Supporting Information for

Water-mediated hydrogen bonds and local side chain interactions in the cooperative collapse and expansion of PNIPAM oligomers

Wanlin Chen,[†] Martin Gruebele,[‡] Martina Havenith,[†] Kurt J. Hebel,[&] Carla Scaletti[&]

[†]Physical Chemistry II, Faculty of Chemistry and Biochemistry, Ruhr University Bochum, Bochum, Germany.

[‡]Department of Chemistry, Department of Physics, and Center for Biophysics and Quantitative Biology, University of Illinois Urbana-Champaign, Illinois, USA.

[&]Symbolic Sound Corporation, Champaign, Illinois, USA.

Contents

Videos S1 to S10

Figures S1 to S14

Tables S1 to S6

Supporting text

Links to SI Videos that are not embedded in the main text

Videos

Video S1. https://youtu.be/LaUjTy7M5k8

The sonification/visualization panel: Comparison of C-P space (PC1 vs PC2) for five, versus three variables. WBs lead DBs and bracket (encompass) them on the sequence. At 975.7 ns WBs (11,22) (8,25) and (3,28) At 976.7 ns WB (15/16, 21/22) encompasses DB(16,19). Trade-off pattern: from 977-1014 ns, WB is replaced by DB (16,19). At 1021 ns, WB (11,22) starts breaking down as system moves back to Intermediate state. At 1053 ns WB (6/7, 27) holds it in compact state.

Video S2. https://youtu.be/QIfFNLvEAEo

Trap pattern: PNIPAM can get trapped in a very long-lived (>275 ns) state of intermediate compactness, which coincides with long-lived DBs (red), SCBs (green), and WBs (blue)

Video S3. https://youtu.be/18_APIR3ewI

An 87.5 ns expansion slowed down by a trap (~ 74 ns). DB, WB, and SCB cooperate to hold it in an intermediate state. Enclosing pattern: DB (21-24) with SCB (22-24), enclosed by WB (21-26).VLR WB pulls it down transiently in frame 20168.

Video S4. https://youtu.be/6kw4vBA-r2Y

'Ball bouncing down stairs' pattern: example of a 14 ns multi-step collapse. Very long range WBs when in collapsed state, joined by long range SCBs. Both of them "bounce" (break and reform a few times). Not much DB activity.

Video S5. https://youtu.be/D6eLKeYuGIo

'Ball bouncing up stairs' pattern: example of a 19.55 ns multi-step expand. Very long range WB (VLR) combined with Enclosing patterns: VLR DB (1-27) in Collapsed, just before transition, "enclosed by" VLR WB (1/2-27) in Collapsed, just before transition. Enclosing pattern of DB + SCB: DB (10-14) during transition, coincides with dip in R_g, "encloses" SCB (10-13) during transition, coincides with dip in R_g.

Video S6. https://youtu.be/F64tq6UiKA8

Direct collapse: example of a 10.45 ns collapse with slight fluctuations. VLR WBs (1-22) form transiently during collapse. WB (12-24) and (3-24) form in collapse.

Video S7. https://youtu.be/QHJTf2nFi3Y

Direct expansion: a 3.25 ns simple expansion from compact to expanded. Medium range WB in collapsed state (9-21).

Video S8. https://youtu.be/GDp-aE_YoPw

See Fig. 6A. Successful transition 1. Successful coil-globule transitions are accompanied by a buildup of several non-local WBs, whereas DBs occur mainly after the onset of the transition to compactness. WBs increasing sequence distance while approaching the compact state. Long range WBs encompass the span of long range

DBs (and WB leads in time): At 615.8 ns WB residues (14, 22) encompasses DB residues (18, 22) at 616 ns. Long range WB anticipates long range SCB, on the same residues: WB (7,30) at 608.7 ns, followed by SCB (7,30) at 609.1 ns.

Video S9. https://youtu.be/BtQ6hS2XAps

See Fig. 6A. Successful transition 2. Pattern of nonlocal WBs leading and encompassing DB (sharing a residue and bridging to one beyond the DB span): WB residues (6, 19/20) at 1077 ns encompass DB residues (6,20) at 1078 ns. Very long sequence distances of several bond types associated with globule state with WBs leading: WBs (1,27) and (3,24) at 1097 ns, followed by DB (1,27) and SCB (1,27) at 1098 ns.

Video S10. https://youtu.be/NT6hfAfvS38

See Fig. S7. Example showing Collapse-Expand-Collapse-Expand sequence, and recurring motifs: “Enclosing or bracketing pattern” WB coincides with (and encloses) DB (or SCB) — shares one residue with DB and spans 1 or 2 residues beyond the DB span; “Very Long-range pattern” Very long range bond or WB (more than 10 residues apart) — corresponds with compact state; “Inflection pattern” Longer range SCBs mark the “inflection points” in the C-P and in R_g (suggesting they may interfere or cause changes in direction of collapse). Sum of Bond distances pattern (SB) long distance bonds or bridges (of all types) associated with collapsed state.

Figures

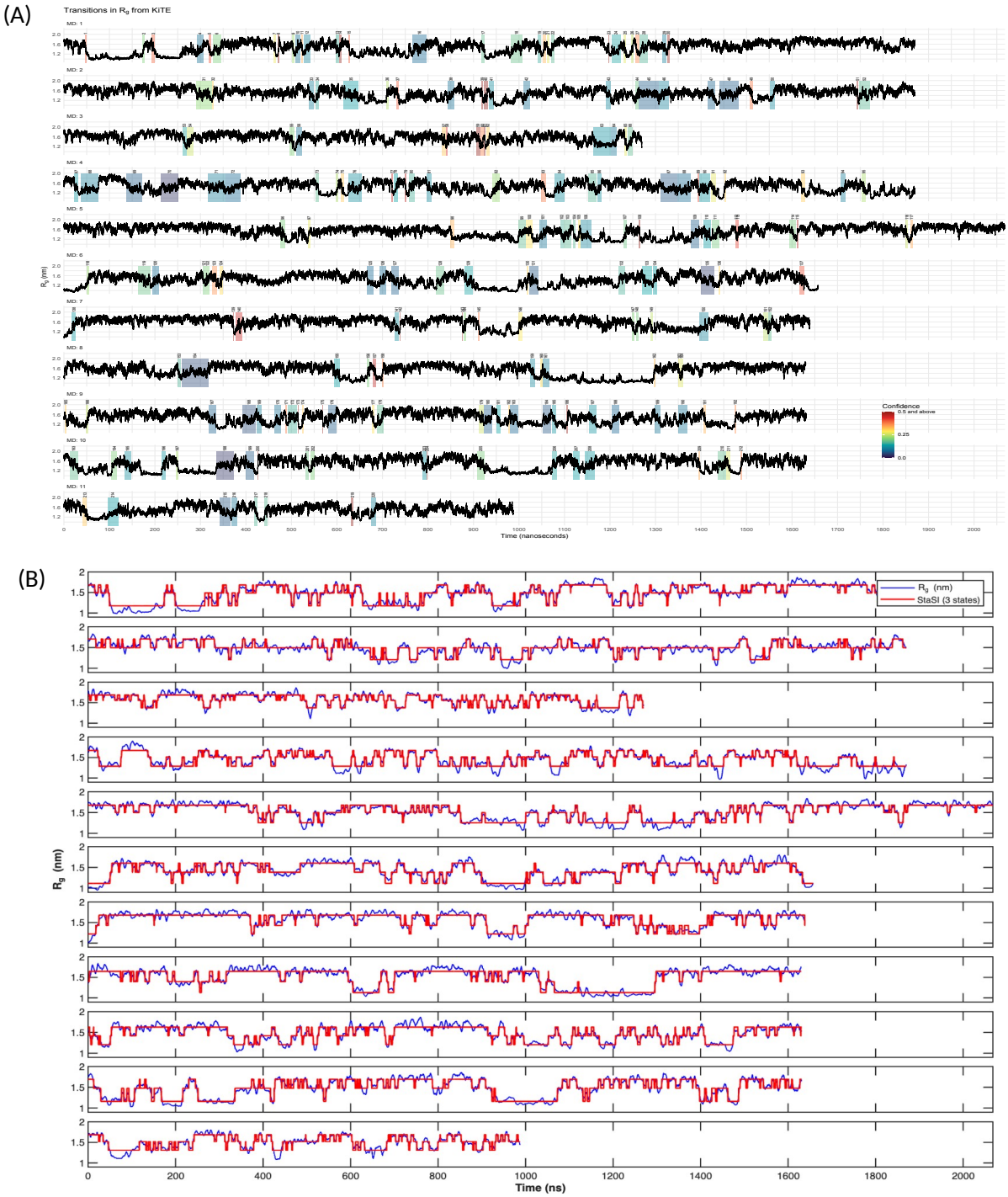


Fig. S1. R_g in nanometers with respect to time. (A) R_g for all 11 trajectories, showing the transitions found using the Kinetic Transition Estimator based on the Gaussian state distribution in Fig. 1C. The color code indicates confidence level for the transition to be a direct transition between globule and coil states (red = more likely; purple = less likely). (B) Analysis using STaSI modeling the data with three distinct states, which assumes Gaussian fluctuations around mean step values for *k* states.

