

RESEARCH ARTICLE

Mechanism of β -hairpin formation in AzoChignolin and Chignolin

Richard L. Zschau | Martin Zacharias 

Physics Department and Center of Protein Assemblies, Technical University of Munich, Garching, Germany

Correspondence

Martin Zacharias, Physics Department and Center of Protein Assemblies, Technical University of Munich, 85748 Garching, Germany.

Email: martin.zacharias@ph.tum.de

Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Number: DFG-ZA153/28-1

Abstract

AzoChignolin is a photoswitchable variant of the mini-protein Chignolin with an azobenzene (AMPP) replacing the central loop. AzoChignolin is unfolded with AMPP in the trans-isomer. Transition to the cis-isomer causes β -hairpin folding similar to Chignolin. The AzoChignolin system is excellently suited for comprehensive analysis of folding nucleation kinetics. Utilizing multiple long-time MD simulations of AzoChignolin and Chignolin in MeOH and water, we estimated Markov models to examine folding kinetics of both peptides. We show that while AzoChignolin mimics Chignolin's structure well, the folding kinetics are quite different. Not only folding times but also intermediate states differ, particularly Chignolin is able to fold in MeOH into an α -helical intermediate which is impossible to form in AzoChignolin. The Markov models demonstrate that AzoChignolin's kinetics are generally faster, specifically when comparing the two main microfolding processes of hydrophobic collapse and turn formation. Photoswitchable loops are used frequently to understand the kinetics of elementary protein folding nucleation. However, our results indicate that intermediates and folding kinetics may differ between natural loops and photoswitchable variants.

KEYWORDS

Markov state models, molecular dynamics, protein kinetics, single-protein analysis, β -Hairpin

1 | INTRODUCTION

For a long time, understanding the dynamics of how a sequence of amino acids folds into and in between intermediate states and its native state has remained an open question. While molecular dynamics allow the simulation of folding at atom-resolution, the massive amount of simulation data provided using today's hardware can be further leveraged with Markov state models (MSMs) to extract information about the dynamics of the system.

AzoChignolin,¹ a photoswitchable Chignolin-mutant, is a short 9-residue peptide (GYDP-AMPP-GTWG) and was derived by substituting the two loop residues in Chignolin (GYDPETGTWG) with

the azobenzene photoswitch AMPP ([3-(3-aminomethyl)phenylazo]phenylacetic acid). The AMPP chromophore exhibits cis-trans isomerism and allows, upon illumination, a change of the global structure of AzoChignolin due to its center position in the loop. Cis-AzoChignolin is able to form a β -hairpin structure similar to native Chignolin, while trans-AzoChignolin mostly exhibits disordered elongated structures, as trans-AMPP has a slightly longer end-to-end distance and thus hinders interaction between the two strands of trans-AzoChignolin.^{1,2} The ability to induce synchronized folding of AzoChignolin by sudden illumination allows one to directly follow the process using techniques such as time resolved IR spectroscopy. Since the folding time scale of the structural motif is also accessible using atomistic molecular

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Journal of Computational Chemistry* published by Wiley Periodicals LLC.

dynamics (MD) simulations it is an ideal system to study the thermodynamics and kinetics of hairpin loop formation. Indeed, in a previous study² utilizing time-resolved IR spectroscopy and MD simulations it was shown that folding of cis-AzoChignolin in MeOH is driven by formation of a hydrophobic core and intrastrand hydrogen bonds, resembling Chignolin's folding.

In the present theoretical study we aim to follow up on the previous study by providing a more detailed description of the folding process of cis-AzoChignolin (in the following simply referred to as AzoChignolin) and how it relates to the behavior of its natural counterpart Chignolin. For this purpose, a significantly larger amount of atomistic MD simulation data was generated and subsequently analyzed with the help of Markov state models. The choice of MeOH as solvent in the previous study was due to the large hydrophobicity of AzoChignolin (particularly AMPP) which easily leads to aggregation. Hence, to remain comparable to the previous TRIR experiments, MD simulations were primarily conducted in MeOH, but also in water, which allows us to further study the effects of solvent on folding.

Note that the Chignolin in this study follows the first peptide design described by Honda et al.³ not the CLN025 mutant where N- and C-terminal Gly were replaced with Tyr.⁴ Due to the small size of the system a rather comprehensive MD sampling for the Chignolin and AzoChignolin systems is possible resulting in well converged Markov state models. It offers a detailed characterization of intermediate states and associated transition kinetics for a protein loop model system that can typically act as nucleus for folding of protein molecules.

2 | METHODOLOGY

2.1 | Simulation setup

The Chignolin and AzoChignolin structures are based on the NMR structure *1uao*.³ Analogous to experiment^{1,3} and previous computational studies,^{5,6} the terminal glycines were left uncapped within H₂O solvent, carrying a positive and negative charge, respectively. AzoChignolin¹ was created by replacing the loop residues Glu5 and Thr6 in Chignolin with azobenzene (AMPP, see Figure 1). For structures in methanol, charged residues (Asp and Glu) were replaced by protonated versions and charged terminal residues (Glu) were replaced by neutral termini. The peptides were parameterized using Amber's ff14SB protein force field,⁷ a force field shown to perform well on IDPs.⁸ Custom residues AMPP and neutral Glu were parameterized using Amber's GAFF and the Antechamber toolkit.^{9,10} tleap was used to build the starting structures and the pmemd.cuda routine of the Amber package was used for performing the MD simulations.¹¹

Simulations were conducted in explicit TIP3P water¹² and in methanol¹³ at 298 K, together with neutralizing Na⁺ counterions in water and no additional salt. The prepared peptide structures were placed within periodic octahedral solvent boxes and a minimum distance between peptide and boundary of 18 Å. The systems were then minimized via Steepest Descent for 10 k steps, heated up step-wise to a target temperature of 298 K within 60 ps and equilibrated within

NVT and NPT conditions for 1 ns each before starting the production runs at NVT conditions. For the thermostat, Berendsen weak coupling¹⁴ was used with a coupling time of 0.1 ps, and for the barostat, isotropic pressure scaling with a reference pressure of 1 bar and a relaxation time of 1.0 ps. Peptide structures were restrained onto backbone atoms during heating and equilibration using harmonic potentials with a force constant of 2.388 kcal/(mol Å²). Hydrogen mass repartitioning¹⁵ and SHAKE bond constraints¹⁶ were used to allow a time step of 4 fs for the production runs. Long-ranged Coulomb interactions were calculated using Particle Mesh Ewald with a cut-off radius of 9 Å.¹⁷

Hydrogen bonds were counted if donor and acceptor are less than 3.5 Å apart with an angle of larger than 135°. The minimum root mean square deviation (RMSD) was calculated using the backbone of the NMR structure *1uao*. The RMSD of AzoChignolin excludes the turn residues Glu5 and Thr6 of *1uao*. Secondary structures were assigned using the DSSP method.¹⁸

2.2 | Markov model estimation

In order to approximate and study the folding dynamics of each system, molecular dynamics simulations were used to construct Markov State Models (MSMs). The use of MSMs allowed the simulation of several trajectories in parallel, compared to one very long trajectory. For a given system, one or more initial trajectories of length 1–2 μs were simulated from which around 4–8 starting structures were seeded by applying density peaks clustering and taking the cluster centroids.¹⁹ These were simulated for 4–5 μs each.

In order to improve the resulting MSMs, systems which did not show sufficient Markovianity at this point, based on the convergence of ITS and Chapman-Kolmogorov tests, were additionally sampled for more rounds, following existing adaptive sampling methods^{20,21} where each new round consists of independent short trajectories of length 1–3 μs. Resampling rounds were performed until the tests showed sufficient converging behavior. Each subsequent round seeds new trajectories from all frames sampled so far (excluding the initial seeding) by first creating a Markov model of all frames, coarse-graining it using PCCA++ and sampling from all metastable states with the probability $p_i = 1/c_i$, where c_i is the sum of stationary probabilities for microstates in metastable state i .²¹ In total, each system of interest was simulated for approximately 50–60 μs, distributed across multiple shorter trajectories (see SI Table S1 for total lengths of each system). MSMs were then constructed by assigning each frame within a set of trajectories to a cluster (discretization step, see below) and counting the transitions between each cluster, separated by a lag time τ_{MSM} . The transition matrix T_{ij} resulting from the count matrix C_{ij} serves as a kinetic model of the system and can be used to derive observables such as relaxation times, which may be compared to experiment.

