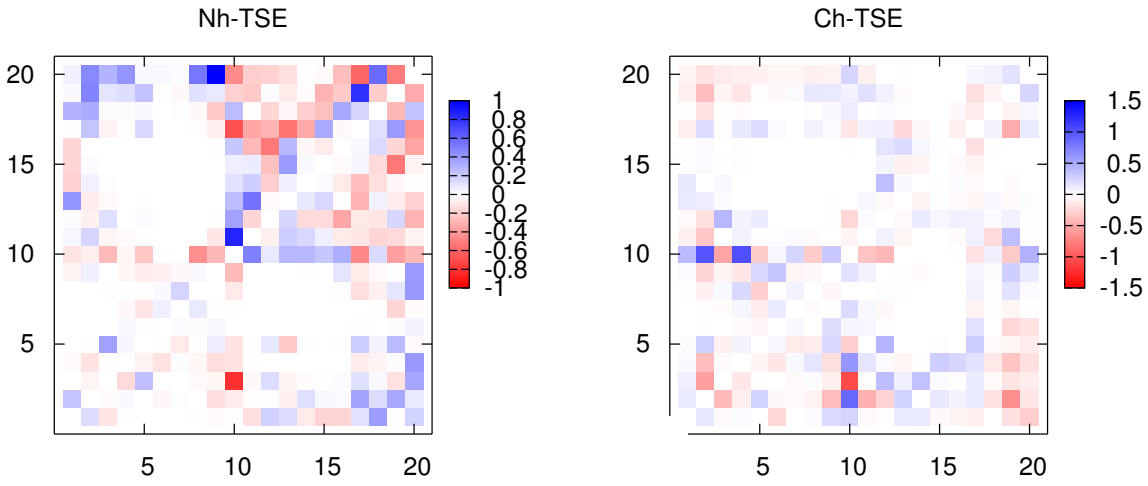

FIG. S9: The pairwise residue interaction energy differences (kcal/mol) between W10V and Beta3s for the Nh-TSE (left, N-hairpin predominant) and the Ch-TSE (right, C-hairpin predominant). The pairwise energy values in the native state are used as reference and subtracted from all values in the matrix. A red square indicates that the corresponding pair of residues has a more favorable interaction energy in the TSE of W10V than Beta3s. The upper and lower triangular matrices show the total and van der Waals energy, respectively, and their similarity indicates that most of the enthalpic effects originate from the difference in van der Waals energy.

D. Sampling of TSE nodes

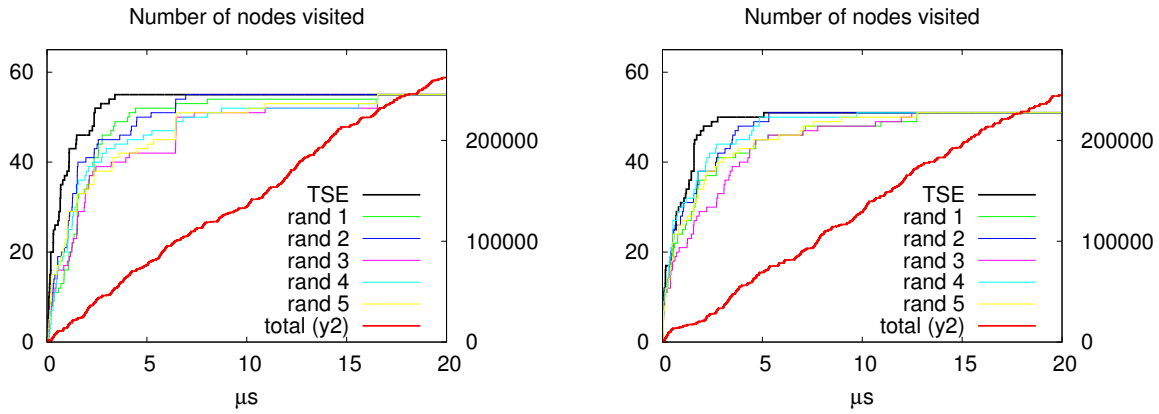

FIG. S10: The number of nodes visited depending on simulation time. The TSE of Beta3s (left) and W10V (right) are accessed completely after 3, respectively 5 μs. Random samples with the same size and weight distribution (weight between 20 and 110) as the two TSEs need in average more than 10 μs to visit all nodes at least once. This observation indicates that TSE nodes are involved in a large number of barrier-crossing events which implies that a node-p_{fold} value of 0.5 does not originate from only very few folding and unfolding events. The total number of nodes sampled during the simulation is shown in red (right axis in both plots). The difference in the evolution of the total number of nodes and the random samples comes from the fact that the latter contain only nodes with weight ≥ 20, while most of the contribution to the increase in total nodenumber comes from lower populated nodes.

