Structural-Kinetic-Thermodynamic Relationships for Peptide Secondary Structure Formation Identified from Transition Network Properties

Joseph F. Rudzinski[1, a)] and Tristan Bereau[1]
Max Planck Institute for Polymer Research, Mainz 55128, Germany

Molecular simulation models have provided immense, often general, insight into the complex behavior of protein systems. Even for very detailed, e.g., atomistic, models, the generation of quantitatively accurate dynamical properties remains a formidable challenge. This lack of consistent dynamics largely hinders simulation models, especially coarse-grained models, from providing structural interpretations for kinetic experiments. In this work, we perform a detailed investigation into the kinetic properties generated by molecular simulation models for a representative system of the helix-coil transition. Inspired by high-throughput studies that aim to uncover structure-property relationships, we systematically vary force-field parameters and monitor their structural, kinetic, and thermodynamic properties. The focus of our investigation is a simple, native-biased coarse-grained model, which accurately represents the underlying structural ensemble by employing near-atomistic steric and intramolecular backbone interactions. From each set of simulations, we construct a Markov state model to efficiently and systematically assess the system's kinetic properties. We observe a rather large restriction in the topology of the resulting kinetic networks, apparently due to the detailed description of the peptide backbone. When screening across force-field parameters, we find that structurally-accurate models, with respect to a higher-resolution reference, also best reproduce the kinetics. This connection suggests *structural-kinetic relationships* for these models—accurate structural features guarantee consistent kinetics. A comparative study against distinct higher-resolution models highlights the generality of our conclusions. These remarkable relationships hinge upon the physics of the model, which shapes the free-energy landscape and restricts the attainable kinetic properties. While structural features determine the kinetics at a single temperature, the temperature dependence of the structure determines the thermodynamics, i.e., cooperativity, of the transition. Interestingly, a topological metric of the kinetic networks characterizing the degree of randomness of pathways traveling between the helix and coil states *at a single reference temperature* dictates the relative cooperativity of the resulting transition.

I. INTRODUCTION

In recent years, a significant overlap in the timescales accessible to experiments and computer simulations probing the dynamics of individual protein molecules has been achieved.[1,2] This overlap provides an array of unexplored opportunities for a closer interplay between these traditionally disparate approaches, with potential for improving methodologies as well as recovering deeper insight from specific applications. While atomically-detailed molecular dynamics simulations have emerged as the gold standard for the theoretical investigation of microscopic and chemically-specific driving forces for particular, e.g., fast-folding, processes, there remains significant challenges for utilizing these detailed simulations as a general tool for interpreting both ensemble-averaged and kinetic experimental observations. Although many studies have aimed at assessing and improving the static properties of all-atom (AA) force fields,[3,4] as well as extending these models to accurately describe a wider range of systems, e.g., intrinsically disordered proteins,[5] investigations to systematically assess kinetic properties have only recently begun.[6] Additionally, despite ever-increasing computational power, an overwhelming gap remains for exclusively applying such detailed models to investigate the large range of timescales (from ps to hours), thermodynamic or chemical conditions (e.g., denaturant concentrations), as well as system variations (e.g., sequence mutations) commonly explored in experimental studies. Indeed, much simpler models, e.g., ensemble construction methods[7,8] or analytically-solvable polymer models,[9,10] are routinely employed to build structural interpretations of experimental observations.

In between these simple models and AA models, coarse-grained (CG) simulation models for proteins may retain specific structural and chemical properties of the underlying system, while removing extraneous details to save on the computational cost of exploring configuration space. In particular, simple physics-based[11] and native-biased[12] models, have largely shaped the foundation for our current interpretation of the protein folding process.[13] Although these models provide a qualitative picture of the dynamical processes sampled by protein systems, it is typically not expected that such simplistic models can quantitatively reproduce the corresponding kinetic properties. Recent advancements in CG methodologies allow increased chemical detail and accuracy, while retaining the sampling efficiency necessary to address problems intractable for AA models.[14–16] A number of advanced CG protein models have been developed,[17–20] with varying representations, interactions and parametrization philosophies, and have been successfully employed to investigate a range of specific folding and aggregation processes.