The general pipeline was the same for each system: The structures were featurized using distances between contacting atoms with a contact life time of 10 ns or above, analogously to Plattner et al.,²²

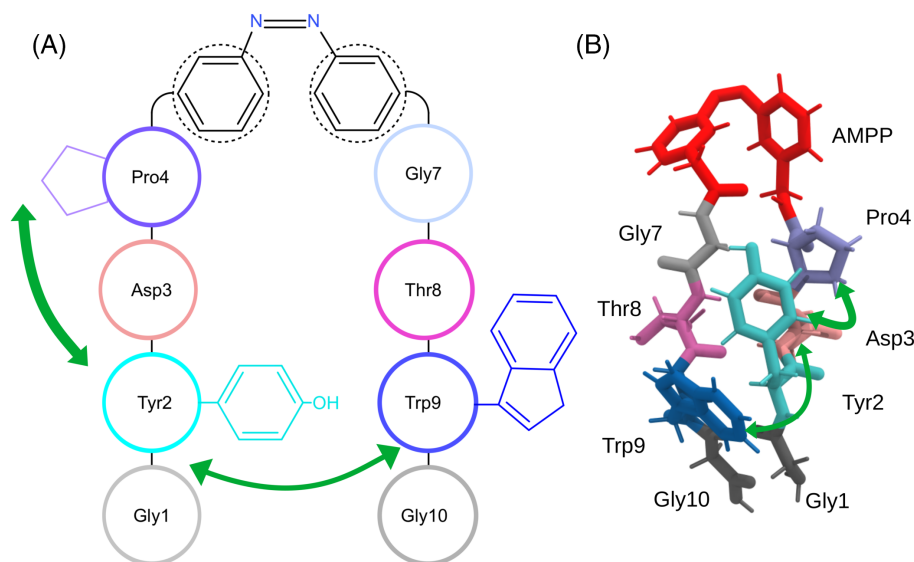


FIGURE 1 (A) Schematic of AzoChignolin (GYDP-AMPP-GTWG) with AMPP inside the loop in its cis-isomer, forming a β -hairpin. Green arrows denote main hydrophobic interactions. (B) Representative 3D structure of β -hairpin AzoChignolin from a simulation, demonstrating the “sandwich” stacking between Trp9-Tyr2 and Tyr2-Pro4. The latter interaction is more frequently observed, indicated by bold arrows

as well as backbone and sidechain dihedrals. Each dihedral angle is represented as a pair of their sine and cosine. The featurization set was chosen among different choices of featurization based on the VAMP-2 score (see SI Figure S1). In a prior CLN025 study the dihedral feature set alone was insufficient to accurately capture the free energy landscape.²³ The featurized trajectories were then reduced to 10–12 independent components (tICs) using time-lagged independent component analysis (tICA)^{24,25} with a lag time of 20 ns and kinetic map.²⁶ The choice of tICA lag time and number of tICs was based on the point of convergence of kinetic variance (see SI Figure S2).

K-means was used to cluster the trajectories into 600 clusters with `kmeans++` seeding. The discretized trajectories were first used to generate Markov models to visualize the slowest implied time scales (ITS) at different lag times (SI Figures S3–5). The timescale where the timescales start to converge was used to generate the actual maximum likelihood Markov model used for analysis. Markovianity of the resulting models was validated via the Chapman-Kolmogorov test²⁷ (SI Figures S6–8), which is implemented with `PCCA++` coarse-graining.²⁸

To cluster microstates into interpretable macrostates, Markov models were further coarse-grained to Hidden Markov Models (HMMs) where the number of coarse-grained macrostates was determined based on gaps between the implied timescales of the microstate model.²⁹

To compare tICA decompositions of two or more systems with slightly differing topology, a similar approach to McKiernan et al.³⁰ was used: A common feature set was created between two systems by taking the union of distance features which are valid for both topologies and similarly the union of backbone and sidechain torsions which exist in both topologies. A “base” tICA model was then estimated on one of the systems and used to transform the feature trajectories of the remaining systems.

For featurization, model estimation and validation, the `PyEMMA` toolkit³¹ was used together with `MDTraj`.³² Additionally, `CPPTRAJ`

and the Python module `pytraj`³³ were used for further trajectory analysis.

3 | RESULTS AND DISCUSSION

Experimental studies on time resolved folding have been performed with AzoChignolin in MeOH² and simulations on this system are also at the focus of our simulation study. In order to investigate the effects of the incorporation of the AMPP photoswitch into AzoChignolin we also compare it with the folding behavior of native Chignolin in MeOH solution. Since the change of solvent from water to MeOH has significant effects on the stability of the systems, for example, decreasing the protein melting temperature, AzoChignolin in MeOH is only directly comparable to Chignolin in MeOH. For further comparison we also conducted simulations of both peptides in water. Note that since AzoChignolin in water was not a primary system of interest, we performed in this case a smaller number of simulations (see SI Table S1).

3.1 | AzoChignolin folding landscape

Based on the 50 μ s of simulations for AzoChignolin in MeOH, we estimated a coarse-grained MSM of HMM states, with which AzoChignolin's kinetic landscape can be split into five meta-stable states or distinct regions of the tICA landscape (see Methods). Figure 2.a shows each HMM state and an overlay of representative structures, together with mean first passage times (MFPTs) for inter-state transitions. For visualization purposes, the states are positioned on top of the free energy landscape spanned by the second and third tIC.

State 3 forms the β -hairpin state which resembles the β -hairpin of native Chignolin, yet only amounts to 18.5% of the stationary distribution. The largest part of this distribution is dominated by disordered

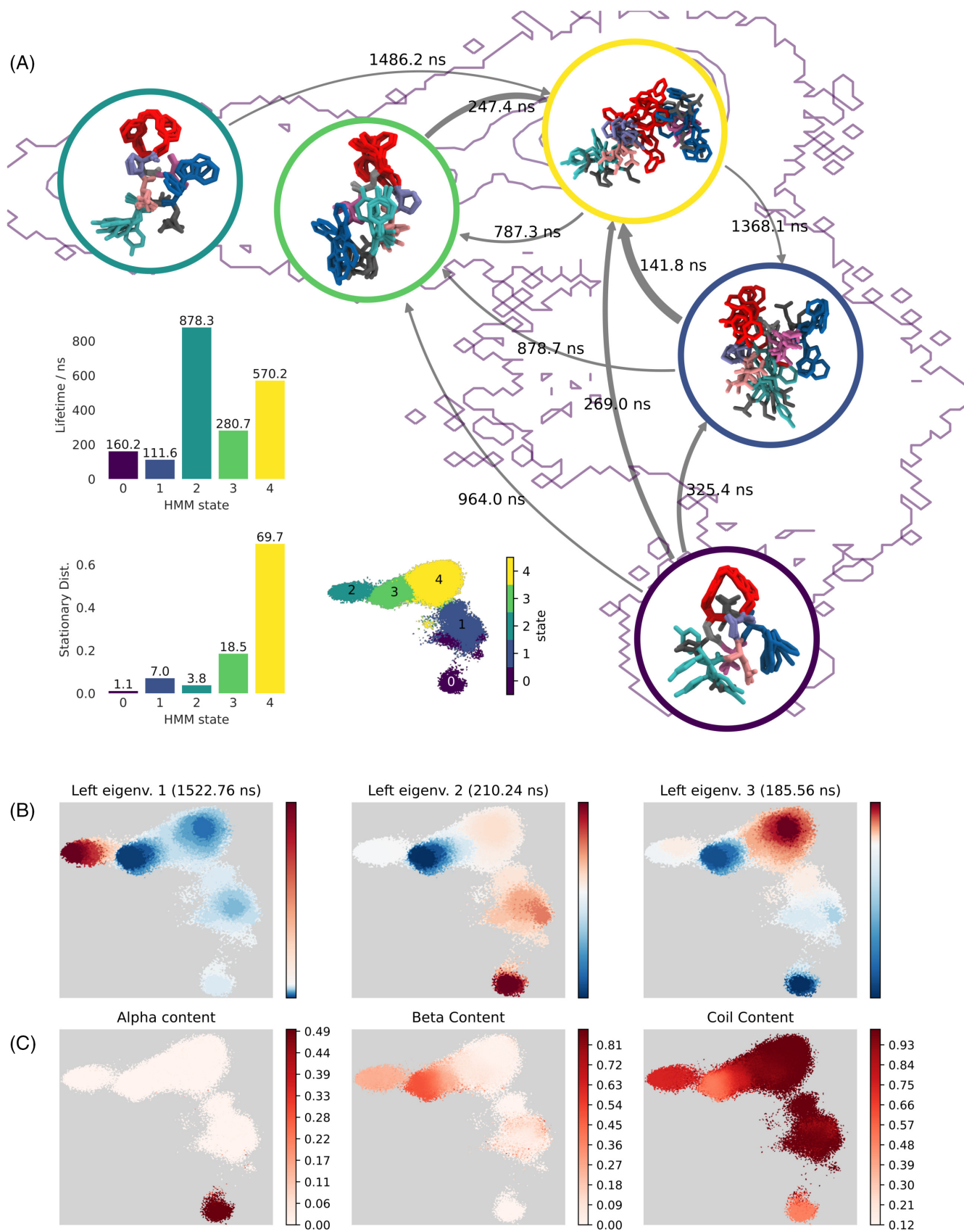


FIGURE 2 Legend on next page.

state 4 which contains mostly disordered structures with high structural diversity, accompanied to a lesser degree by state 1 which similarly features disordered structures. Together, these two states occupy 76.9% of the stationary distribution. Structurally, the main feature to distinguish between the two disordered states is a different ϕ -backbone dihedral of Trp9 and a resulting difference in local curvature of the second strand end. The preference for AzoChignolin to stay unfolded in MeOH suggests that the protein melting temperature lies below 298 K for the used solvent and force field. Most trajectories move between states 3 and 4, with partially folded structures located in between these states. No stable misfolded states are identified by the HMM along this major path, which matches observations in previous MSM studies for the similar protein CLN025.^{30,34,35}

Apart from these three major states, AzoChignolin also forms two rarer intermediates which were only visited after longer sampling: State 2 resembles a misfolded hairpin at first, but lacks many of the interactions of the β -hairpin. Its two strands are oriented approximately perpendicular to each other, in such a way that Tyr2's side-chain is mostly isolated and unable to participate in a hydrophobic core. However, Trp9 is able to interact with the second benzene ring of AMPP. Its radius of gyration is the lowest of all states.