VII. THE HELICAL ENSEMBLE

Interestingly, the distribution of helical content along the sequence shows that Beta3s is more helical than W10V in the central segment, i.e., residues 7-13 (Fig. S11). This observation is consistent with the fact that the side chain of valine has a destabilizing effect on the helical structure[6] because of the branching at the C_β carbon.

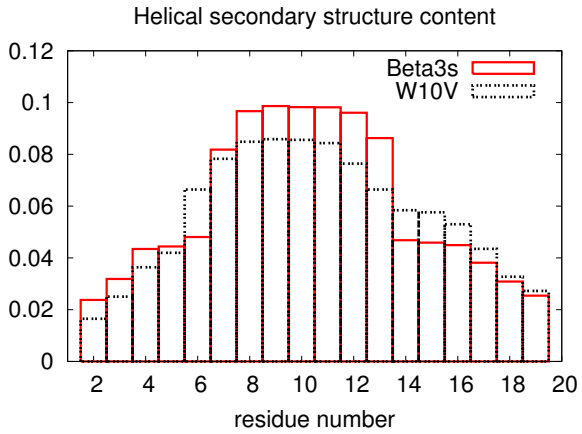

FIG. S11: The distribution of helical content along the sequence shows that Beta3s is more helical than W10V in the central segment.

The helical ensemble of W10V shows a slightly faster decay for the distribution of node weights (i.e., higher entropic character[7]) than the one of Beta3s (Fig. S12), which again reflects the destabilization of helical structure due to the valine side chain.

In order to assure that the distributions of node weights in the helical regions of the considered peptides are not affected by an undersampling problem, two tests have been carried out. Fig. S13 shows the helical weight distribution for the Beta3s simulation with a 50 times higher saving frequency nsavc (left), as well as the distribution of the first and second 10 μs against the full 20 μs simulation (right). In all cases the slope does not change significantly, thus a higher saving frequency and more sampling do not change the distribution, providing evidence that the entropic character of the helical region (i.e., the pronounced decay of the node-weight distribution) does not suffer from undersampling.

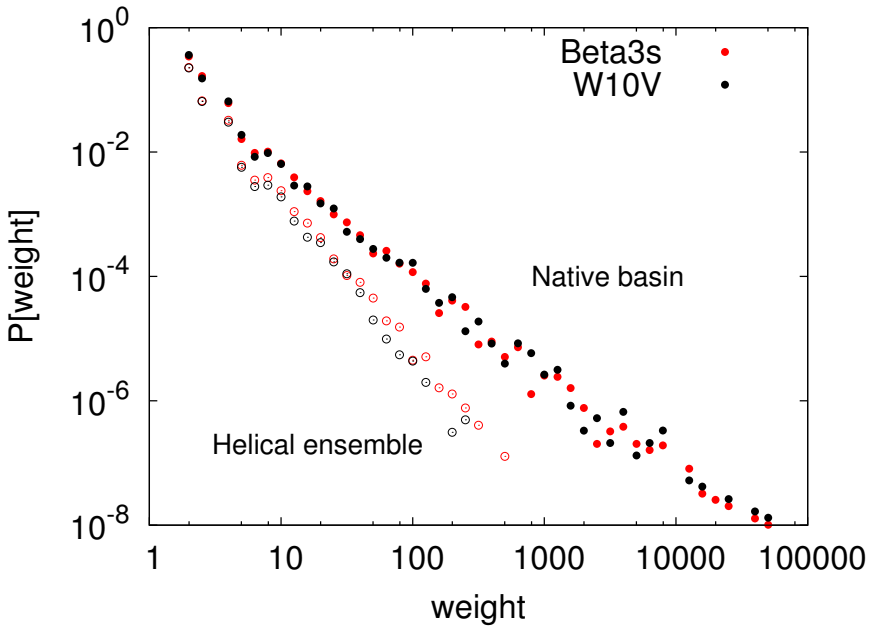

FIG. S12: Distribution of node weights. Logarithmic binning is used to reduce noise. The distributions of the enthalpic free energy basins Ns-or, Cs-or, Nh-curl, and Ch-curl show a very similar decay as the native basin and are not shown to avoid overcrowding.