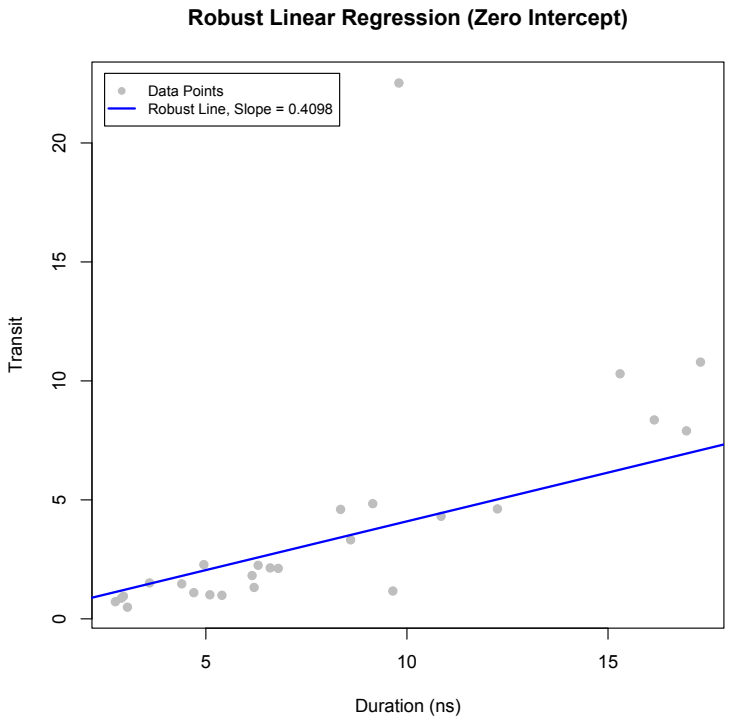


Fig. S2. Zero-intercept robust linear regression used to compute the relation between KiTE transition durations and STaSI transit times. While the STaSI logistic fit indicates $1\text{-}e^{-1}$ completion (~63% completion), KiTE is more conservative by a factor of 1/0.04098 ≈ 2.4, corresponding to about $1\text{-}e^{-2.4}$ or about 90% completion of a transition state passage.

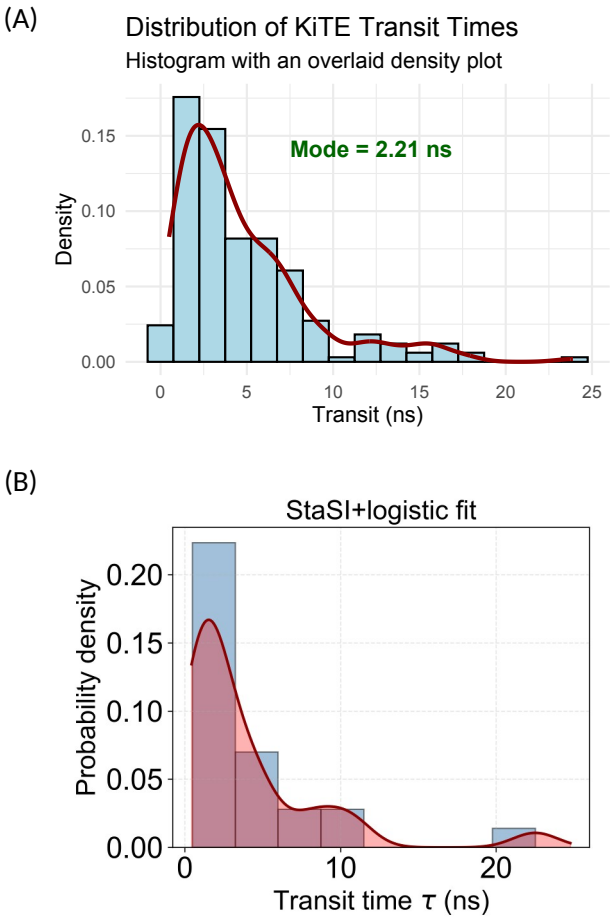
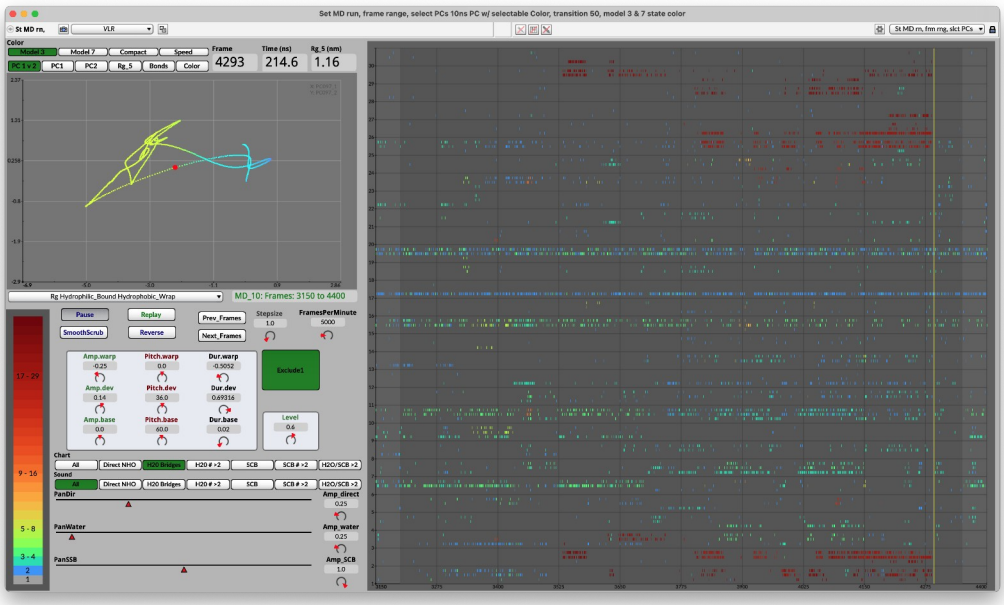
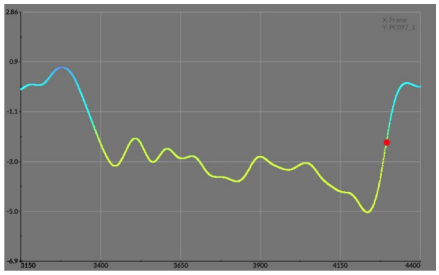


Fig. S3. Histogram of transition state passage times for direct globule-coil and coil-globule transitions using (A) the KiTE estimator (all transitions); and (B) the STaSI analysis for state identification combined with a logistic fit for transit time (26 representative transitions).

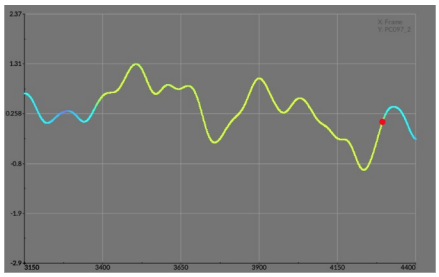
(A)



(B)



(C)



(D)

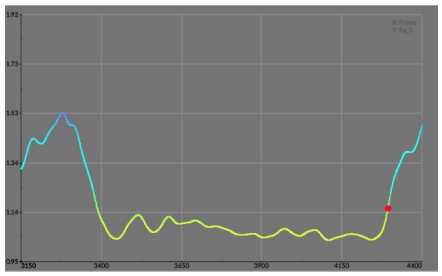


Fig. S4. User interface for trajectory 10, frames 3200 to 4350, and plots comparing PC1, PC2 and R_g for a transition between intermediate (cyan) and compact (yellow) state. (A) C-P state space portrait (left) with WBs in piano roll (right) of visualization/sonification panel. Comparison of (B) PC1, (C) PC2, and (D) R_g .