The beneficial speed-up of CG models, attained

[a)] Electronic mail: rudzinski@mpip-mainz.mpg.de

through a combination of reduced molecular friction and softer interaction potentials, comes at the cost of obscuring the connection to the true dynamical properties of the underlying system. Unlike many polymer systems, where a homogeneous dynamical rescaling factor is capable of recovering the correct dynamics of the underlying system,[21,22] the rescaling of CG dynamical processes may generally be a complex function of the system's configuration. This lost connection to the true system dynamics represents a severe limitation for CG models, which may not only prevent quantitative prediction of kinetic properties, but may also lead to qualitatively misleading or incorrect interpretations generated from CG simulations. For example, Habibi *et. al.*[23] recently demonstrated that three different CG models provide disparate descriptions of the forced unfolding process of a 110 residue peptide, despite the capability of all models to fold the peptide to the proper native structure. Furthermore, the fidelity of the folding process, with respect to an AA reference simulation, did not correlate with the complexity of the model. Perhaps more troubling, Rudzinski *et. al.*[24] recently demonstrated that even for a tripeptide of alanines, the hierarchy of local dynamical processes along the peptide backbone, i.e., transitions between metastable states on the Ramachandran plot, is qualitatively misrepresented by a transferable, physics-based CG model.

Although it is possible in principle to rescue the dynamics of a CG model via a generalized Langevin formalism, this approach offers a daunting computational and conceptual challenge for complex biological molecules that give rise to hierarchical dynamics, i.e., kinetic processes coupled over various timescales. As an alternative, it may be possible to rescue the kinetic properties generated from a CG simulation via an a posteriori reweighting of simulation data in order to reproduce a set of provided reference observables, e.g., experimental measurements.[24,25] However, the extent to which experimental data can correct for deficiencies in simulation models through a reweighting is obviously limited. Thus, a detailed understanding of the link between given target observables, e.g., structural properties, and the accurate reproduction of relative kinetic quantities is required. Moreover, because there are established methods for systematically addressing the structural accuracy of CG models,[16] identification of robust relationships between structural and kinetic properties provides a practical route for improving the dynamical accuracy of CG models for biomolecular systems.

In this work we investigate relationships between structural and kinetic properties in the helix-coil transition networks generated by various simulation models. As a fundamental process in protein folding, investigation of α-helical secondary-structure formation represents a logical step in understanding the kinetic properties generated by CG protein models. Remarkably, we find that the physics of the model enforces implicit constraints on the free energy landscape, which results in simple connections between structural and kinetic properties—accurate structural features guarantee consistent kinetics. These structural-kinetic relationships ease the interpretation of the CG helix-coil kinetics, justifying the use of a homogeneous speed-up factor. Our approach extends upon previous works that have performed systematic scans of model parameters for relatively simple CG models to obtain insight into the driving forces for emergent condensed-phase properties.[26–28] Conceptually, this strategy mirrors recent efforts at uncovering structure-property relationships by systematic high-throughput studies,[29–31] but focuses on probing property interrelationships by simultaneously monitoring various emergent features of the model. As a primary tool for investigation, we employ a native-biased model, whose parameters may be easily tuned to elucidate the connection between specific interactions, e.g., hydrophobic attraction between side chains, and emergent properties of the resulting kinetic network. However, in order to sample a physically-realistic ensemble of structures (likely a necessary condition for the consistent reproduction of kinetic properties), this model retains a near-atomistic description of backbone steric interactions. To assess the generality of the results, we additionally consider a transferable, physics-based CG peptide model, which also retains near-atomistic backbone resolution but employs phenomenological interactions in order to reproduce the balance of α/β structural propensities in a number of distinct peptide systems.[19,32]