The very rare state 0 contains α -helical structures according to DSSP (see Figure 2B, SI Figure S9). It features curled up strands, Tyr2 and Trp9 on opposing sites which are unable to interact with each other, and a hydrophobic core between Tyr2, Pro4 and one benzene ring of AMPP. Due to its globular shape, its radius of gyration is lower than in hairpin state 3, but slightly higher than in state 2.

The notably higher stationary distribution of β -hairpin state 3 over states 0 and 2, despite both of these exhibiting a lower radius of gyration, suggests that the specific interactions found in state 3, particularly the cross-strand hydrophobic Tyr2-Trp9 interaction missing in the other states, may be more important for structural stability than compactness. The alternative hydrophobic cores from states 0 and 2 resemble the mechanism found in native Chignolin folding, where one of the participants of the hydrophobic core (Tyr2, Pro4, and Trp9) is bound to another part of the protein and thus rendered unavailable, also known as “packing frustration”⁶ which is suspected of delaying the formation of the native hairpin.

Kinetically, disordered state 4 acts similar to a hub state, connecting most other metastable states. Despite the structural similarity between state 2 and state 3 (hairpin-like and hairpin), there is no direct connection between these states. Based on the HMM's transition matrix, the only path to enter and leave state 2 is via hub state 4. The lifetimes further implicate that state 2 acts as a trap state, with

the highest lifetime of 878 ns, while folded state 3 only has a lifetime of 280 ns (Figure 2A, bottom left).

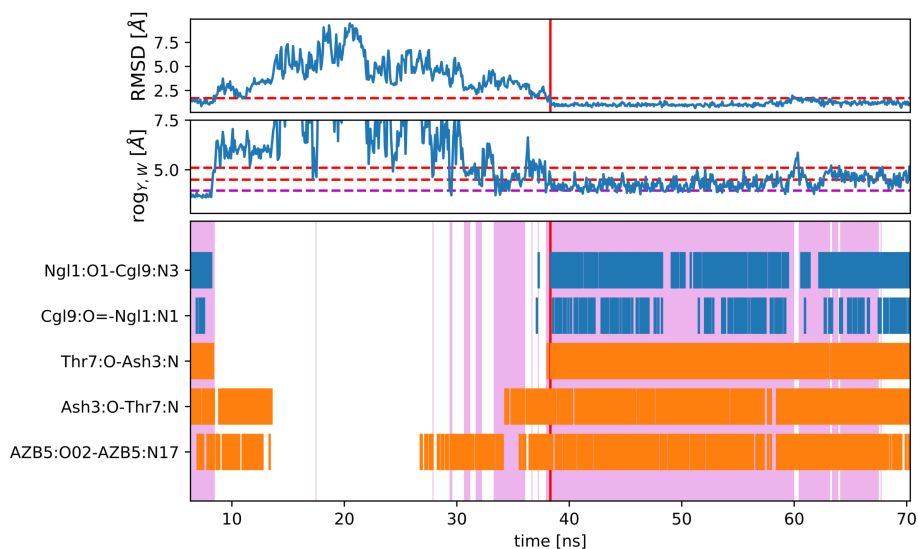
The dynamical processes of the system can be extracted from the Markov model by decomposing its transition matrix into eigenvectors and their eigenvalues. Figure 2C shows the first three weighted (left) eigenvectors, projected onto the tIC2-tIC3 landscape, excluding the 0-th eigenvector which describes the stationary process. Each eigenvector describes a transition process between microstates with positive and negative value, with an associated timescale given by their eigenvalue: $t_i = -\tau / \ln |\lambda_i|$. For visual clarity we chose not to use the first tIC, as it mostly only separates between trap state 3 and all other states, while the second and third tIC are better at separating the larger states and are likely more related to folding.

The first eigenvector is by far the slowest, with a timescale of 1523 ns, and describes transitions between trap state 2 and the remaining states. The slow timescale of this process, and similarly the slow MFPTs between states 2 and 4, can be explained with the relatively high energy barrier separating state 2 from the other states, with approximately $\Delta G_{\text{barrier}} = 2.4$ kcal/mol based on the tIC free energy (see SI Figures S10 and S11). The next two eigenvectors are more related to the folding process, as they involve mainly transitions from and to the hairpin state 3 (eigenvector 2) with a timescale of 210 ns and mainly transitions from and to the disordered hub state 4 (eigenvector 1) with a timescale of 185.5 ns, matching autocorrelation timescales of a previous MD study.²

The fully folded β -hairpin of AzoChignolin is stabilized by hydrophobic interactions between mainly Tyr2-Trp9 and to a lesser extent Tyr2-Pro4, similar to native Chignolin. While additional hydrophobic interactions would be possible with the introduced AMPP turn as well, they are only observed within the other states. Additionally, the hairpin forms a number of hydrogen bonds between Gly1-Gly10, Asp3-Thr8 and an intra-residue bond within AMPP, matching the findings in the previous study² (SI Figure S12). However, different to their progression of hydrogen bonds, we found that the intra-H-bond AMPP-AMPP usually precedes the key bond Asp3-Thr8. The formation of this hydrogen bond is likely a nucleation event analogous to turn nucleation in Chignolin, as it initiates and stabilizes the formation of the turn region by constraining the shape of the AMPP. Once the AMPP-AMPP hydrogen bond has formed, the rest of the hydrogen bonds mostly form in a zipper-like fashion, progressing from the turn towards the strand ends with Asp3:O-Thr8:N, Thr8:O-Asp3:N and finally Gly10:O-Gly1:N and Gly1:O-Gly10:N to form usually last (see Figure 3). Interestingly, rare instances of folding were observed as well where zipping happens from the strand ends towards the turn,

FIGURE 2 (A) Mean first passage times (MFPTs) between HMM states of the AzoChignolin (MeOH) ensemble, placed on the tICA landscape spanned by second and third tIC. Thicker arrows denote faster passage times. Transitions with MFPTs above 1500 ns were excluded (this includes the slow transition into state 2 from state 4). States 0, 2 and 3 are structurally homogeneous, while states 1 and 4 are largely disordered with high structural variability. State 4 has the highest stationary distribution and also acts as a hub, connecting all the other states. State 3 corresponds to the native β -hairpin. Bottom-left insets: Lifetimes and stationary distribution of each HMM state. (B) First three left eigenvectors (excluding the stationary process) projected onto tICA landscape with their relaxation timescales, indicating transitions between positive (red) regions and negative (blue) regions. The folding process is mostly expressed by eigenvectors 2 and 3. (C) Amount of secondary structure distributed across the tICA landscape, showing states 2 and 3 containing β -sheet structure and state 0 α -helical structure

FIGURE 3 Example of a folding event. Upper panels: Backbone RMSD to the folded peptide and radius of gyration of Tyr2 and Trp9 (describing side-chain packing). The red vertical line denotes the moment where the RMSD reaches below 1.5 Å. Lower panel: Key hydrogen bond events. Bonds near the tail are in blue, near the turn in orange. A magenta background denotes where the radius of gyration of Tyr2, Trp9 is below 5.1 Å



initiated by the salt bridge Gly10:O-Gly1:N (SI Figure S13). Different from Chignolin, Asp3 in Azochignolin's β -hairpin lacks many of the native interactions with residues on the other strand other than Thr8.

AzoChignolin exhibits the same major folding pathways as seen in Chignolin, the hydrophobic collapse model and the zipper model. A more thorough examination of the timescales of each microfolding process is described in section "Comparison of dynamics," revealing faster relaxation times associated with hydrophobic collapse compared to turn formation and confirming the observations above.

3.2 | Speed of folding after light-induced isomerization

The speed at which AzoChignolin folds (and unfolds) in MeOH was measured in the experimental study of Hofmann et al.² using time-resolved infrared spectroscopy (TRIR) where a trans-ensemble of the molecule was flipped to mostly cis via excitation at 355 nm, allowing the formation of β -hairpin. A timescale of folding was then quantified by fitting the IR absorption changes at 1632 cm^{-1} and around 1700 cm^{-1} within the amide I band which both correspond to antiparallel β -structure in the backbone,³⁶ yielding timescales of $\tau_{\text{str},1} = 54$ ns (stretched exponential fit) and $\tau_{\text{str},2} = 54$ ns and $\tau_2 = 90$ ns (sum of previous stretched exponential and single exponential fit), respectively.²

To answer the original question of the speed at which AzoChignolin folds using conventional MD simulations, we replicated the isomerization event by applying a harmonic dihedral restraint on the backbone of AMPP on 50 different structures of trans-AzoChignolin. The trans-Azochignolin structures were seeded from a short 1 μs simulation, sufficiently long due to the large structural variability of trans-Azochignolin. After flipping the AMPP from trans to cis, the resulting elongated structures of cis-AzoChignolin were relaxed and used as starting points for 50 short MD simulations with 1000 ns each.

Out of all 50 short simulations most (40 simulations) come close to the fully folded hairpin within the simulation time, measured by a backbone-RMSD threshold of 1.5 Å, but only 20 simulations reach an RMSD of 1.0 Å (Figure 4B, SI Figure S14). The average arrival time (the minimum time needed to arrive below the RMSD threshold) is 389 ns for a 1.5 Å threshold and 453 ns for a 1.0 Å threshold.

Note that these mean arrival times, particularly the latter, would be much higher if all simulations were included in the average by prolonging the simulations beyond 1000 ns until all have reached the low-RMSD region. The Markov model estimated on the longer ensemble of simulations allows us to also quantify the transition with the mean first passage time (MFPT). Since the unfolded structures right after isomerization are all located within the large unfolded HMM state 4, the relevant MFPT would be the time to go from state 4 to the folded state 3, which is 787 ns (see Figure 2A).