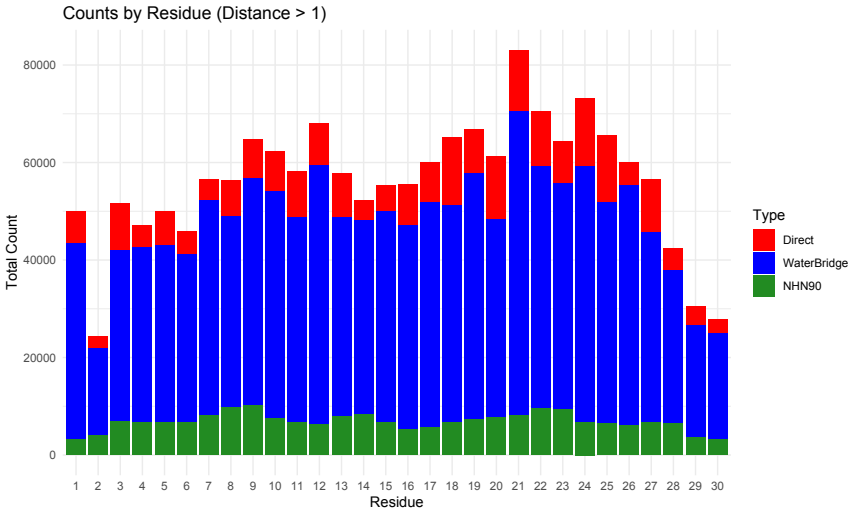


Fig. S5. Total counts of each interaction type — DB (red), WB (blue), SCB (green) — by residue number (position in the sequence) for non-adjacent residues. WBs outnumber all other bond types, and the excess of bonds around residues 10 and 20 suggests that the PNIPAM polymer preferentially folds into a three-part structure.

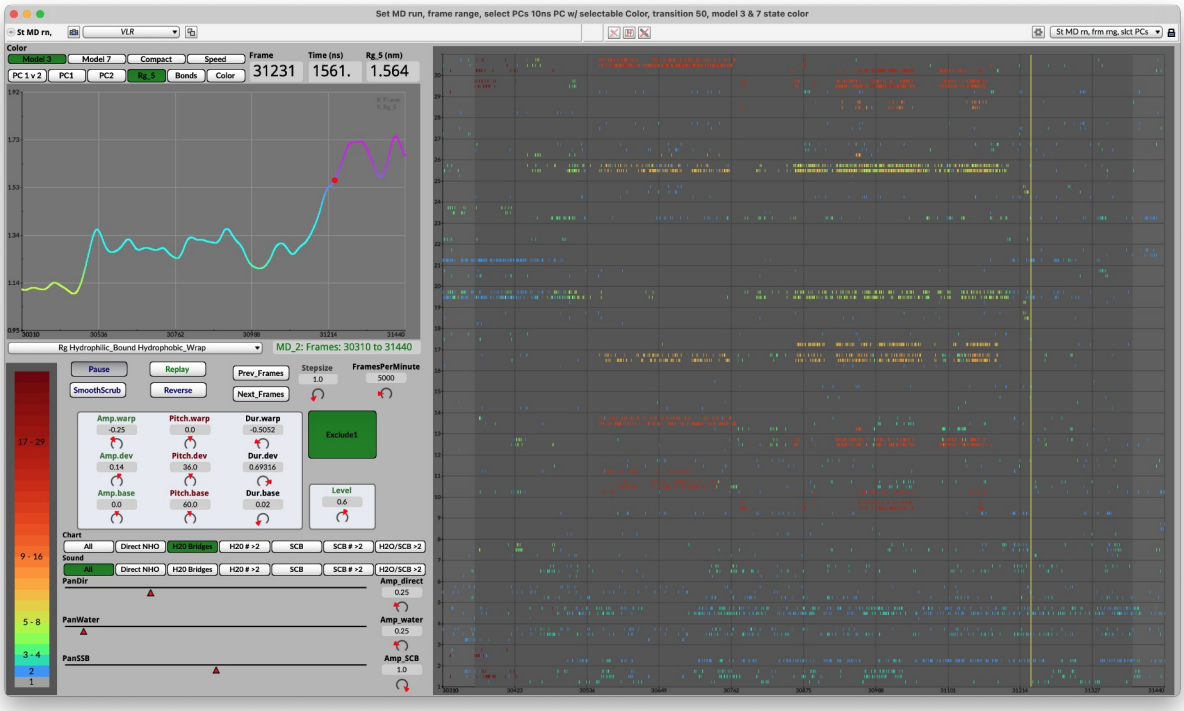


Fig. S6. A two-step transition from globule (yellow) to coil (purple) is observed in the R_g trajectory (Fig. S1A, time interval after transition 49 through transition 50). The transition is characterized by a two-step transition from globule to coil, residing for about 700 ns in an intermediate state held together by water-bridged hydrogen bonds that are highlighted on the 'piano roll' (sequence vs. time) on the right. The color code of the piano roll (shown on the bottom left of the window) shows how non-local the bonding is, from 2 residues away (blue) to 29 away (dark red).

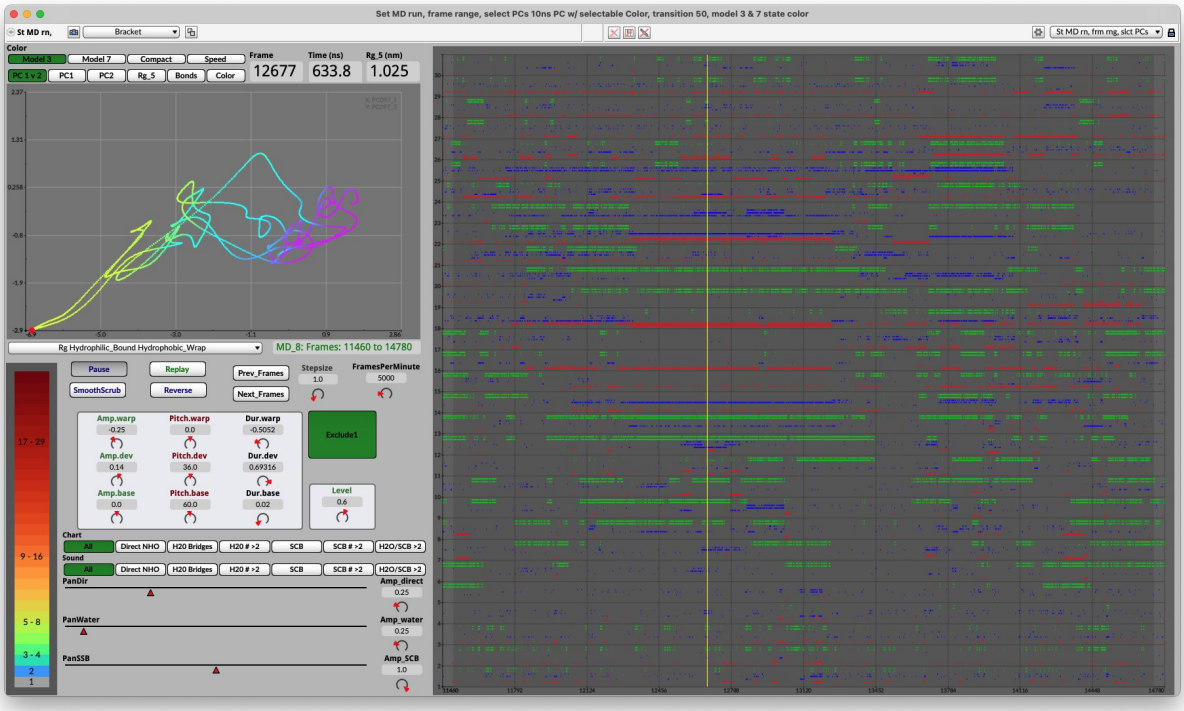


Fig. S7. Example of the 'Bracket' pattern: At the time cursor, two adjacent DBs (in red, from 16-18 and 18-22) are encompassed by a single WB (blue) that spans residues 14-22. In the upper left, the corresponding time in C-P space is indicated by a red dot. In C-P space, the collapsed states are toward the left, expanded states toward the right, and the indicates the probability of being in coil (magenta), Intermediate ensemble (cyan) and globule (yellow). This example shows the H-bonds during a collapse-expand-collapse-expand sequence (see Video S10).