As a model system, we examine an uncapped heptapeptide of alanines, for which reference AA simulations display a complex, disordered ensemble of pathways between the helix and coil states, resulting in low helical content and strong end effects—a challenge for the simple native-biased model. We demonstrate, using standard Lifson-Roig models as a structure-characterization tool, that matching of the Lifson-Roig parameters guarantees the reproduction of certain relative timescales, given a weak structural constraint on the system as well as the implicit constraints provided by the underlying physics of the model. We then characterize the temperature dependence of the various CG models, further validating the structural-kinetic relationships, and demonstrate that constraining simple structural properties at a single temperature is not enough to dictate the cooperativity of the resulting transition. However, an investigation of the detailed properties of the transition network, afforded by Markov state models built directly from the CG simulations, reveals connections between the network topology and the thermodynamic properties of the model. Surprisingly, the average conditional path entropy—a graph measure quantifying the degree of randomness for trajectories traveling between the helix and coil states through particular intermediate states—determined at a single reference temperature, provides sufficient information for determining the relative cooperativity of each CG model.

II. METHODS

A. Coarse-grained (CG) models

1. Hybrid Gō (Hy-Gō)

To investigate the relationship between structural and kinetic properties generated from CG simulation models, we wanted a simple model (i.e., with few, physically-motivated parameters) which sampled all the relevant conformational states. For this purpose, we proposed a flavored-Gō model with three interactions: *(i)* a native contact (nc) attraction, $U_{\rm nc}$, employed between pairs of C_α atoms which lie within a certain distance in the native structure, i.e., the α-helix, of the peptide. *(ii)* a desolvation barrier (db) interaction, $U_{\rm db}$, also employed between native contacts, and *(iii)* a hydrophobic (hp) attraction, $U_{\rm hp}$, employed between all pairs of C_β atoms of the amino acid side chains. $U_{\rm nc}$ and $U_{\rm hp}$ are necessary for sampling the correct ensemble of conformations (i.e., helix, coil, and swollen structures), while $U_{\rm db}$ assists in providing cooperativity in the resulting transitions. We employed the same functional forms as in many previous studies,[33] with a tunable prefactor for each interaction as described below.

These three Gō-type interactions may be viewed as the minimum resolution necessary to *roughly* sample the correct conformational ensemble for short peptides. However, because we are interested in characterizing kinetic properties, it is important that we reproduce the underlying conformational ensemble more precisely. More specifically, we wanted to ensure that the model *(i)* does not sample conformations that are sterically forbidden in the all-atom (AA) model and *(ii)* samples the relevant regions of the Ramachandran plot, while retaining barriers between metastable states. Thus, in addition to the simple interactions described above, the model also partially employs a standard AA force field, AMBER99sb,[34] to model both the steric interactions between all non-hydrogen atoms and also the specific local conformational preferences along the chain. More specifically, the bond, angle, dihedral, and 1–4 interactions of the AA force-field are employed without adjustment. To incorporate generic steric effects, without including specific attractive interactions, we constructed Weeks-Chandler-Andersen potentials (i.e., purely repulsive potentials) directly from the Lennard-Jones parameters of each pair of atom types in the AA model. For simplicity of implementation, we then fit each of these potentials to an r^{-12} functional form. The van der Waals attractions and all electrostatic interactions in the AA force field were not included and water molecules were not explicitly represented.

Although the local sterics of the backbone are modeled with near-atomistic resolution, the Hy-Gō model substitutes the full atomistic description of the peptide with a highly coarse-grained representation by modeling the complex combination of dispersion and electrostatic peptide-peptide and peptide-solvent interactions with a limited set of simple interactions between C_α and C_β atoms. The total interaction potential for the model may be written: $U_{\rm tot} = \epsilon_{\rm nc}U_{\rm nc} + \epsilon_{\rm db}U_{\rm db} + \epsilon_{\rm hp}U_{\rm hp} + \epsilon_{\rm bb}U_{\rm bb}$, where the backbone (bb) interaction includes both the intramolecular and steric interactions determined from the AA force field. The first three coefficients represent the only free parameters of the model, while $\epsilon_{\rm bb} = 1$. The relative impact of of the backbone interactions may then be characterized by $E_{\rm tot} \equiv \epsilon_{\rm nc} + \epsilon_{\rm db} + \epsilon_{\rm hp}$.