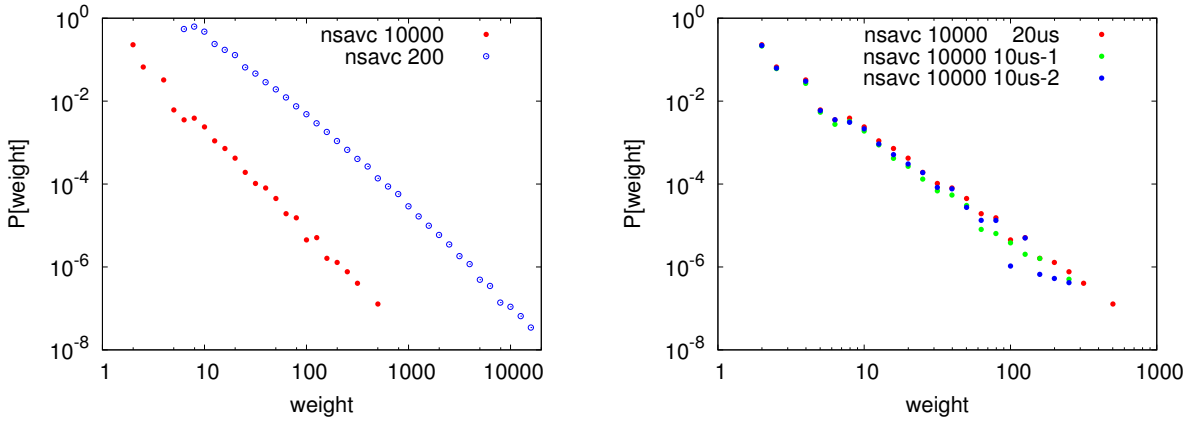

FIG. S13: Weight distribution of nodes in the helical basin. (left) A 50 times higher saving frequency (nsavc), as well as the splitting of the 20 μs simulation into two 10 μs-parts (right) do not change sensibly the slope of the distribution, indication that no undersampling problem exists in this case.

VIII. KINETIC GROUPING OF THE ALANINE DIPEPTIDE

Kinetic grouping is explained here using the alanine dipeptide which is a simple system yet containing the key features of a polypeptide chain. A total of 5×10^7 snapshots saved along a 1 μs MD trajectory at 300 K is used for this purpose. The main degrees of freedom are the dihedral angles ϕ and ψ. In the continuum solvent approximation used here[8] the projection of the free energy landscape onto ϕ and ψ shows four basins (see Fig. S14): C_{7eq}, α_R, C_{ax} and α_L. The most natural discretization of the phase space splits the (ϕ, ψ) space into cells. Using a 50×50 discretization of the Ramachandran map, 1821 nodes and 53995 links are visited during the 1 μs trajectory[7]. The most populated node in the system corresponds to the bottom of the C_{7eq} basin with coordinates ϕ=-86.4 and ψ=136.8.

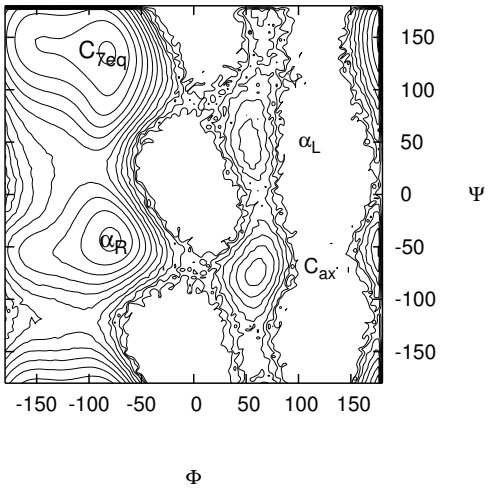

FIG. S14: The ϕ - ψ projection of the 1 μs MD simulation of the alanine dipeptide with the ACE2 implicit model of solvation[8]. Each contour line represents k_BT=0.6 kcal/mol.