The TRIR experiment above can be described as a perturbation-relaxation experiment in which an ensemble with an off-equilibrium distribution is created, populated by unfolded structures right after trans-cis isomerization. The measured time is the time needed for the ensemble to relax to the equilibrium distribution, which is a mixture of folded and unfolded structures. The MSM can similarly be decomposed into eigenvectors, describing kinetic processes occurring at timescales given by their eigenvalues. The experimentally measured relaxation times of the given observables can be related to the timescales of the MSM, a concept also known as dynamic fingerprints.^{37,38} A dynamic fingerprint, a vector of coefficients, describes the impact of each eigenprocess in a relaxation profile seen in experiment $f(t) = \gamma_1 + \sum_{i=2}^m \gamma_i \exp(-t/t_i)$, and can be calculated for a perturbation experiment with $\gamma_i = (\mathbf{a}^T \mathbf{l}_i) ([\mathbf{p}'(0)]^T \mathbf{l}_i)$ where a_k is the mean observable for microstate k , \mathbf{l}_i the i -th left eigenvector and $\mathbf{p}'(0)$ the initial perturbed distribution.³⁷

Assuming that the absorption changes ΔA in the amide I band at wavenumbers related to β -structure (the true observable) can be approximated as proportional to the increase of β -content in the experimental ensemble, we define an analogous observable a_k as the

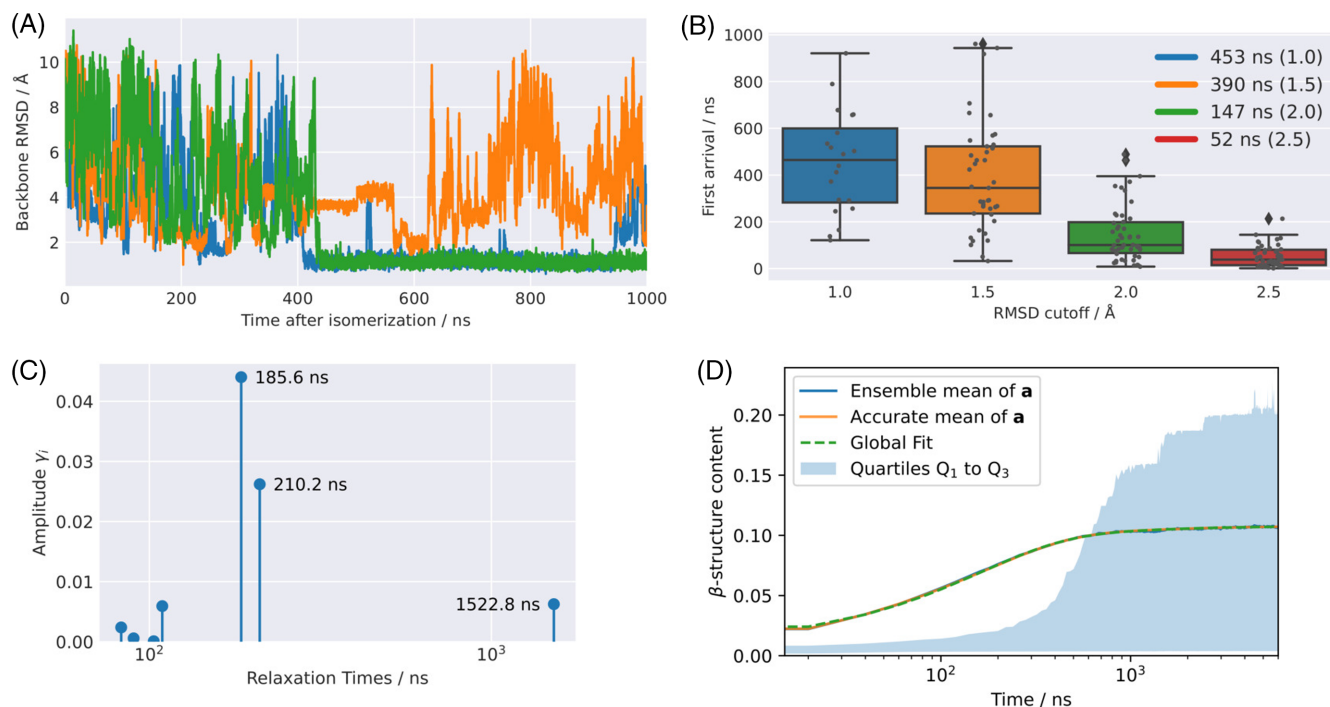


FIGURE 4 (A) Backbone RMSD to the folded hairpin of three exemplary trajectories of AzoChignolin (MeOH). While two trajectories converge to a structure with an RMSD below 1 Å after around 400 ns, the other fails to stay in the low-RMSD region. (B) Minimum arrival times for trajectories to reach a certain RMSD region, denoted with gray dots. Box width denotes quartiles, box line the median, whiskers the extent excluding outliers. (C) Dynamic fingerprint calculated using MSM of long trajectories and β -content based observable, suggesting that the TRIR experiment is dominated by two timescales at 186 ns and 210 ns. (D) TRIR-perturbation-relaxation experiment replicated with MSM, using β -content based observable related to the absorption change ΔA . The quartile range (Q_1 – Q_3) is better at showing the time propagation of the kinetic ensemble

mean number of β -structure residues in MSM state k , given by DSSP. The initially perturbed distribution $\mathbf{p}'(0)_k$ is defined by the distribution of the first 250 frames (20 ns) of all 50 short trajectories after isomerization. The resulting fingerprint (see Figure 4C) is dominated by two timescales at $\tau_3 = 185.6$ ns and $\tau_2 = 210.2$ ns, while implicating that the slowest timescale at 1522 ns as well as faster timescales are barely expressed in the given observable. We can further replicate the experiment by simulating an ensemble of $N = 100,000$ synthetic state trajectories with the Markov model given starting states sampled from $\mathbf{p}'(0)_k$ and visualizing the observable across all N trajectories. The average of the resulting observables closely resembles the relaxation profile of the change in absorption ΔA in experiment (see Figure 4D). This average can also be more accurately calculated using $f(t)$ from above or by calculating the expectation value $\mathbb{E}_a(n\tau) = \mathbf{p}'_0 \mathbf{P}(\tau)^n \mathbf{a}$ for values of n (where $\mathbf{P}(\tau)$ is the probability matrix of the MSM), which matches the previous average in the limit of infinite synthetic trajectories.³¹ Further, we also estimated a fit using a sum of a stretched exponential and single exponential onto the expectation value, analogously to the fit of ΔA for 1700cm^{-1} in experiment, yielding a fit with timescales $\tilde{\tau}_1 = 133.8$ ns and $\tilde{\tau}_2 = 166.8$ ns which are noticeably faster than the dominating underlying timescales τ_3 and τ_2 of the ground truth profile. Differences to the timescales in experiment may be due to bias of the used force fields, differences between the true experimental

observable and our observable, the purity of cis-AzoChignolin in experiment as well as statistical uncertainties in our Markov model.

3.3 | Chignolin folding landscape

In order to make a meaningful comparison between the kinetics of Azochignolin and Chignolin, we also estimated a Markov model on trajectory data of Chignolin solvated within MeOH. Chignolin's main kinetic differences stem from the different middle region where Glu5 and Thr6 are located instead of AMPP. Since these two residues lack the intrinsic turn-like shape which AMPP would provide, the kinetics of Chignolin's turn-formation are more complex than in Azochignolin, leading to an easier tendency to form misfolds due to an incorrectly shaped or not correctly positioned turn. A comparative analysis was also performed for Chignolin in water (further details are provided in the SI).

The tendency to misfold is demonstrated by the significantly different energy landscape of Chignolin, which can be split up into 6 regions given by a coarse-grained HMM, visualized on the tICA landscape spanned by the first and second tIC (see Figure 5A, SI Figure S15). State 5 is the native β -hairpin, amounting to 32.2% of the stationary distribution, which is significantly lower than for Chignolin in water at the same temperature.