2. PLUM

As an alternative CG model, we considered the PLUM model, which also describes the protein backbone with near-atomistic resolution, while representing each amino acid side chain with a single CG site, within an implicit water environment.[19] In PLUM, the parametrization of local interactions (e.g., sterics) aimed at a qualitative description of Ramachandran maps, while longer-range interactions—hydrogen bond and hydrophobic—aimed at reproducing the folding of a three-helix bundle, without explicit bias toward the native structure.[19] The model is transferable in that it aims at describing the essential features of a variety of amino-acid sequences, rather than an accurate reproduction of any specific one. After parametrization, it was demonstrated that the PLUM model folds several helical peptides,[19,35–38] stabilizes β-sheet structures,[19,32,39–41] and is useful for probing the conformational variability of intrinsically disordered proteins.[42]

B. Simulation Details

1. AA

We employed AA simulations of an uncapped heptamer of alanine residues (Ala_7), previously published by Stock and coworkers.[43] In short, these simulations employed the GROMOS 45A3 force field[44] along with the SPC water model[45] to obtain an 800 ns trajectory, sampled every 1 ps, for Ala_7 in the zwitterionic state at 300 K.

2. Hy-Gō

CG molecular dynamics simulations of Ala_7 with the Hy-Gō model were performed with the Gromacs 4.5.3 simulation suite[46] in the constant NVT ensemble, while employing the stochastic dynamics algorithm with a friction coefficient $\gamma = (2.0\ \mathcal{T}^{\rm S})^{-1}$ and a time step of $1\times10^{-3}\ \mathcal{T}^{\rm S}$. For each model, 20 independent simulations were performed with starting conformations varying from full helix to full coil. Each simulation was performed for $100,000\ \mathcal{T}^{\rm S}$, recording the system every $0.5\ \mathcal{T}^{\rm S}$. The CG

unit of time, $\mathcal{T}^{\mathrm{S}}$, can be determined from the fundamental units of length, mass, and energy of the simulation model, but does not provide any meaningful description of the dynamical processes generated by the model. In this case, $\mathcal{T}^{\mathrm{S}} = 1$ ps.

3. PLUM

CG simulations of Ala_7 with the PLUM force field[19,37,47] were run using the ESPResSo simulation package.[48] For details of the force field, implementation, and simulation parameters, see Bereau and Deserno.[19] A single canonical simulation at temperature $k_{\mathrm{B}}T = 1.0\ \mathcal{E}$ was performed for $200,000\ \mathcal{T}^{\mathrm{P}}$ with at timestep of $0.01\ \mathcal{T}^{\mathrm{P}}$, recording the system every $0.5\ \mathcal{T}^{\mathrm{P}}$, where $\mathcal{T}^{\mathrm{P}} \sim 0.1$ ps. Temperature control was ensured by means of a Langevin thermostat with friction coefficient $\gamma = (1.0\ \mathcal{T}^{\mathrm{P}})^{-1}$.