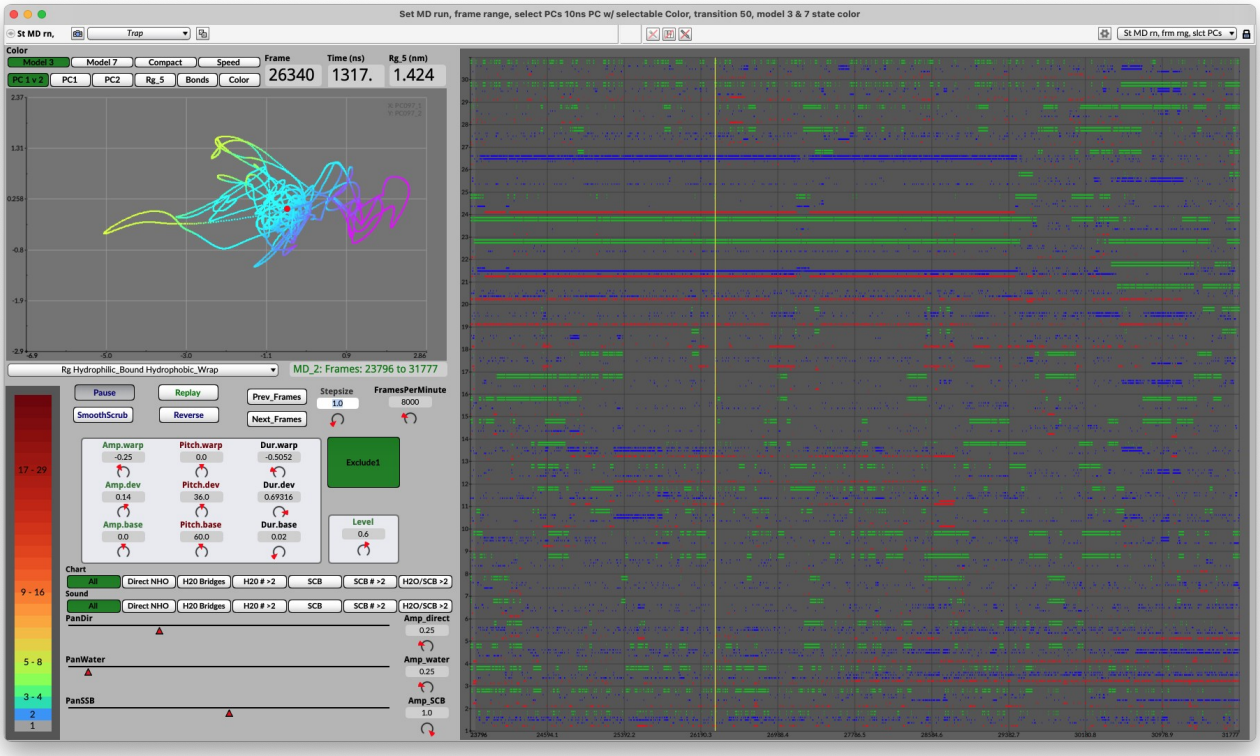


Fig. S8. In this transition from coil to globule and back, long-duration DB, SCB, and WB bonds keep PNIPAM 'circling around' in an intermediate trap in the middle of the C-P space for over 275 ns. When the bonding pattern finally breaks, PNIPAM extends; then long-range WB (residues 1+2 to 29+30) and SCB (1 to 29) form and the oligomer enters a different trap state. The snapshot shows the red dot in the middle of the "trap-like state"; the collapsed state *C* is marked in yellow, the intermediate ensemble of states is blue, and the expanded state is purple. See Video S2.

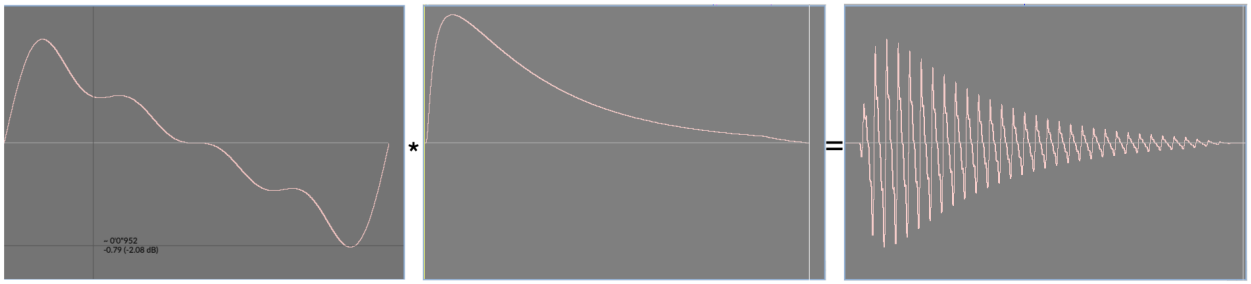


Fig. S9. Each audio event is a synthetic impulse response generated by multiplying the first four harmonics of a sawtooth waveform by a smoothed exponential amplitude envelope.

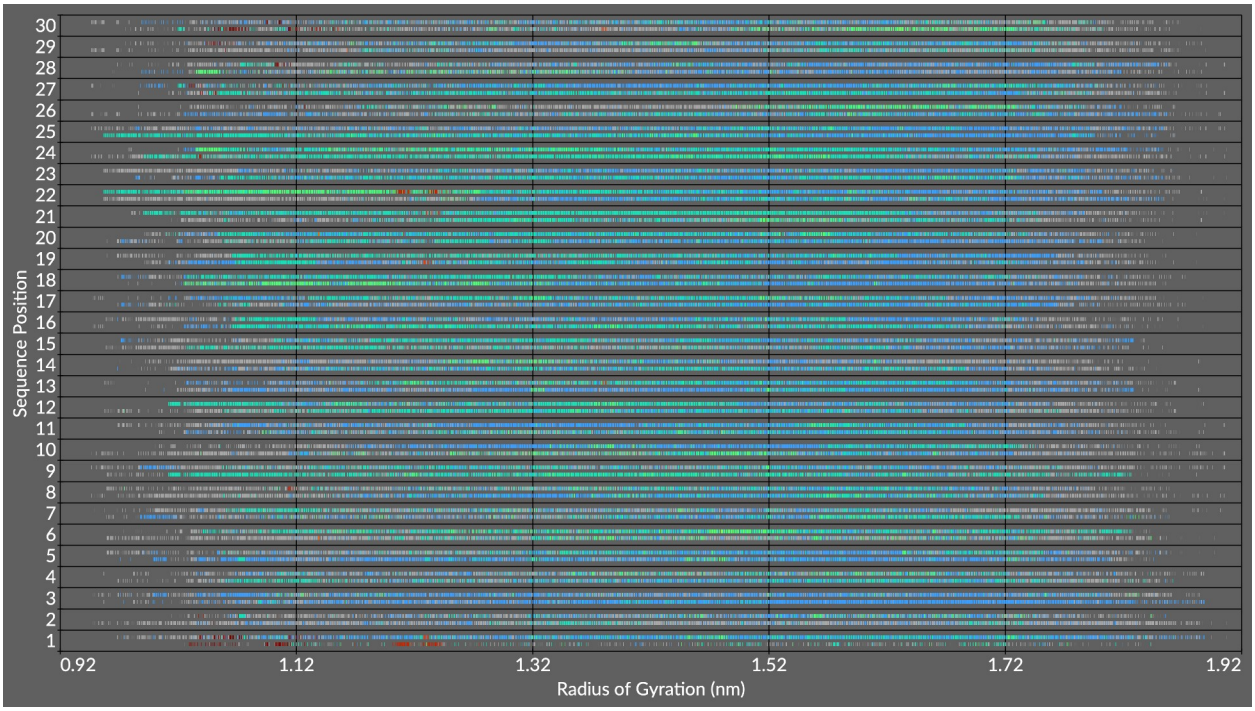


Fig. S10. All direct H-bonds for each residue as a function of R_g and sequence position. Long range direct bonds are scarce and primarily occur only when R_g < 1.2 nm.