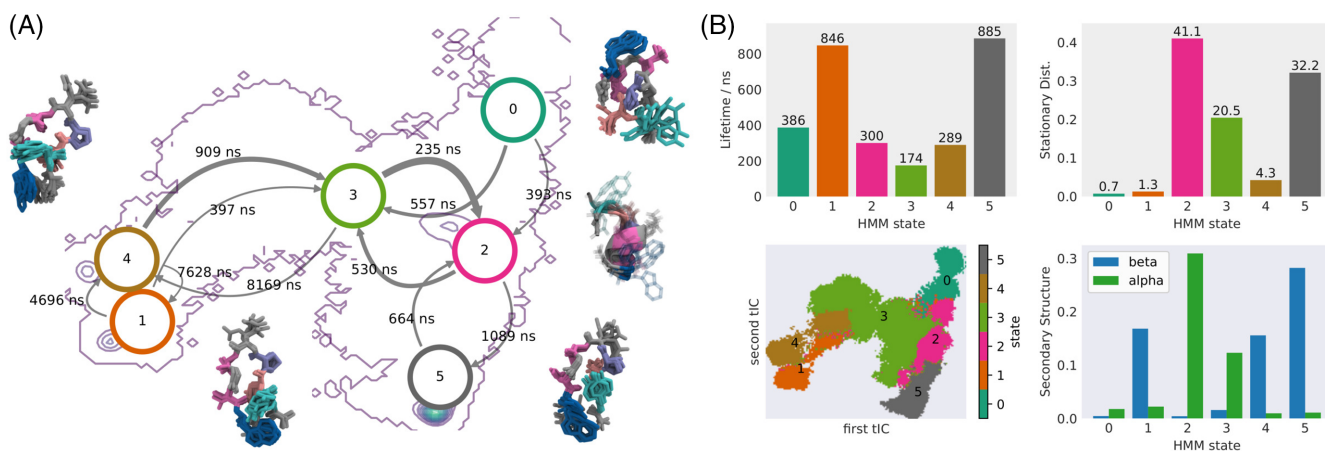


FIGURE 5 (A) Mean first passage times (MFPTs) between HMM states of Chignolin (MeOH) ensemble, placed on tICA energy landscape together with representational snapshots. Thicker arrows denote faster passage times. Transitions with a transition probability below 0.2% were excluded. State 3 has a very high structural variability and is therefore not pictured. State 2 features helical substates among many other disordered structures, but only helical substate is pictured. The native β -hairpin is state 5. (B) Lifetimes, stationary distribution and secondary structure distributions for all HMM states. The somewhat helical state 2 dominates the energy landscape

The landscape is dominated by state 2, which is populated by a mixture of disordered and α -helical structures (SI Figure S16). Based on the transition matrix, state 2 functions as a gateway for state 5, as other states need to pass through state 2 before they can reach the β -hairpin. The α -helical substate of state 2 is mainly stabilized via hydrogen bonds, since the hydrophobic side-chains Tyr2, Pro4 and Trp8 are positioned so far from each other that they would be unable to form a hydrophobic core. There is a significant discrepancy between the weights of the α -helical states in Azochignolin (1.1%) and Chignolin (41%), which can be explained by the most-significant H-bond in Chignolin's α -helix, Asp3-Thr6, which is absent in Azochignolin due to the missing Thr6 (SI Figure S17). Additionally, AMPP's central location and shape likely have a helix-breaking effect on the global structure.

States 1 and 4 are two similar misfolded β -hairpins, with a very small stationary distribution each. Yet state 1 features a relatively high lifetime (846 ns), similar to the native β -hairpin (885 ns), suggesting that state 1 and, to a lesser degree, state 4 function as trapped states. Tyr2 is only able to form a hydrophobic core with Trp8 via edge-to-face stacking, while Pro4 is isolated. The difference in states 1 and 4 lies in a rotated Thr6 backbone dihedral, leading to a slightly skewed shape in state 1 around the turn, allowing the formation of additional H-bonds, particularly between Asp3-Gly7 and Pro4-Thr6, potentially explaining the higher lifetime of state 1. The low stationary distribution of state 1 is likely due to state 1 only being accessible through state 4. These states do not have a structural analogue in AzoChignolin, but are kinetically similar to Azochignolin's trapped state 2.

State 3 is the third-largest state going by the stationary distribution (20.5%) and features a high structural variability, containing mostly disordered structures with a large radius of gyration. It functions as a partial hub, as it connects all states other than the native β -hairpin and state 1. Both structurally and kinetically, it most

resembles the disordered state 4 in Azochignolin, but is much less populated.

State 0 is a rare state which resembles a misfolded β -hairpin, but does not contain any β -sheet content according to DSSP. The center of its turn is shifted to the end, with Thr6 and Gly7 forming the center of this shape, while Pro4 is bent away, leading to the Tyr2-strand pointing outwards with Tyr2 and Pro4 forming an infrequent hydrophobic core. Since the main stabilizing H-bonds are bonds between Glu5-Thr8 and Glu5-Trp9 (the former frequently found in state 3), this state is similarly not possible for Azochignolin.

Curiously, key hydrogen bond interactions of the native β -hairpin between Asp3 and Thr6 and between Gly1 and Gly10 are also found in all other metastable states apart from state 0 and disordered state 3 (SI Figure S17). The key interaction Asp3-Thr8, which is also found in Azochignolin, is only found in hairpin-like states 5, 1, and 4.

Going back to the native β -hairpin state 5, its structure matches the NMR structure 1UAO³ within 1 Å of backbone RMSD. Furthermore, the stabilizing hydrogen bonds of this hairpin state in MeOH are analogous to the bonds found for Chignolin in water, with key hydrogen bonds being formed between residue pairs Asp3-Thr8, Asp3-Gly7, Asp3-Thr6, Asp3-Glu5, Thr6-Thr8, and Gly1-Gly10, despite their different charges compared to Chignolin in water, confirming that both solvents and both protonation states lead to the same β -hairpin, albeit with different kinetics. The main differences to AzoChignolin are the hydrogen bonds involving loop residues Glu5 and Thr6, yet Gly7:N-Asp3:O which would be possible for Azochignolin is also not found in Azochignolin. Since AMPP's shape does not precisely match that of Thr6:Gly7, particularly as its end-to-end distance is slightly smaller (~ 4 Å vs. ~ 4.7 Å), the resulting shape distortion pulls Gly7 slightly more to the turn region in Azochignolin and additionally twists it around, disabling the key contact between Asp3:O and Gly7:N-H, which might be one reason for the comparably smaller β -hairpin population in Azochignolin.

Summarizing the comparison between Chignolin in MeOH and water (with further details provided in the SI, Figures S18–22), we find that the folded β -hairpin exhibits the same structure in both solvents, but that unfolded states are more populated and more diverse in MeOH, with a higher population of α -helical structures. Notably, reptonally-shifted^{39,40} misfolded β -hairpins, also found in prior Chignolin studies,^{6,41,42} are absent in MeOH.

Many of the slower processes described by the eigenvectors for Chignolin are not directly related to the actual β -hairpin folding. Similar to the dynamical fingerprint for Azochignolin, we extracted the most important eigenvectors for Chignolin using an observable a_k related to the amount of β -sheet content to simulate a perturbation-relaxation experiment designed to observe hairpin folding. However, unlike Azochignolin, the starting probability distribution $\mathbf{p}(0)$ was chosen based on all microstates with an average RMSD to the hairpin of more than 7.5 Å, weighted by the stationary distribution. The only eigenvectors with a significant fingerprint amplitude γ_i are vectors 4 and 6 (see Figure 6). Their processes mainly describe transitions between β -hairpin state 5, α -helical state 2 and disordered state 3, with relaxation times of around 559 ns and 364 ns, respectively, which are considerably slower than the relaxation times for Azochignolin (210 ns and 186 ns). Similarly, the MFPT from state 2 to state 5 is also slower for Chignolin (Figure 5A), compared to the MFPT of state 4 to state 3 in Azochignolin (Figure 2A).

3.4 | Comparison of dynamics

While the general folding behavior is similar between AzoChignolin and Chignolin in both solvents, there are differences regarding the population of native, disordered and intermediate states. In water, the β -hairpin population of Chignolin dominates (~80%) the trajectories, while AzoChignolin has a more equally balanced population with only ~60% hairpin content. Simulations in methanol show a large shift away from the hairpin, with only ~30% hairpin content for Chignolin and ~20% for AzoChignolin. In contrast, the α -helical intermediate in Chignolin becomes the largest energy basin, and α -helical structures are even sampled for AzoChignolin, albeit at a very low population, which are unobservable for AzoChignolin in water.

The difference in structural distribution of the different metastable states can be visualized by applying tICAs estimated on one system on another system (see Methods) to make the energy landscapes of two systems comparable, which was done for Azochignolin and Chignolin in MeOH (Figure 7). Since the tICA is applied on the union of their features, this excludes features differing between the systems, particularly distances and dihedrals involving AMPP in Azochignolin and the loop residues Glu5, Thr6 in Chignolin. While this does not have a large effect on tIC1 and tIC2 of Azochignolin (compare Figure 7C, Figure 2), excluding loop features rotates and partially compresses the energy landscape spanned by tIC0 and tIC1 of Chignolin (compare Figures 7A and 5B).

In the new energy landscape created by transforming Azochignolin's data with tIC0 and tIC1 of Chignolin (Figure 7B), tIC0

separates the single β -hairpin state from all others, while tIC1 is unable to separate any macrostate of Azochignolin. The position of the energy minimum of Azochignolin's hairpin is shifted to lower tIC1 values (SI Figure S23) compared to Chignolin's native hairpin minimum.

The landscape of Chignolin, transformed by Azochignolin's tIC1 and tIC2 shows a larger similarity to Azochignolin's energy landscape: tIC1 separates the β -hairpin state 5 and the hairpin misfolds 1 and 4 from other Chignolin metastable states (Figure 7D). Compared to Azochignolin's landscape, these hairpin states are located at the same position as Azochignolin's hairpin state 3. However, the region made up by Azochignolin's helical state 0 (tIC2 values below -3) is completely absent in the landscape of Chignolin, showing that the α -helical structures encountered for Chignolin and Azochignolin are structurally distinct. The same holds true for the absence of Azochignolin's compact state 2 (tIC1 values below -2, also see SI Figure S23).