C. Markov State Models

Markov state models (MSMs) are kinetic models that characterize the probability of transitioning between a finite set of microstates, chosen to represent the configuration space of the underlying system.[49–51] The transition probabilities may be estimated directly from molecular dynamics trajectories via a Bayesian scheme to enforce relevant physical constraints. In the present work, MSMs are built from the Ala_7 helix-coil trajectories generated from each simulation model. To determine the MSM microstate representation, Principle Component Analysis was performed on the configurational space characterized by the ϕ/ψ dihedral angles of each residue along the peptide backbone. A density clustering algorithm was then applied to the five "most significant" dimensions in order to determine the number and placement of microstates, following previous investigation of the AA trajectory.[52] This procedure yields 32 states in all cases, corresponding to the enumeration of all possible helical/coil state combinations for each of the 5 peptide bonds. As described further in the Supporting Information section, the helical (h) and coil (c) states for individual peptide bonds roughly correspond to the α and β regions, respectively, of the corresponding Ramachandran plot. MSM construction and analysis was performed using the pyEmma package.[53] (See Supporting Information for more details and MSM validation.) We also considered lower resolution MSMs, characterized by the number of helical residues in a conformation, Q_{h}. These MSMs were constructed from the "full" MSM via a direct mapping of the transition probability matrix. Such a construction compromises the quantitative description of kinetics. However, we found that this reduced resolution provides a useful tool for assessing the differences between various kinetic descriptions of Ala_7.

III. RESULTS AND DISCUSSION

In this study, we investigate the relationship between structural and kinetic properties of helix-coil transition networks generated by microscopic simulation models. The equilibrium statistical mechanics of helix-coil transitions is well-characterized by a 1D-Ising model, which represents the state of each residue as being either helical, h, or coil, c.[54] These simple equilibrium models employ two parameters, w and v, according to the Lifson-Roig (LR) formulation,[55] which are related to the free energy of helix propagation and nucleation, respectively. These parameters may be determined directly from simulation data using a Bayesian approach,[56] and describe the overarching structural characteristics of the underlying ensemble (see Supporting Information for more details). Perhaps the most important quantity, the average fraction of helical segments, $\langle f_{\mathrm{h}} \rangle$, i.e., propensity of sequential triplets of h states along the peptide chain, may be measured directly from calorimetry experiments. The average number of helical residues, regardless of the particular sequence of states along the peptide chain, denoted $\langle N_{\mathrm{h}} \rangle$, provides a complimentary observable to $\langle f_{\mathrm{h}} \rangle$. Although residue- or sequence-specific LR parameters may be determined in order to more faithfully reproduce the helix-coil properties generated from a simulation of a particular peptide system, in this work we determine a single set of $\{w, v\}$ for each simulation model. These simple models are incapable of describing certain features, e.g., end effects, of the underlying systems; however, we utilize the LR parameters only as a characterization tool and demonstrate that the sequence-independent parameters are sufficient to effectively distinguish between various characteristics of the underlying simulations.

The kinetics of helix-coil transitions are often interpreted in terms of a kinetic extension of the Ising model.[57,58] These kinetic models typically assume a simple relationship between the LR parameters and the on/off rate of helix formation and have been widely successful in describing the emergent kinetic properties, e.g., folding/unfolding rates, observed in experiments.[58,59] However, the kinetics of the helix-coil transition may demonstrate drastic divergences from this behavior, e.g., when misfolded intermediates complicate the network of transition pathways.[60] Even in simpler cases, the precise impact of the model's assumptions on the fine details of the resulting kinetic network is not well understood. Here, we construct kinetic models directly from the simulation trajectories, i.e., Markov state models, allowing both a more complex relationship between the LR parameters and the kinetic properties of the system and also providing a direct assessment of the assumptions made by the more approximate kinetic models.

As a model system, we consider an uncapped heptapeptide of alanine residues (Ala_7), i.e., 5 peptide bonds. From the LR point of view, there are 32 states, determined by enumerating the various sequences of h's and c's. We employ an all-atom (AA) simulation, previously