Two procedures can be used for kinetic grouping. The first approach is a one-step all-against-all procedure which requires only one parameter determined by plotting fpt-distributions as in Fig. S6 and Fig. 6 of the main text for Beta3s and W10V. The second approach iteratively extracts basins (Fig. S17) by taking into account relaxation times of individual basins (Fig. S16).

A. Kinetic grouping: Simultaneous detection of basins

The simultaneous detection of basins is the procedure that was used for the denatured state of Beta3s and W10V (main text) because it is simpler and requires only a single τ_{commit} value. For all nodes populated above a certain cutoff, an all-against-all commitment probability (p_{commit}) matrix is calculated using a system-typical commitment time τ_{commit}. The weight-cutoff is introduced to make the analysis faster and robust, i.e., to avoid errors caused by nodes lying in high-energy regions. Two nodes are grouped if $p_{commit} \geq 0.5$, using the 80% criterion to reduce false negatives. The results for the alanine dipeptide nodes with $\tilde{w} \geq$300 are shown in Fig. S15. The four basins are identified correctly with $\tau_{commit} = 5$ ps which is the relaxation time to α_R (see Fig. S16). Notably, the nodes in the transition state region between C_{7eq} and α_R are not assigned to either of the two basins (black nodes). Using a commitment time significantly shorter (τ_{commit}=1.5 ps) or longer (τ_{commit}=10 ps) than 5 ps leads to a too detailed or too coarse split of the energy landscape, respectively (Fig. S15 top left and bottom), which shows how the choice of the commitment time is related to the allowed ruggedness of the surface.

B. Kinetic grouping: Iterative detection of basins

The existence of a single τ_{commit} for the simultaneous detection of all basins in more complicated systems is not necessarily guaranteed. A rigorous way for the isolation of basins by the use of kinetic information is to determine relaxation times for each region individually and then perform the kinetic grouping iteratively. The advantage of this approach is that heterogeneous relaxation times within a system are taken into account. Furthermore, there is no need to introduce a weight-cutoff to reduce the number of nodes as required in the simultaneous detection procedure. On the other hand, the analysis becomes more complicated and it is not clear how to automatize the choice of different values of the relaxation time. The procedure works as follows: in a first step, the distribution of the first passage times (fpts) to the most populated node is calculated. As indicated in Fig. S16 A, a value of τ_{commit}=10 ps is chosen to isolate the full C_{7eq} attractor region using $p_{commit} \geq$0.5 with the 80%-criterion (red region in Fig. S17). In a second step, the procedure is repeated for the most populated node that has not been grouped to the first basin. This node has coordinates