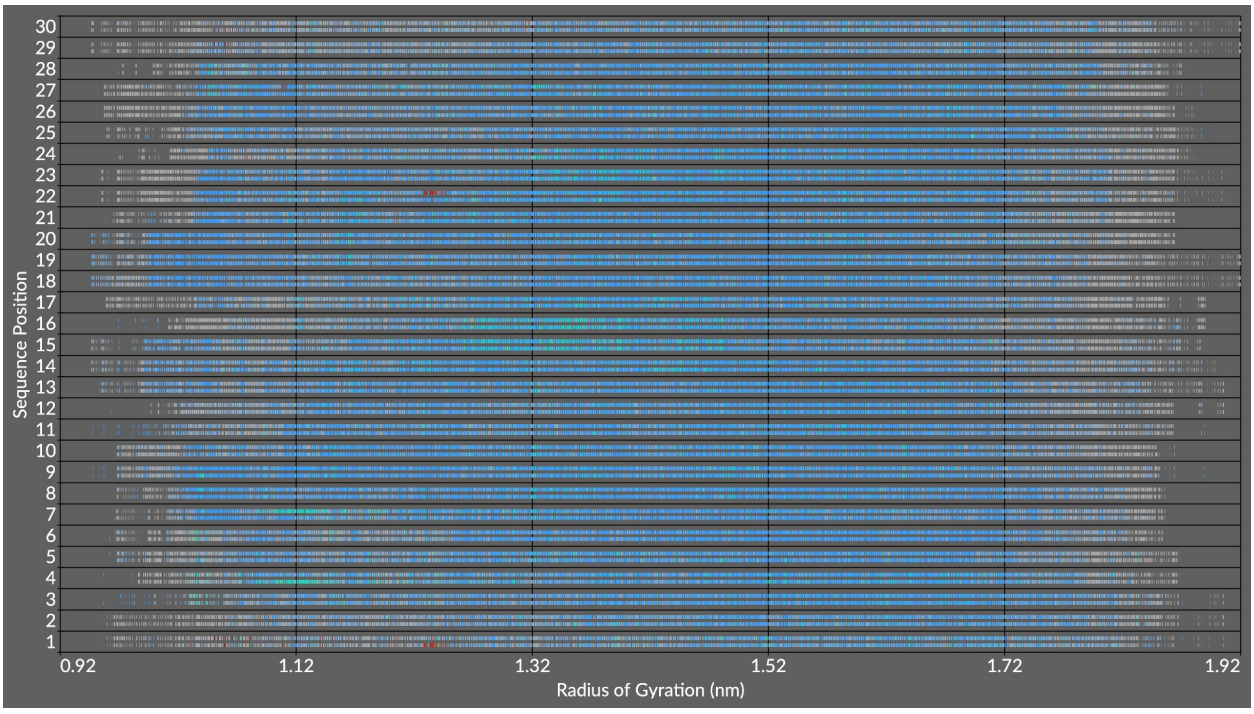


Fig. S11. All SCBs as a function of R_g and sequence position.

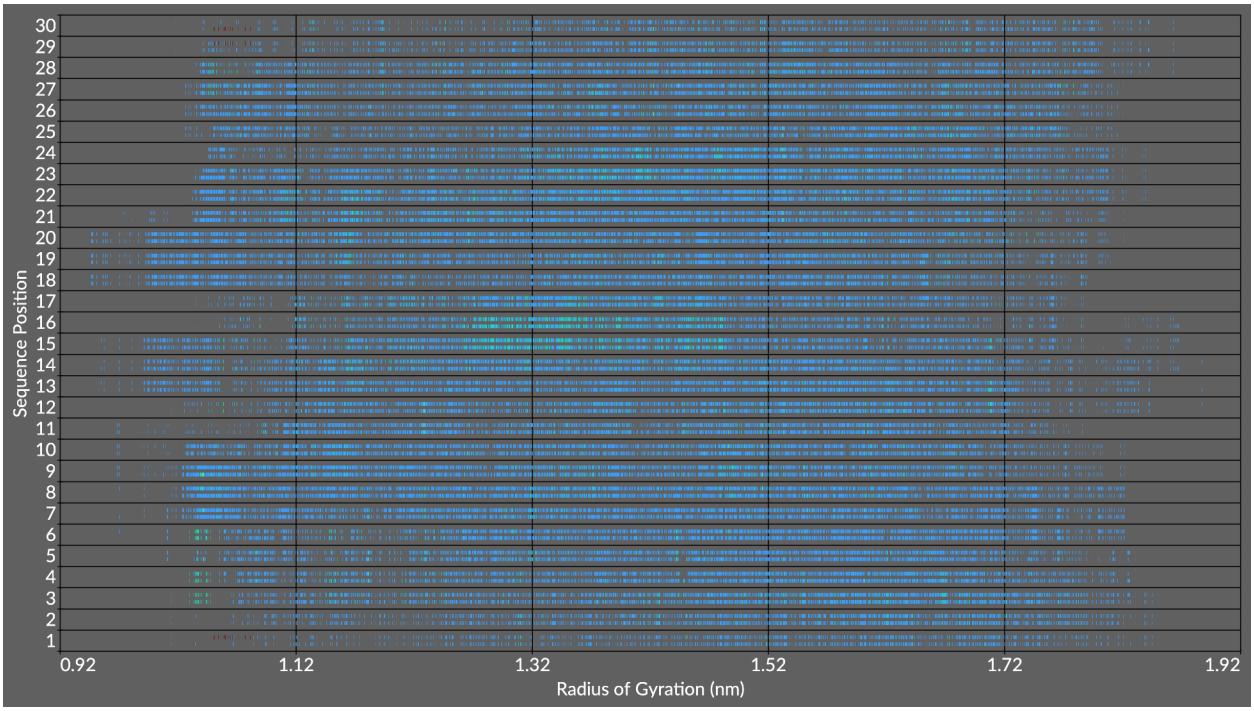


Fig. S12. SCBs between non-adjacent residues as a function of R_g and sequence position. Note decreasing number of bonds for R_g < 1 nm and with the few concentrated toward the center of the polymer, rather than the ends. The green color indicates a slight excess of non-local SCBs in intermediate states (the mid-range of R_g).

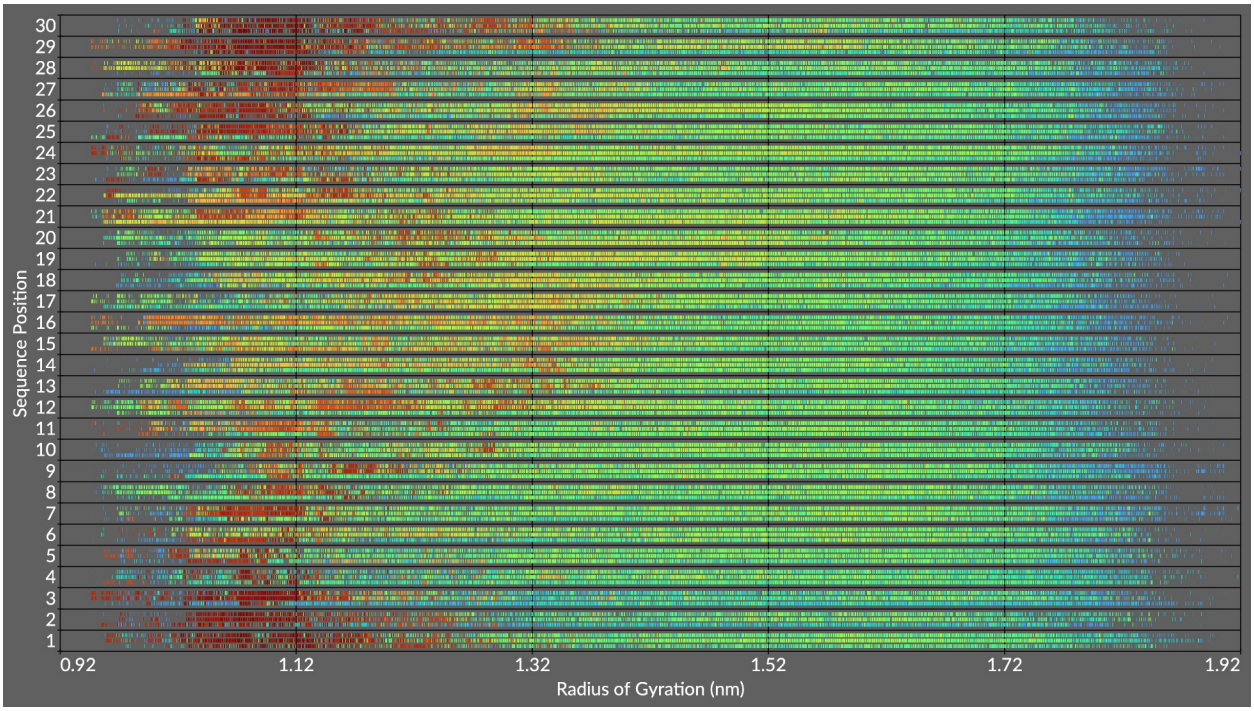


Fig. S13. All WBs for each residue for each residue as a function of R_g and sequence position. Note, as in Fig. 4 of the main text, the concentration of long range bonds linking the two ends of the chain between R_g ~ 1.0 to 1.12 nm. Also, there is a tendency for mid-range water bridges in the center of the chain to move outward toward the ends as R_g decreases from 1.32 to 1.12 nm. For R_g> 1.8 nm there are almost no long-range WBs (only blue color code on the right).