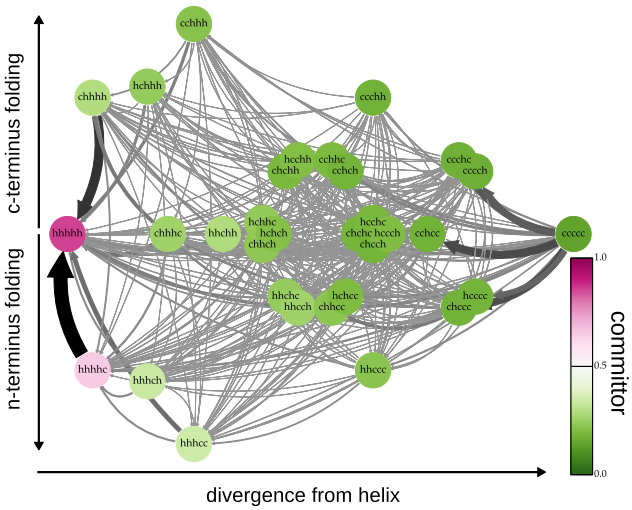

FIG. 1. Network representation for the Ala_7 helix-coil transition generated from an atomically-detailed simulation. The structural ensemble is representation by 32 microstates, corresponding to each possible sequence of the 5 peptide bonds attaining either a helical, h, or coil, c, state. The horizontal axis characterizes the difference between the sequence of each state and the native (full helix) state, while the vertical axis quantifies the extent to which the state corresponds to n- or c-terminus folding. The thickness of the arrows represents the fractional probability flux passing through a pair of states for trajectories which begin in the coil state and end in the helix state. The color of each state denotes the value of the committor—the probability of reaching the helix state before returning to the coil state.

studied by Stock and coworkers,[43,52,61] as a reference in order to assess to what extent CG models are capable of reproducing the kinetic properties of this more detailed model. We note that it is well-known that distinct AA force fields yield widely varying results, e.g., in terms of helical propensities, for short peptide systems.[56] The AA simulation employed in this work serves as an ideal reference since it samples a largely disordered ensemble with a diverse collection of pathways from the coil to helix state, representing a challenge for the CG models considered. Fig. 1 presents a network representation of the AA simulation, plotted along simple reaction coordinates that characterize the helicity of each conformation on the horizontal axis and the direction (c- or n-terminus) of folding on the vertical axis. The thickness of the arrows denote the relative probability flux passing between pairs of states for trajectories starting from the coil state and ending at the helix state. The "middle" of the network appears quite random, with more directed transitions towards the ends of the graph. Interestingly though, the residue dynamics remain highly coupled (see Fig. S4). The color of each state denotes its committor value—the probability of reaching the helix state before returning to the coil state. Clearly the landscape is largely tilted toward the unfolded region, as the committor values are very close to 0 for all but a few states.

A. Identifying structural-kinetic relationships at a single temperature using the Hy-Gō model

The focus of our study is a relatively simple, native-biased CG model. The hybrid Gō (Hy-Gō) model, is a flavored-Gō model, with native contact (nc) interactions (i.e., hydrogen-bonding-like interactions between $i/i+4$ residues) and associated desolvation barriers (db) between C_α atoms, as well as generic hydrophobic (hp) attractions between all pairs of C_β atoms. The relative strengths of these interactions are determined by the set of 3 model parameters: $\{\epsilon_{\rm nc}, \epsilon_{\rm db}, \epsilon_{\rm hp}\}$. Fig. 2 presents a visualization of the model representation and also the corresponding interaction potentials. The model also employs physics-based interactions in the form of sterics (see bottom-right panel of Fig. 2) and torsional preferences along the backbone. We denote this model "hybrid" since it employs both traditional Gō-type interactions as well as detailed physics-based interactions determined from an AA model. Further details of the model and corresponding simulations are described in the Methods section.