Since it is difficult to interpret all eigenprocesses of a Markov model and to find an isolated eigenprocess which can either describe the process of turn formation (of the central turn) or the hydrophobic collapse in order to evaluate the associated timescale through their eigenvalue, an approach similar to that in McKiernan et al.³⁰ was applied to our systems:

Separate smaller MSMs, so-called miniMSMs, were estimated on much smaller hand-picked feature sets extracted from the same trajectories as above where each feature set is specifically tailored to the micro-folding event in question. For the hydrophobic collapse, the feature set comprises the individual solvent-accessible surface area contributions of the sidechains of hydrophobic residues Tyr2 and Trp9 as well as their distances (three features only). For the turn formation the feature set comprises the distances between donor-acceptor pairs of key hydrogen bonds near the turn as well as the backbone dihedral angles at the turn (see SI Table S2 for full listing). The feature sets are then transformed with tICA, returning 2 and 3 components, respectively. The lag times of the miniMSMs have been set to 50 ns, adapted from McKiernan et al.³⁰ since a model with a slower lag time would have difficulties capturing the dynamics of the relatively fast collapse and turn formation events. This procedure was repeated for each system.

Due to the sparsity of the feature sets, the resulting MSMs only describe their respective folding dynamics. The largest timescale of each MSM therefore approximates the relaxation time of their folding event. For Azochignolin in MeOH, the timescales of around 88 ns associated with the hydrophobic collapse and 238 ns for the turn formation show a clear separation, with turn formation being more than twice as slow as hydrophobic collapse. This ordering holds true for all other systems investigated (see Figure 8). A similar ordering of timescales was experimentally found for CLN025 as well.⁴³ Further, hydrophobic collapse happens at a faster rate for AzoChignolin (around 70–95 ns faster) compared to Chignolin, and slightly faster in water compared to methanol. Turn formation is much faster for AzoChignolin (around 200–400 ns faster) compared to Chignolin, and is significantly faster in water compared to methanol.

FIGURE 6 Left: Dynamic fingerprint for Chignolin (MeOH). For the observable (A) based on β -sheet content, only eigenvectors 4 and 6 show a significantly high amplitude γ_i . Right: Left eigenvectors 4 and 6 and their associated timescales, projected onto the energy landscape spanned by first two tICs, describing transitions between positive (red) and negative (blue) regions, mainly between states 5 (β -hairpin), 2 (partially α -helical) and 3 (disordered)

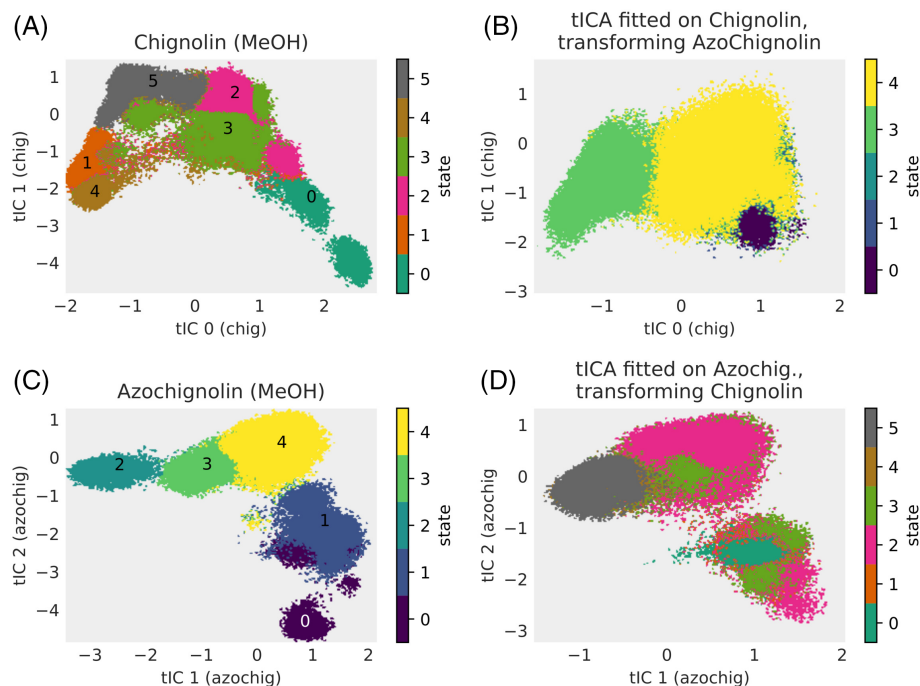
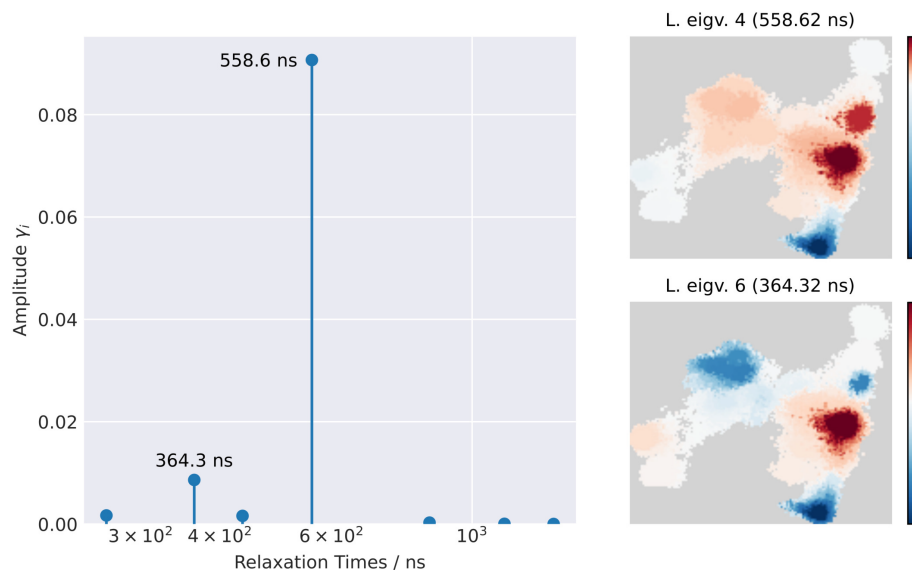


FIGURE 7 Top: The first two tICs, estimated on Chignolin (MeOH), used to transform data of Chignolin itself (A) and of Azochignolin (B). On the right plot, states 0, 1, 2 and 4 are overlapping. Bottom: The components tIC1 and tIC2, estimated on Azochignolin, transforming Azochignolin data (C) and Chignolin data (D). On the right plot, states 5, 1 and 4 are overlapping. Note that Chignolin's β -hairpin states are states 5 (native), 1 and 4 (misfold). Azochignolin's β -hairpin state is state 3

The comparatively high relaxation time for Chignolin in methanol could be related to its strong preference for α -helical structures with entirely different folding dynamics. The much faster turn formation for AzoChignolin compared to Chignolin is likely promoted by the shape of the cis-azobenzene which already exhibits a turn-like structure and makes it easier to form the actual turn.

3.5 | Effects of methanol solvent on folding

The slowest folding processes of the systems of interest have noticeably larger relaxation times in methanol compared to water. Further, the amount of natively folded β -hairpin structures is significantly

reduced, leading to a preference of non-native and disordered structures in methanol. Methanol is known to weaken hydrophobic interactions, but strengthen hydrogen bonds,⁴⁴ leading to the assumption that the weakening of the hydrophobic core Tyr, Trp, and Pro (and AMPP in AzoChignolin) outweighs the stabilization through cross-strand hydrogen bonds. In previous studies^{44–46} it has been shown as well that the addition of alcohols to water can stabilize α -helical structures while suppressing or denaturing others which is in line with our observation where the largest intermediate state of Chignolin in methanol is a mostly α -helical structure while the β -hairpin occurs at a smaller fraction. Helical content is also absent for AzoChignolin in water, but exists for AzoChignolin in methanol, albeit at a very low population due to the helix-breaking azobenzene in the loop.

Other possible reasons for the difference in folding behavior include the residues specifically used for simulations in methanol: The deprotonated residues Glu in Chignolin and Asp in both systems, as well as the neutralized terminals Gly, which were replaced with their charged versions for simulations in water. To determine the effects of the differing residues, additional simulations in water were conducted for Chignolin and AzoChignolin and subsets of these neutral residues, with a cumulative simulation time of 4–8 μ s per system (see SI Table S1).

Notably, Chignolin in water hardly reaches the native hairpin once all residues are replaced by their neutral variants, with hairpin content being even lower than in methanol. Selective residue replacements show that protonation of Glu and neutral terminals both moderately decrease hairpin content. Protonation of Asp seems to slightly increase hairpin content.

A similar picture is painted by AzoChignolin: In water, neutralizing the terminal Gly seems to entirely remove AzoChignolin's ability to fold the hairpin. In contrast, protonation of Asp moderately increases

the amount of hairpin content, more so than it does for Chignolin. Combining both modifications leads to an AzoChignolin which is still able to fold the hairpin, albeit at a reduced rate.