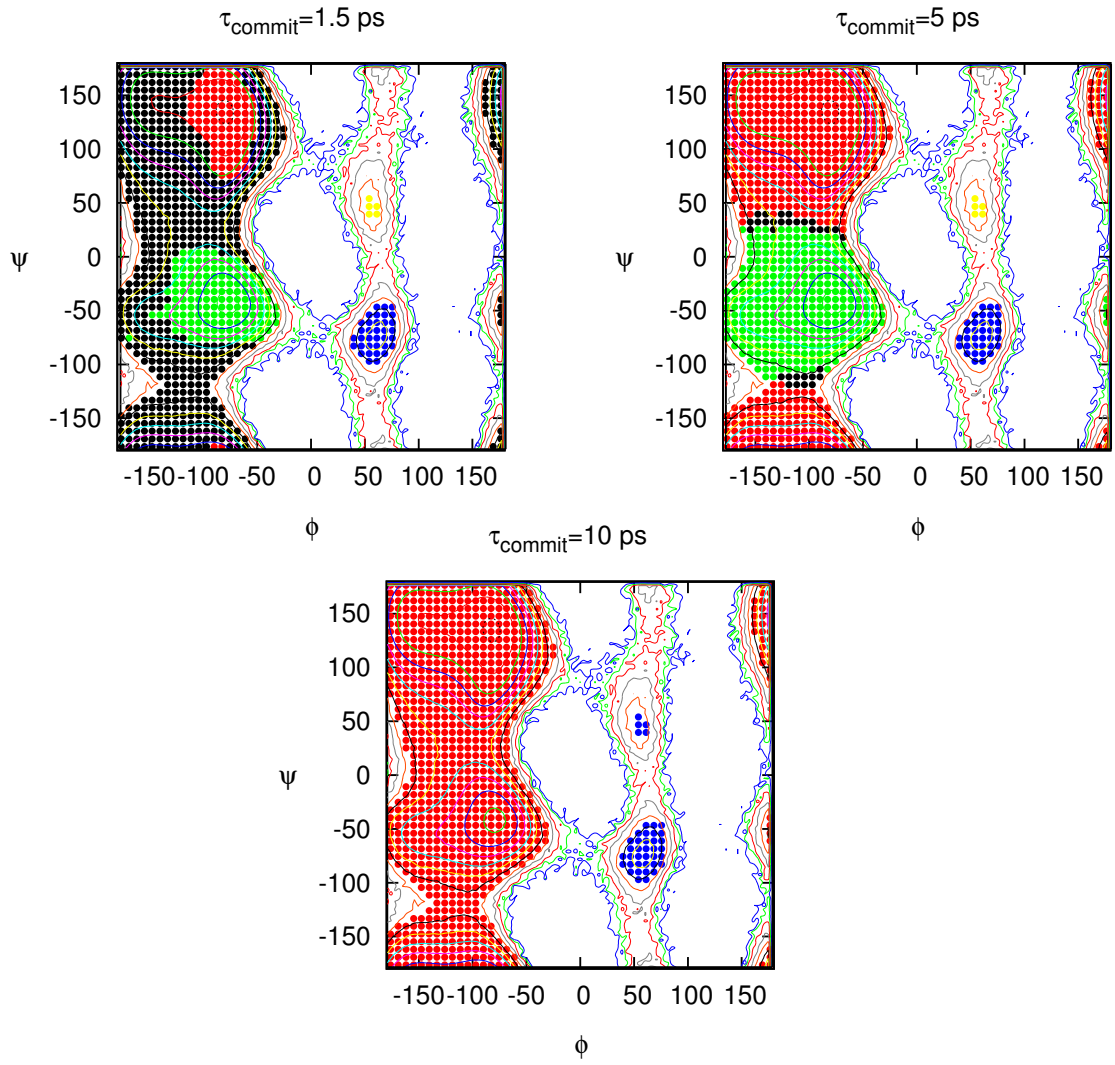

FIG. S15: Results of the simultaneous detection approach to isolate the basins in the alanine dipeptide using a single τ_{commit} value and only nodes with $\tilde{w}$ $\geq$300. Red, green, blue and yellow mark the region containing the bottom of the C_{7eq}, α_R, C_{ax} and α_L basin, respectively. Black dots are used for the remaining nodes which are split into several small groups for τ_{commit}=1.5 ps and 5 ps. The partition into four basins obtained using τ_{commit}=5 ps is very similar to the result of the iterative kinetic grouping (Fig. S17). On the other hand, using τ_{commit}=10 ps, the C_{7eq} and α_R basins as well as the C_{ax} and α_L basins are merged (red and blue regions).

ϕ=-79.2 and ψ=-43.2 and represents the bottom of the α_R region. The commitment time to identify the corresponding attractor region is 5 ps (Fig. S16 B) and a set of nodes is isolated (green region in Fig. S17) that has a marginal overlap with the C_{7eq} region. This intersection

contains putative transition state nodes. The procedure can be continued iteratively with τ_{commit}=1.5 ps for C_{ax} and α_L (Fig. S16 C and D), which leads to the splitting as indicated in Fig. S17. Interestingly, the regions isolated by the kinetic grouping analysis using either procedure (simultaneous or iterative detection) are comparable to those found by Markovian clustering[7] with granularity parameter $p = 1.2$.