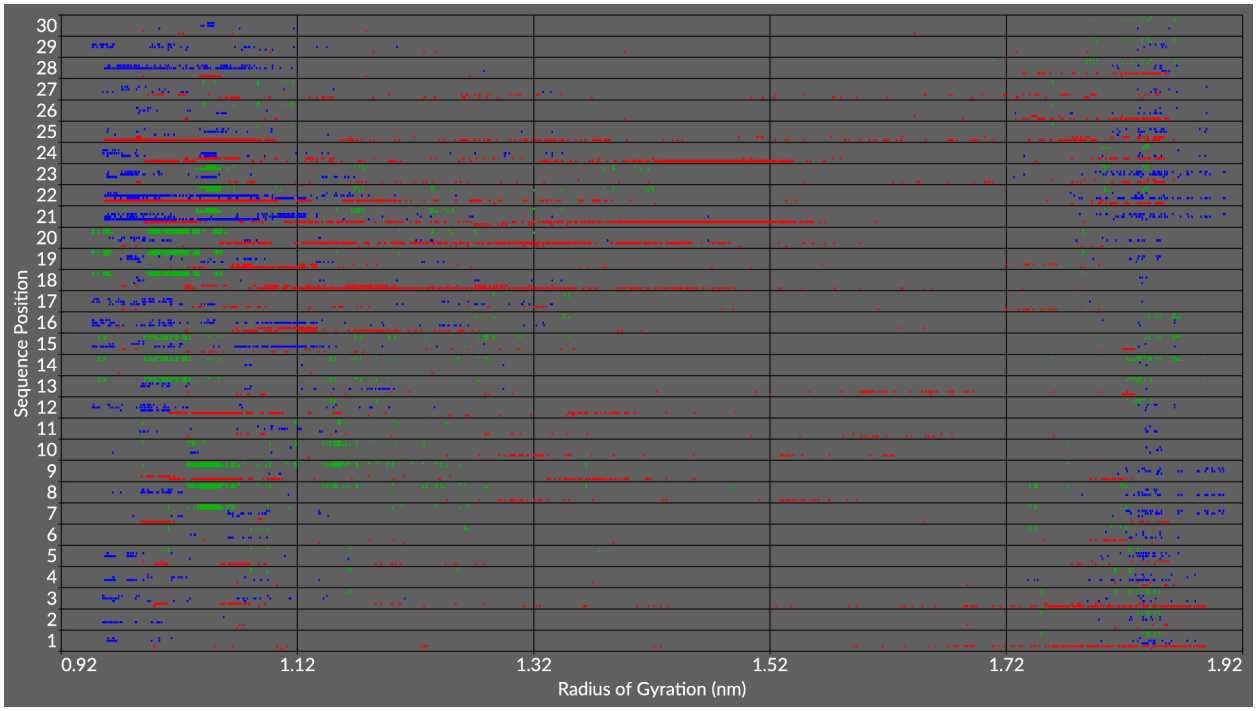


Fig. S14. Summary of DBs (red), WBs (green) and SCBs (blue) as a function of R_g and sequence position.

Tables

Table S1. Triple Gaussian fit to the radius of gyration distribution in Fig. S1. Uncertainties are one standard deviation of the fit.

Mean	Width in $e^{-\frac{(x-x_0)^2}{w^2}}$	Amplitude
1.102 ± 0.001	0.080 ± 0.003	162 ± 8
1.42 ± 0.01	0.25 ± 0.01	462 ± 11
1.673 ± 0.002	0.155 ± 0.004	501 ± 40

Table S2. Transit times obtained from logistic function fits, with transition intervals identified using STaSI analysis.

		Rg (nm)		Time (ns)		
Transition	MD	Begin	End	Begin	End	Transit
1	1	1.558	1.042	46.42	53.06	1.51
2	1	1.106	1.554	171.93	176.30	0.99
3	1	1.579	1.073	194.95	204.26	2.12
4	1	1.640	1.175	625.33	628.48	0.72
5	1	1.566	1.075	920.67	926.46	1.32
6	2	1.713	1.158	1506.96	1514.95	1.82
7	2	1.260	1.670	1557.16	1562.29	1.17
8	4	1.634	1.114	553.80	563.68	2.25
9	6	1.083	1.603	46.72	56.76	2.28
10	6	1.066	1.666	809.63	844.33	7.90
11	6	1.821	1.030	864.57	911.99	10.79
12	7	1.049	1.687	9.45	29.64	4.60
13	7	1.668	1.223	909.11	912.94	0.87
14	8	1.633	1.150	590.40	610.71	4.62
15	8	1.793	1.125	980.33	1079.29	22.52
16	9	1.163	1.726	49.41	55.86	1.47
17	9	1.586	0.803	316.56	353.32	8.36
18	9	1.606	1.199	1406.66	1411.08	1.01
19	9	1.170	1.518	1473.70	1475.86	0.49
20	10	1.038	1.646	211.09	225.66	3.32
21	10	1.638	1.170	247.39	252.24	1.10
22	10	2.059	1.167	885.89	931.17	10.30
23	10	1.125	1.650	1066.35	1085.28	4.31
24	10	1.111	1.673	1485.78	1489.95	0.95
25	11	1.673	1.119	35.55	56.82	4.84
26	11	1.685	1.062	418.12	427.51	2.14

Table S3. Transit times for transition intervals obtained from KiTE analysis that correspond to the STaSI intervals of Table S2.

		Rg (nm)		Time (ns)		
Transition	MD	Begin	End	Begin	End	Transit
1	1	1.472	1.205	45.80	49.35	1.48
2	1	1.222	1.625	172.95	178.30	2.21
3	1	1.650	1.192	193.45	200.20	2.79
4	1	1.523	1.285	625.40	628.10	1.13
5	1	1.604	1.119	918.55	924.70	2.54
6	2	1.465	1.201	1507.30	1513.40	2.52
7	2	1.236	1.566	1552.15	1561.75	3.96
8	4	1.654	1.245	553.70	559.95	2.58
9	6	1.212	1.561	50.05	54.95	2.03
10	6	1.143	1.477	818.90	835.80	6.95
11	6	1.590	1.194	879.95	897.20	7.09
12	7	1.190	1.569	17.90	26.20	3.42
13	7	1.569	1.188	910.50	913.35	1.19
14	8	1.586	1.174	594.05	606.25	5.02
15	8	1.553	1.202	1024.35	1034.10	4.02
16	9	1.222	1.604	49.40	53.75	1.80
17	9	1.602	1.172	317.45	333.55	6.62
18	9	1.598	1.248	1405.65	1410.70	2.09
19	9	1.179	1.491	1473.95	1476.95	1.25
20	10	1.248	1.594	215.00	223.55	3.52
21	10	1.556	1.172	246.60	251.25	1.93
22	10	1.576	1.204	908.35	923.60	6.27
23	10	1.170	1.591	1072.50	1083.30	4.45
24	10	1.262	1.521	1486.15	1489.05	1.21
25	11	1.594	1.154	41.60	50.70	3.75
26	11	1.587	1.232	418.95	425.50	2.71

Table S4. Principal component weights obtained from five starting coordinates described in the main text, covering 93.5% of the variance.

Vector	Rg (nm)	EPHOA (Å²)	EPHIA (Å²)	Bound Water	Wrap Water
PC1	**0.430324**	**0.464572**	0.434162	0.426205	**0.478379**
PC2	-0.428804	-0.331444	**0.53175**	**0.581192**	-0.292797

Table S5. Principal component weights obtained from three starting coordinates described in the main text, covering 95.8% of the variance.

Vector	Rg (nm)	Bound Water	Wrap Water
PC1	**0.581328**	0.533479	**0.614376**
PC2	-0.549694	**0.814198**	-0.186864

Table S6. Excerpt from time-stamped event list for trajectory 11.