3.6 | MMGBSA stability energies

To investigate and quantify the influence of the protonated residues (Asp and Glu) on the stability of Chignolin and AzoChignolin, the molecular mechanics Generalized Born/Surface Area (MMGBSA) method was applied on the previously simulated structures to calculate the difference in stabilization energies $\Delta\Delta E$. MMGBSA was conducted using the MMPBSA.py software⁴⁷ from the AmberTools package, using the implicit water model $igb = 8$ together with the *mbondi3* radii set.⁴⁸ Since an implicit water solvent was used, only simulations in water were processed by MMGBSA. Entropy calculation was omitted due to the high computational demand and the introduction of large statistical uncertainties. MMGBSA was

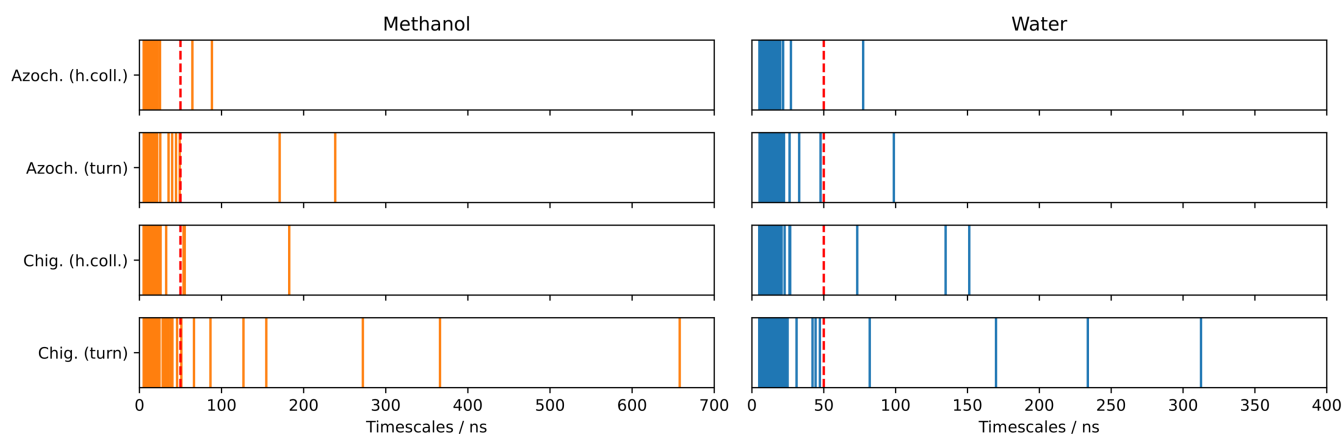


FIGURE 8 Relaxation times describing hydrophobic collapse and turn formation in MeOH (orange) and water (blue) for Azochignolin and Chignolin, calculated by estimating MSMs on respective sparse feature sets. The red dashed line denotes the MSM lag time. The slowest relaxation times can be compared with each other for each system, showing that hydrophobic collapse consistently occurs on a faster timescale than turn formation for all four systems

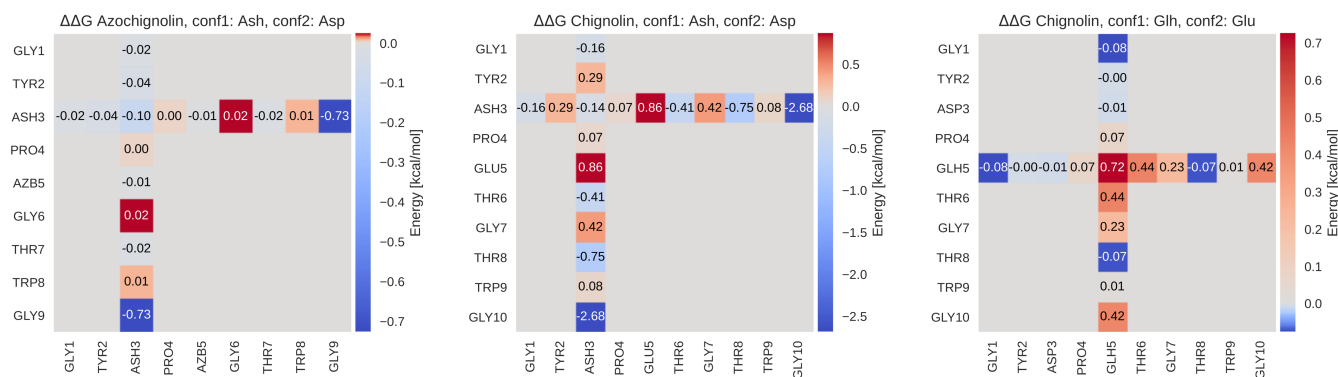


FIGURE 9 Pairwise-decomposed MMGBSA energy differences for Ash/Asp systems (Azochignolin and Chignolin) and Glh/Glu systems (Chignolin). Negative (blue) energy differences denote lower folding energies for protonated Ash or Glh, positive (red) energy differences denote lower folding energies for charged Asp or Glu. Note that the underlying simulations were always conducted within the protonated system (conf1)

performed with a per-residue and a pairwise decomposition scheme on AzoChignolin and Chignolin.

$$\Delta\Delta E = \langle (E_{\text{conf}_1, f} - E_{\text{conf}_1, u}) - (E_{\text{conf}_2, f} - E_{\text{conf}_2, u}) \rangle_{\text{conf}_1}, \quad (1)$$

Similar to alanine scanning, the free simulations for AzoChignolin and Chignolin with protonated Asp (subsequently called Ash), as well as the simulation for Chignolin with protonated Glu (Glh) were used to generate two sets of energies each, respectively, by applying MMGBSA a second time on a modified topology where partial charges for the proton of protonated Ash (or alternatively Glh) were set to zero, effectively creating a topology which is the same as the native AzoChignolin or Chignolin. The pair of energy sets created this way will have the same van der Waals and non-polar solvation contributions as well as largely same internal energies (only direct neighbors of Asp/Glu will have different internal energies due to 1–4 electrostatics). The difference in stability energy is then calculated according to Equation (1), where the suffix conf_1 , for example, refers to the topology with Ash, and conf_2 refers to the topology with Asp. The outer bracket is denoted with conf_1 since the simulations were carried out with the original Ash variant. Subfixes *f* and *u* denote folded and unfolded structures which were filtered using a lower RMSD threshold of 1 Å and an upper threshold of 2 Å to exclude intermediate states. Note that since Azochignolin's Ash variant barely formed any unfolded structures above the upper threshold, additional simulations in the unfolded regime had to be performed to improve statistical uncertainty.

As shown in Figure 9, for Chignolin's Ash/Asp system, the largest energetic pairwise differences are seen for the interaction between Asp3 and the charged C-terminal Gly ($\Delta E_{\text{pol} + \text{eel}} \approx -2.7$ kcal) which comparably stabilize the neutral Ash variant, and a positive interaction (≈ 0.9 kcal) between Asp3 and Glu5, destabilizing the Ash variant compared to Asp. For Chignolin's Glh/Glu system, smaller pairwise differences are mainly seen with Glu5's self-interaction and interaction of Glu5 with Thr6 and the charged C-terminal Gly (between +0.7 and +0.4 kcal), denoting a more favorable stability for anionic Glu versus neutral Glh. AzoChignolin's Ash/Asp system is similar to Chignolin's system, with the largest difference seen for the interaction between Asp3 and C-terminal Gly ($\Delta E_{\text{pol} + \text{eel}} \approx -0.7$ kcal), however without the positive interaction between Asp3 and Glu5, since Glu5 does not exist in AzoChignolin. This difference may explain why AzoChignolin's Ash variant is significantly more stable than Chignolin's Ash variant in water.

4 | CONCLUSIONS

Protein loops and β -hairpins are important elements to nucleate protein folding. Understanding the folding of such motifs including intermediate states and transition kinetics in atomic detail is of major importance for understanding the protein folding process.

Frequently, photoswitchable loop and hairpin peptidomimetics have been used to induce synchronized photo-induced folding to follow the folding process using ultra-fast spectroscopic methods. Atomistic simulations can supplement such studies giving insight into the molecular details of the sampled conformations and transition kinetics. It is also important to compare the folding behavior of a photoswitchable hairpin with its natural counterpart. The photoswitchable AzoChignolin and native Chignolin have been extensively characterized experimentally and thus are excellent model systems for simulations.

Our extensive MD simulations (aggregated simulation time $\approx 50 \mu\text{s}$) and Markov state analysis extends previous studies on the system² and result in folding kinetics for AzoChignolin in good agreement with experiment.³⁸ It has been possible to characterize all relevant intermediates, order of contacts and transition kinetics along the structure formation process including misfolded conformations. The simulations suggests that the AzoChignolin in its *cis* isomer can serve as a photoswitchable variant of the original Chignolin miniprotein forming similar β -hairpin structures. However, there are differences in hairpin population as well as in the processes which form the hairpin. β -hairpin content is lower for AzoChignolin, both in water and in MeOH. The helical intermediates of Chignolin, particularly encountered in MeOH, are also not observable for AzoChignolin in water and barely in MeOH, as its central azobenzene has a helix-breaking effect. With the help of Markov modeling we were able to extract timescales from the slowest eigenprocesses which match previous shorter MD simulations² and can be compared to experiment.³⁸

Both general MSMs and miniMSMs showed that the kinetics of AzoChignolin are generally significantly faster compared to Chignolin, especially for turn formation, likely due to the turn-promoting shape of *cis*-azobenzene, but also for the hydrophobic collapse involving Tyr and Trp. Hence, the very rapid photoswitching induced folding indicates an upper limit for the β -hairpin folding probably not reached by natural β -hairpins. As shown by McKiernan et al³⁰ however, the use of different force fields can also modulate the timescales of microfolding processes. Further work may include a verification of the timescales in other protein and solvent force fields. Our simulation and analysis protocol might also be useful and applicable for systematic studies on other protein folding motifs.