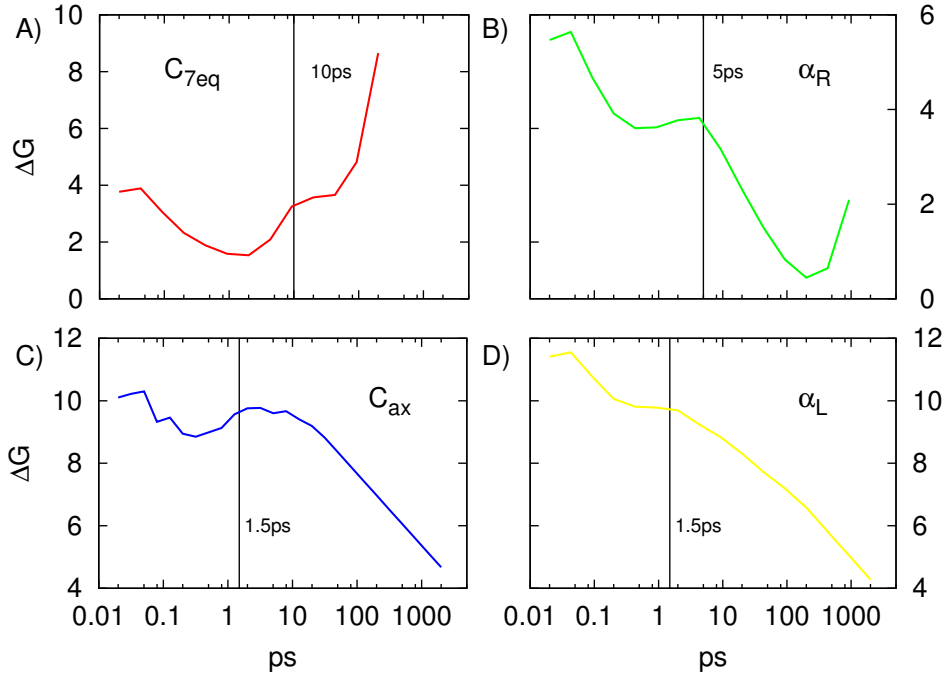

FIG. S16: Free energy profile as a function of temporal distance from the most populated node of (A) C_{7eq}, (B) α_R, (C) C_{ax} and (D) α_L. ΔG is calculated as $-k_BT \cdot ln(P(fpt))$ and plotted in kcal/mol. The fpt can be considered as a geometrically unbiased reaction coordinate. This projection is very useful to determine the transition between intra- and inter-basin relaxation, which is emphasized by the logarithmic binning without normalization of the bin size.

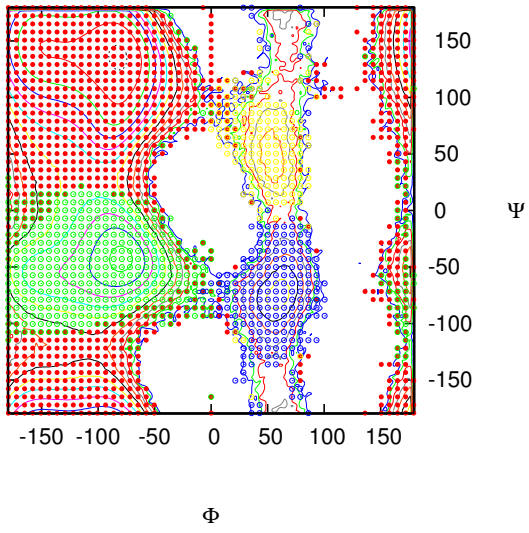

FIG. S17: Isolation of the free energy basins with the iterative kinetic grouping analysis. Coloring is consistent with Fig. S16. Black nodes are due to overlapping attractor regions of basins and occur close to transition state regions. Each contour line represents k_BT=0.6 kcal/mol.

[1] R. Koradi, M. Billeter, K. Wüthrich, MOLMOL: a program for display and analysis of macromolecular structures, J. Mol. Graphics Modell. 14 (1) (1996) 51–55.

[2] F. Rao, A. Caflisch, The protein folding network, J. Mol. Biol. 342 (2004) 299–306.

[3] I. Witten, T. Bell, The zero-frequency problem: estimating the probabilities of novel events in adaptive text compression, IEEE Trans. Inf. Theroy 37 (1991) 1085–1094.

[4] Y. Sawada, S. Honda, Structural diversity of protein segments follows a power-law distribution, Biophys. J. 91 (2006) 1213–1223.

[5] F. Rao, G. Settanni, E. Guarnera, A. Caflisch, Estimation of protein folding probability from equilibrium simulations, J. Chem. Phys. 122 (2005) 184901.

[6] M. H. Hecht, B. O. Zweifel, H. A. Scheraga, Macromolecules 11 (1978) 545–551.

[7] D. Gfeller, P. De Los Rios, A. Caflisch, F. Rao, Complex network analysis of free-energy landscapes, Proc. Natl. Acad. Sci. USA. 104 (2007) 1817–1822.

[8] M. Schaefer, C. Bartels, F. Leclerc, M. Karplus, Effective atom volumes for implicit solvent models: comparison between Voronoi volumes and minimum fluctuation volumes, J. Comput. Chem. 22 (15) (2001) 1857–1879.