Frame	Time (ns)	Element	Distance	Arity	Residue	Endpoint	Color	Type	Type_num
0	0.0	2	1	1	3	3.25	1	Direct	1
0	0.0	1	1	1	4	4.125	1	Direct	1
0	0.0	5	4	3	21	21.625	4	WaterBridge	2
0	0.0	3	4	3	22	22.375	4	WaterBridge	2
0	0.0	4	4	3	25	25.5	4	WaterBridge	2
0	0.0	7	1	2	15	15.875	1	NHN90	3
0	0.0	7	1	2	23	23.875	1	NHN90	3
0	0.0	7	1	2	24	24.875	1	NHN90	3
0	0.0	7	1	2	25	25.875	1	NHN90	3
0	0.0	7	1	2	26	26.875	1	NHN90	3
0	0.0	6	1	2	14	14.75	1	NHN90	3
0	0.0	6	1	2	23	23.75	1	NHN90	3
0	0.0	6	1	2	24	24.75	1	NHN90	3
0	0.0	6	1	2	25	25.75	1	NHN90	3
0	0.0	6	1	2	26	26.75	1	NHN90	3
1	0.05	2	1	1	3	3.25	1	Direct	1
1	0.05	1	1	1	4	4.125	1	Direct	1
1	0.05	5	4	3	21	21.625	4	WaterBridge	2
1	0.05	3	4	3	22	22.375	4	WaterBridge	2
1	0.05	5	4	3	25	25.625	4	WaterBridge	2
1	0.05	7	1	2	5	5.875	1	NHN90	3
1	0.05	7	1	2	6	6.875	1	NHN90	3
1	0.05	7	1	2	14	14.875	1	NHN90	3
1	0.05	7	1	2	15	15.875	1	NHN90	3
1	0.05	7	1	2	23	23.875	1	NHN90	3
1	0.05	7	1	2	24	24.875	1	NHN90	3
1	0.05	7	1	2	25	25.875	1	NHN90	3
1	0.05	7	1	2	26	26.875	1	NHN90	3
1	0.05	6	1	2	5	5.75	1	NHN90	3

Supporting Information Text

Distinguishing Coil from Globule states based on R_g values

To differentiate between the more collapsed states (globule) and more expanded states (coil), we looked at the Rg distribution (Fig. 1C), which shows a secondary peak indicating time spent in the collapsed state. We identified a trough in the histogram at 1.2 nm, between the coil and globule peaks. The histogram can be fit minimally to three Gaussians as shown in Fig. 1C, although the intermediate compactness population is heterogeneous and should be classified as an 'intermediate compact ensemble', not a structurally defined 'intermediate state' as in the case of protein folding.

Two methods were used to compute the location and duration of transitions: KiTE and STaSI.

1. Kinetic Transition Estimator (KiTE)

We developed the Kinetic Transition Estimator (KiTE) to estimate the time intervals and durations of transitions between coil and globule states.

This method begins by using a Gaussian Mixture Model (GMM) to characterize the distribution of R_g values. The GMM is composed of three Gaussian distributions, corresponding to the globule, intermediate, and coil states, that maximize the likelihood of the R_g data (16 µs, 50 ps steps). For each data point, a backward state (S_b) and a forward state (S_f) are assigned based on the state that maximizes the joint likelihood of the previous and next 1 ns of data, respectively. A final state is assigned only when the forward and backward predictions agree (S_f=S_b). When the predictions disagree, the state is labeled as 'unknown'.

After each data point is assigned a state label, the algorithm sequentially processes the data to identify transitions. The start of a potential transition is marked when the system enters either the coil or globule state. The frame number is then recorded and held until the opposite state is reached. A full transition is identified when the system moves from one stable state (e.g., coil) to the opposite stable state (globule), at which point the start (f_{start}) and end (f_{end}) frames are labeled. In cases of a frustrated transition—where the system leaves a stable state but returns to it without reaching the opposite state—the frame counter is reset, and these partial events are not included in the final analysis.

Since state labels are only assigned when the forward (S_f) and backward (S_b) predictions agree, the resulting transitions are a coarse estimate that includes undefined boundary regions. The resulting time interval contains three distinct phases: a pre-transition period where the system is leaving its starting state, the transition passage itself where the system is in a transitional state, and a post-transition period as it settles into the final state. This coarse definition yields an overestimate of the true transition time.

To estimate a time interval containing only the transition passage, we trim the boundaries of the coarse interval. We increment the start frame until the backward prediction, S_b, no longer matches the starting state, S_{start}, and decrement the end frame until the forward prediction, S_f, no longer matches the ending state, S_{end}. The resulting interval, from f'_{start} to f'_{end}, is a refined estimate that represents an underestimate of the full transition.

By averaging the overestimated and underestimated intervals, we obtain the final estimate of the transition interval:

$$\hat{f}_{start} = \frac{f_{start} + f'_{start}}{2} \qquad \hat{f}_{end} = \frac{f_{end} + f'_{end}}{2}$$

The final interval, $(\hat{f}_{start}, \hat{f}_{end})$ provides a balanced estimate of the transition interval. This approach allowed the KiTE algorithm to identify and characterize 220 transition intervals, as shown in Fig. S1A.

Transition Durations and Transit Times. From the obtained transition intervals, we compute the transition duration: $\hat{f}_{end} - \hat{f}_{start}$. These computed durations differ from the commonly used transit time τ (the time constant of a logistic fit) by a scale factor of 0.4098.

Efficiency. Once identified, each full transition is further characterized by its 'efficiency' computed as the ratio of 'R_g distance travelled' to 'R_g spanned by that transition':

$$\text{efficiency} = \frac{\left| R_{\text{g}}(\hat{f}_{end}) - R_{\text{g}}(\hat{f}_{start}) \right|}{\sum_{f=\hat{f}_{start}}^{\hat{f}_{end}-1} \left| x_{f+1} - x_f \right|}$$

Efficiency is a quantitative measure of the directness of the transition — an indication of how much time the molecule lingers in an intermediate state on its passage to or from the globule state. A threshold based on efficiency could be used to eliminate transitions with long dwell times in intermediate states.

Confidence. 'Confidence' estimates the likelihood that the transition interval is a direct transition between coil and globule:

$$\text{confidence} = \frac{\sum_{\hat{f}_{start}}^{\hat{f}_{end}} S_b \neq S_f}{\hat{f}_{end} - \hat{f}_{start}}$$

Confidence is large when the transition interval contains a large number of frames in the 'unknown' state (indicating that it is not lingering in a stable intermediate state). Fig. S1A shows all KiTE-obtained transitions, color coded by this confidence value.

KiTE locates transition intervals and computes transition durations: $\hat{f}_{end} - \hat{f}_{start}$; these differ from the commonly used transit time τ (the time constant of a logistic fit) by a scale factor of 0.4098.

Cooperativity index. From the transit times, total number of frames, and number of identified transitions, we computed the mean dwell time as 77.3 ns. The mean transit time is 4.71 ns. The cooperativity index is defined as the ratio of mean dwell time to mean transit time, which in this case is 16.4.

2. STaSI

For confirmation, we employed the step transition and state identification (STaSI) method (B. Shuang et al., "Fast Step Transition and State Identification (STaSI) for Discrete Single-Molecule Data Analysis", *J. Phys. Chem. Lett.*, 2014, 5, 18, 3157–3161, https://doi.org/10.1021/jz501435p), initially developed for discrete single-molecule data analysis, to identify different states of PNIPAM from R_g trajectories. STaSI detects step transitions using a Student's t-test and groups segments into states via hierarchical clustering, optimizing the number of states using a minimum description length (MDL) criterion. All analyses were performed in Matlab using the source code published as a supporting information in the above STaSI reference. We fitted the R_g trajectories to a three-state model, corresponding to the coil, intermediate, and globular states, which yielded roughly consistent state R_g values across independent trajectories. While increasing the number of states beyond three provided marginal improvements in fitting quality, the detection of transition periods remained consistent.

For a selection of representative direct coil-to-globular transitions, we quantified the transit time by fitting the trajectories to a logistic function:

$$R_g(t) = B + \frac{2h}{1+\exp(-t/\tau)}$$

where *B* and *h* are step size fitting values, and τ is transit time. We obtained 26 direct coil-to-globule (or globule-to-coil) transitions, with transit time (τ) summarized in Table S2.

Comparison to KiTE. Using the 26 corresponding KiTE transitions (summarized in Table S3), we used robust zero-intercept linear regression to determine the scaling coefficient between KiTE transition durations and STaSI transit times (Fig. S2). The distribution of all transit times obtained using KiTE has a mode at 2.21 ns (Fig. S3A), and the τ values obtained with STaSI (Fig. S3B) peaks at 2-3 ns.

C-P State Space Representation

To visualize the dynamics in the C-P state space, we first smoothed the time series of the five observables by convolving each time series with a Gaussian-weighted window of 10 ns. This window gives the greatest weight to the central ~3 ns, and is defined as:

$$w(n) = e^{-\frac{1}{2}\left(\frac{2\alpha n}{N}\right)^2}$$

where α = 4, *N* = 200 frames (10 ns) and *n* = -*N*/2 … *N*/2.

We then performed a principle components analysis (PCA) on these five smoothed observables. See Table S4.

A second PCA on R_g, Bound Water, and Wrap Water alone resulted in two principal components, PC1 (assigned to the horizontal axis) and PC2 (assigned to the vertical axis), which together covered ~96% of the total variance. See Fig. S4 and Table S5.

Listening to dynamic hydrogen bond motifs

To investigate the dynamics of hydrogen bonds as the PNIPAM 30-mer transitions between coil and globule conformations, we developed a data sonification tool (Fig. S4A). This tool displays the polymer's state visually within the C-P state space while providing a data-driven soundtrack of the time-dependent state of three distinct hydrogen bond motifs. These motifs include direct, intramolecular bonds (0th order), water bridges formed by a single water molecule linking up to four functional groups (1st order water bridges), and direct N-H…N bonds between side-chain functional groups (SCB interactions). The data-driven audio track of each molecular dynamics trajectory allows for a detailed analysis of which specific hydrogen bonds form, at what time, and in what order.