ACKNOWLEDGMENTS

We are grateful to Christina V. Frost for insightful discussions and help with the simulation setup. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, ZA153/28-1). HPC resources provided by the High Performance Computing Center of Friedrich-Alexander Universität Erlangen-Nürnberg (NHR project b118bb) are gratefully acknowledged. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors have no conflicts to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Martin Zacharias  <https://orcid.org/0000-0001-5163-2663>

REFERENCES

- [1] T. Podewin, M. S. Rampp, I. Turkanovic, K. L. Karaghiosoff, W. Zinth, A. Hoffmann-Röder, *Chem. Commun.* **2015**, 51, 4001 ISSN 1359-7345, 1364-548X, <http://xlink.rsc.org/?DOI=C4CC10304A>
- [2] S. M. Hofmann, C. V. Frost, T. Podewin, M. Gailer, E. Weber, M. Zacharias, W. Zinth, A. Hoffmann-Röder, *J. Phys. Chem. B* **2020**, 124, 5113 ISSN 1520-6106, 1520-5207.
- [3] S. Honda, K. Yamasaki, Y. Sawada, H. Morii, *Structure* **2004**, 12, 1507 [https://www.cell.com/structure/abstract/S0969-2126\(04\)00242-4](https://www.cell.com/structure/abstract/S0969-2126(04)00242-4)
- [4] S. Honda, T. Akiba, Y. S. Kato, Y. Sawada, M. Sekijima, M. Ishimura, A. Oishi, H. Watanabe, T. Odahara, K. Harata, *J. Am. Chem. Soc.* **2008**, 130, 15327 ISSN 0002-7863, 1520-5126.
- [5] D. Satoh, K. Shimizu, S. Nakamura, T. Terada, *FEBS Lett.* **2006**, 580, 3422 ISSN 00145793.
- [6] S. Enemark, R. Rajagopalan, *Physical Chemistry Chemical Physics* **2012**, 14, 12442 ISSN 1463-9076, 1463-9084, <http://xlink.rsc.org/?DOI=c2cp40285h>
- [7] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, 11, 3696.
- [8] V. H. Man, X. He, P. Derreumaux, B. Ji, X.-Q. Xie, P. H. Nguyen, J. Wang, *J. Chem. Theory Comput.* **2019**, 15, 1440 ISSN 1549-9618, publisher: American Chemical Society.
- [9] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, 25, 1157 ISSN 1096-987X, eprint.
- [10] J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graphics Model.* **2006**, 25, 247 ISSN 1093-3263, <https://www.sciencedirect.com/science/article/pii/S1093326305001737>
- [11] D. A. Case, H. M. Aktulga, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, G. A. Cisneros, V. W. D. Cruzeiro, T. A. Darden, et al., *Amber 2021*, University of California, San Francisco **2021**.
- [12] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, 79, 926 ISSN 0021-9606, publisher: American Institute of Physics.
- [13] P. Cieplak, J. Caldwell, P. Kollman, *Journal of Computational Chemistry* **2001**, 22, 1048 ISSN 1096-987X, eprint.
- [14] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *The Journal of Chemical Physics* **1984**, 81, 3684 ISSN 0021-9606, publisher: American Institute of Physics.
- [15] C. W. Hopkins, S. Le Grand, R. C. Walker, A. E. Roitberg, *J. Chem. Theory Comput.* **2015**, 11, 1864 ISSN 1549-9618, publisher: American Chemical Society.
- [16] S. Miyamoto, P. A. Kollman, *J. Comput. Chem.* **1992**, 13, 952 ISSN 1096-987X, eprint.
- [17] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, 103, 8577 ISSN 0021-9606, publisher: American Institute of Physics.
- [18] W. Kabsch, C. Sander, *Biopolymers* **1983**, 22, 2577 ISSN 1097-0282, eprint.
- [19] A. Rodriguez, A. Laio, *Science* **2014**, 344, 1492. Publisher: American Association for the Advancement of Science. <https://doi.org/10.1126/science.1242072>
- [20] G. R. Bowman, D. L. Ensign, V. S. Pande, *J. Chem. Theory Comput.* **2010**, 6, 787 ISSN 1549-9618, 1549-9626.
- [21] E. Hruska, J. R. Abella, F. Nüske, L. E. Kavrakci, C. Clementi, *J. Chem. Phys.* **2018**, 149, 244119 ISSN 0021-9606, 1089-7690.
- [22] N. Plattner, S. Doerr, G. De Fabritiis, F. Noé, *Nature Chem.* **2017**, 9, 1005 ISSN 1755-4349. bandiera_abtest: a Cg_type: Nature Research Journals Number: 10 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Molecular dynamics;Single-molecule biophysics Subject_term_id: molecular-dynamics;single-molecule-biophysics. <https://www.nature.com/articles/nchem.2785>
- [23] D. Mendels, G. Piccini, Z. F. Brotzakis, Y. I. Yang, M. Parrinello, *J. Chem. Phys.* **2018**, 149, 194113 ISSN 0021-9606, publisher: American Institute of Physics.
- [24] C. R. Schwantes, V. S. Pande, *J. Chem. Theory Comput.* **2013**, 9, 2000 ISSN 1549-9618, 1549-9626.
- [25] G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé, *J. Chem. Phys.* **2013**, 139, 15102 ISSN 0021-9606, 1089-7690.
- [26] F. Noé, C. Clementi, *J. Chem. Theory Comput.* **2015**, 11, 5002 ISSN 1549-9618, 1549-9626.
- [27] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, F. Noé, *J. Chem. Phys.* **2011**, 134, 174105 ISSN 0021-9606, 1089-7690.
- [28] S. Röblitz, M. Weber, *Adv. Data Anal. Classif.* **2013**, 7, 147 ISSN 1862-5347, 1862-5355.
- [29] F. Noé, H. Wu, J.-H. Prinz, N. Plattner, *J. Chem. Phys.* **2013**, 139, 184114 ISSN 0021-9606, 1089-7690.
- [30] K. A. McKiernan, B. E. Husic, V. S. Pande, *J. Chem. Phys.* **2017**, 147, 104107 ISSN 0021-9606, 1089-7690.
- [31] M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz, F. Noé, *J. Chem. Theory Comput.* **2015**, 11, 5525 ISSN 1549-9618, 1549-9626.
- [32] R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane, V. S. Pande, *Biophys. J.* **2015**, 109, 1528 ISSN 0006-3495, Publisher: Elsevier. [https://www.cell.com/biophysj/abstract/S0006-3495\(15\)00826-7](https://www.cell.com/biophysj/abstract/S0006-3495(15)00826-7)
- [33] D. R. Roe, T. E. Cheatham, *J. Chem. Theory Comput.* **2013**, 9, 3084 ISSN 1549-9618, publisher: American Chemical Society.
- [34] K. Lindorff-Larsen, S. Piana, R. O. Dror, D. E. Shaw, *Science* **2011**, 334, 517 ISSN 0036-8075, 1095-9203.
- [35] L. Bonati, G. Piccini, M. Parrinello, *Proc. Natl Acad. Sci* **2021**, 118, e2113533118 Publisher: Proceedings of the National Academy of Sciences.
- [36] A. Barth, *Biochim. Biophys. Acta (BBA)-Bioenerget* **2007**, 1767, 1073 ISSN 0005-2728, <https://www.sciencedirect.com/science/article/pii/S0005272807001375>
- [37] F. Noe, S. Doose, I. Daidone, M. Lollmann, M. Sauer, J. D. Chodera, J. C. Smith, *Proc. Natl. Acad. Sci.* **2011**, 108, 4822 ISSN 0027-8424, 1091-6490.
- [38] B. G. Keller, J.-H. Prinz, F. Noé, *Chem. Phys.* **2012**, 396, 92. ISSN 0301-0104, <https://www.sciencedirect.com/science/article/pii/S0301010411003892>
- [39] G. Wei, P. Derreumaux, N. Mousseau, *J. Chem. Phys.* **2003**, 119, 6403 ISSN 0021-9606, publisher: American Institute of Physics.
- [40] G. Wei, N. Mousseau, P. Derreumaux, *Proteins* **2004**, 56, 464 ISSN 1097-0134, eprint.
- [41] A. Suenaga, T. Narumi, N. Futatsugi, R. Yanai, Y. Ohno, N. Okimoto, M. Taiji, *Chemistry* **2007**, 2, 591 ISSN 18614728, 1861471X.
- [42] R. Harada, A. Kitao, *J. Phys. Chem. B* **2011**, 115, 8806 ISSN 1520-6106, publisher: American Chemical Society.
- [43] C. M. Davis, S. Xiao, D. P. Raleigh, R. B. Dyer, *J. Am. Chem. Soc.* **2012**, 134, 14476 ISSN 0002-7863, publisher: American Chemical Society.

- [44] S. Hwang, Q. Shao, H. Williams, C. Hilty, Y. Q. Gao, *J. Phys. Chem. B* **2011**, *115*, 6653.
- [45] N. Hirota, K. Mizuno, Y. Goto, *J. Mol. Biol.* **1998**, *275*, 365 ISSN 0022-2836, <https://www.sciencedirect.com/science/article/pii/S0022283697914686>
- [46] M. Buck, *Quart. Rev. Biophys.* **1998**, *31*, 297 ISSN 1469-8994, 0033-5835, publisher: Cambridge University Press. <http://www.cambridge.org/core/journals/quarterly-reviews-of-biophysics/article/trifluoroethanol-and-colleagues-cosolvents-come-of-age-recent-studies-with-peptides-and-proteins/C72805CFAB8922CBBA8FD3246E9D48E>
- [47] B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke, A. E. Roitberg, *J. Chem. Theory Comput.* **2012**, *8*, 3314.
- [48] H. Nguyen, D. R. Roe, C. Simmerling, *J. Chem. Theory Comput.* **2013**, *9*, 2020 ISSN 1549-9618, publisher: American Chemical Society.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: R. L. Zschau, M. Zacharias, *J. Comput. Chem.* **2023**, *44*(9), 988. <https://doi.org/10.1002/jcc.27059>