In order to assess the capability of the Hy-Gō model to reproduce structural and kinetic properties of the AA model, we considered various combinations of $\epsilon_{\rm nc}$, $\epsilon_{\rm db}$,

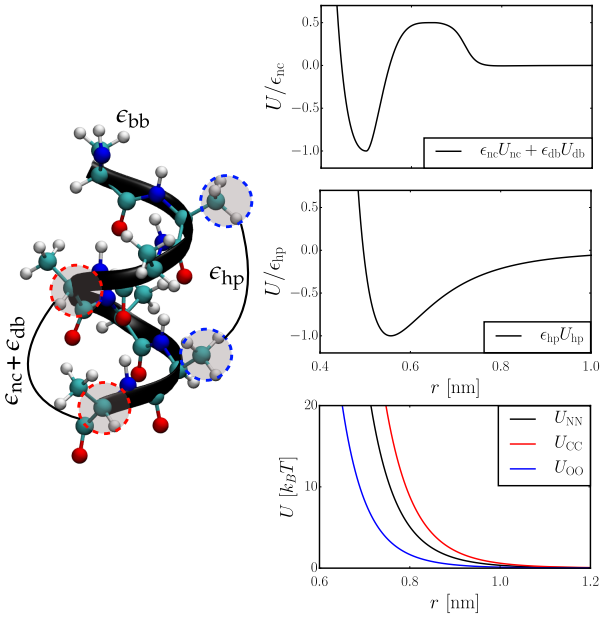

FIG. 2. A visualization of the Hy-Gō model representation and interactions for Ala_7. (Left) Illustration of a native contact between C_α atoms and a generic contact between C_β atoms, along with the corresponding parameters, $\{\epsilon_{\rm nc}, \epsilon_{\rm db}, \epsilon_{\rm hp}\}$, associated with these interactions. (Right) The top two panels present the interaction potentials for the Gō-type interactions as a function of the model parameters. In the top panel, $\epsilon_{\rm db} = 0.5\epsilon_{\rm nc}$. The bottom panel presents the Weeks-Chandler-Andersen-like potentials employed to model sterics along the peptide backbone.