Using angle and distance search criteria, and identifying water bridges by molecule ID, for each bond motif, we generated a sparse 30×30 connection matrix for each frame of the simulation. The sparse connection matrix was reshaped as a time-stamped event list, where multiple bonds of each type can occur in each frame. See excerpt in Table S6. Fig. S5 shows the total count of each bond type with respect to sequence position.

Piano Roll visualization and sonification of H-bond dynamics

The “Piano Roll” shows time on the horizontal axis and sequence number (1 to 30) on the vertical axis. Each potential bond contact point is assigned to a fraction of the distance between residues in order to display multiple bond types and contact points for each residue.

In frames where a bond forms, its color in the piano roll plot indicates the distance between donor and acceptor in terms of sequence number [1, 30]. The visual piano roll shows the states of the bonds over a given time span, and the instantaneous bond states are conveyed by the sonification (Figs. S6-8).

Each position on the vertical axis is assigned a fixed frequency in log space (pitch). On each frame, each bond formed in that frame triggers a sound event: a “grain” constructed by multiplying an oscillator at the pitch (mapped to sequence number) by an adjustable duration “grain envelope” (Fig. S9).

The distance spanned by that bond is mapped to the amplitude and the duration of the grain. As many as 288 grains can be sounding at any given point in time (so the decay of one can overlap with the initiation of a newer one).

Each bond type (DB, WB, SCB) can be panned to any position in the stereo field to assist the listener in tracking streams of bond types when all are selected. Bond motif charts can be selected independently of the bond motif sound events, so you can listen to one type of bond while visually monitoring another.

Pitch, amplitude, and duration base values, deviations (how much the data affects the value’s deviation from the base value) and warping (shaping the deviation curves from log to linear to exponential) are all adjustable while listening.

The audio signal of each data-driven event is defined as:

$$x(t) = a_e \cdot w\left(\frac{t - t_e}{d_e}\right) \cdot \sum_{n=1}^{4} \frac{1}{n} \sin\left(2\pi n f_e t\right),\ f_e = 440 \cdot 2^{\frac{p_e - 69}{12}}\ Hz$$

where a_e, d_e, p_e, and t_e (amplitude, duration, pitch, start time) are mapped from the event data:

$$a_e = \frac{amp_{base} + amp_{dev} \cdot warp\left(\frac{Distance_e}{29}, amp_{warp}\right)}{Arity_e},$$

$$d_e = dur_{base} + dur_{dev} \cdot warp\left(\frac{Distance_e}{29}, dur_{warp}\right),$$

$$p_e = pitch_{base} + pitch_{dev} \cdot warp\left(\frac{Endpoint_e - 1.125}{29.75}, pitch_{warp}\right),$$

$$Endpoint_e = Residue_e + \frac{Element_e}{8},$$

$$Element_e = \begin{cases} 1 & \text{if DB N} \\ 2 & \text{if DB O} \\ 3 & \text{if WB N} \\ 4 & \text{if WB 1 O} \\ 5 & \text{if WB 2 O} \\ 6 & \text{if SCB N-H}\cdots\text{N} \\ 7 & \text{if SCB N}\cdots\text{H-N} \end{cases},$$

$Distance_e$, $Arity_e$, $Endpoint_e$ (span of the contacts, number of contacts, and the sequence number) are derived from the MD simulation. Control parameters amp_{base}, amp_{dev}, amp_{warp}, dur_{base}, dur_{dev}, dur_{warp}, $pitch_{base}$, $pitch_{dev}$, $pitch_{warp}$ are adjustable by the listener.

Composite hydrogen bond motifs over all 11 trajectories

To visualize the density of the hydrogen bond motifs as a function of R_g [0.92, 1.92 nm], we generated composite piano rolls over all 11 MD trajectories for 3000 R_g value bins. For R_g values where more than one bond of that bond type occurs in a bin, we show the maximum value of the bond distances within that 1/3000 nm wide bin. Alternative displays are available that show the median or mean value in each bin, or the mean, median, or maximum bonds with respect to PC1 on the horizontal axis.

Figs. S10-S14 show the various bond types and some combinations as a function of sequence (1-30 on vertical axis) and compactness (R_g from 0.92 to 1.92 along the horizontal axis).

Links to SI Videos not embedded in the main text

Additional Video examples: HEADPHONES recommended for stereo information

Recommended YouTube settings: Gear menu: Quality 1080p HD (minimum 720p60)
Headphones recommended for stereo information.
See Fig. S1A to see the MD trajectory and time in R_g

Abbreviations for Recurring motifs or patterns in video captions
• **EP** "Enclosing pattern": WB coincides with (and encloses) DB (or SCB) — shares one residue with DB & spans 1 or 2 residues beyond the DB span
• **VLR** "Very Long-range pattern": Very long range bond or WB (>10 residues apart) — corresponds with compact state
• **IP** "Inflection pattern (IP)": Longer range SCBs mark the "inflection points" in the C-P and in Rg
• **SB** "Sum of Bond distances pattern": Long sequence distance bonds(of all types) associated with collapsed state.
• **BBS** "Bouncing ball down (or up) stairs": maintains general trend of transition but with "bounces" or fluctuations
• **Trap** Rg starts transition, then stalls in intermediate state
• **TO** "Trade-off": Occasionally bond types trade off with one another during transition (e.g., WB with SCB, DB with WB)

Excerpt 2: MD 1 frames 3311 to 3823, 25.65 ns expand with slight fluctuation
https://youtu.be/04L91O-uatk

VLR WBs form successively shorter distance bridges as it expands
In collapse (1-24/27) (1-26/29)
then (5-12) (15-23)
~mid transition (8-12/13) (21/22-25)

TO Trade-off between WB & DB
WB (21-25) ends 3754
DB (22-25) starts 3754

IP SCB>2 at inflection points in the C-P curve

Excerpt 18: MD 2 frames 30464 to 31264, 40.05 ns expand with trap-like intermediate state
https://youtu.be/77iIvL0UiJ0

VLR WB & SCB cooperate to hold it in intermediate state for ~30 ns
Not much DB activity

Excerpt 20: MD 3 frames 10169 to 10489, 16.05 ns expand with small diversion
https://youtu.be/JWiRFAAQzhs

VLR WB (2/3 – 29/30) when in collapsed at the beginning
During expand they form (7-13/14) coincident with a small detour down before it finishes the expand

Excerpt 21: MD 4 frames 10983 to 11609, 31.35 ns collapse by soft stair steps
https://youtu.be/w4vfke1lqN4

VLR WB bridges increase in distance during collapse (1/2-30 by collapsed)
frame 11604: WB (5-24) coincides with a bounce followed by return to collapse

DB (14-17) forms during the collapse transition
gaps correspond to bounces in Rg value
transient (24-28) coincides with final drop into collapsed
In collapsed state, they are mostly absent

Excerpt 52-53: MD 8 frames 20425 to 26259, Collapse, sustain, expand
https://youtu.be/GFIKxT8499c

BBS ball bouncing down stairs in frames: 20475-21564
VLR DB (6 - 20) at bottom of the stairs cooperating with
VLR WB (1/3/4/6/ to 19-24) at the same time

EP DB (15-18) alternating bond and break (corresponds to bouncing in collapse state)
WB (14-18) break and form with DB

@23026 **EP** sustains (with a break) for 75.25 ns
DB (9-12)
WB (8-12)

@24579 **VLR** WB from 2/7 to 25/26 and 13 to 22/23
Corresponding to C-P movement in anti-diagonal direction (low PC1 and low PC2)
Possible distinct state marked by **VLR** WBs: lower/left region of the C-P space

Excerpt 69: MD 10 frames 6624 to 7498, 43.75 ns expand with ~34.9 ns trap
https://youtu.be/Dw3cEMg5ZQ8

VLR WB in collapsed state at beginning (6/7/9 - 27/28/30)
Sustained WB (18,22,24) during trap
Undulations WB (12/13-20) and (25-28)

Excerpt 78: MD 11 frames 8363 to 8579, 10.85 ns collapse with gentle bounces
https://youtu.be/JXyHgVviBVQ

Sequence of successive water bridges form as it collapses:
(10 - 14/15), (18-21), (20-22), (9,14,15)
WB **VLR** pattern at end of transition (1/5-27/30) and in collapsed state (5/6-27)

BBS slight bouncing down stairs pattern