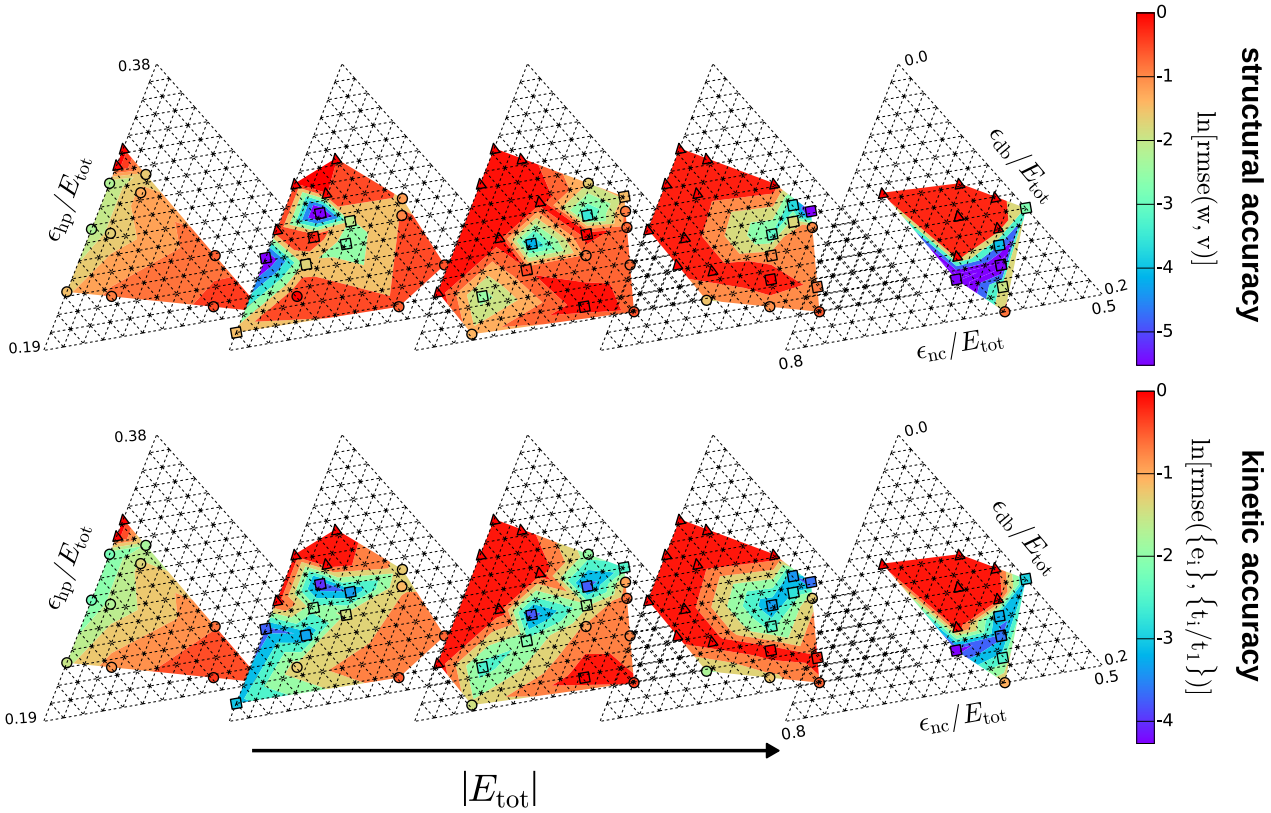

FIG. 3. "Parameter landscape" for the Hy-Gō model. The axes of the ternary plots characterize the relative importance of each of the three Gō-type interactions: $\epsilon_i/E_{\rm tot}$, where $E_{\rm tot} \equiv \epsilon_{\rm nc}+\epsilon_{\rm db}+\epsilon_{\rm hp}$. Because the relative impact of the backbone interactions depends on $E_{\rm tot}$, the plot is discretized along $E_{\rm tot}$. Squares indicate models that reproduced $\langle N_{\rm h}^{\rm AA}\rangle$ at some temperature T^*, circles indicate models similar to squares but whose native state is the coil state, and triangles indicate models with a helix native state but that could not achieve $\langle N_{\rm h}^{\rm AA}\rangle$ at any temperature. The color of each marker denotes the error with respect to some property of the AA model (cooler colors represent lower errors). Note that the filled color grid between points is included only to guide the eye. (Row 1) Root mean square error (rmse) with respect to the LR parameters of the AA model. (Row 2) rmse with respect to the slowest 2 dynamical processes of the system (as characterized by the corresponding eigenvectors of the MSM at the $Q_{\rm h}$ level of resolution) and also the ratio of timescales between these two processes. Because the errors of the models marked with triangles were considerably larger than the other models, their error values were reduced to the maximum of the other models for easier visual comparison.

and $\epsilon_{\rm hp}$ and then adjusted the temperature to reproduce the average helicity of the AA model, $\langle N_{\rm h}^{\rm AA}\rangle$. We found three classes of CG models: *(i)* those that reproduced $\langle N_{\rm h}^{\rm AA}\rangle$ at some temperature T^*, *(ii)* same as *(i)*, but whose native state was the coil state, and *(iii)* those where the helix state was native but the model could not achieve such a low value of helicity at any temperature. For each model, we determined the two LR parameters, $\{w, v\}$, and constructed a 32-state Markov state model (MSM) directly from the simulation data.

Fig. 3 presents a "parameter landscape" for the Hy-Gō model, plotted as a ternary diagram with each axis characterizing the relative importance of the three Gō-type interactions: $\epsilon_i/E_{\rm tot}$, where $E_{\rm tot} \equiv \epsilon_{\rm nc}+\epsilon_{\rm db}+\epsilon_{\rm hp}$. Because the relative impact of the backbone interactions depends on $E_{\rm tot}$, we discretize along $E_{\rm tot}$ and plot ternary diagrams for distinct values along this coordinate. Squares indicate models of type *(i)*, circles of type *(ii)*, and triangles of type *(iii)*. The color of each model denotes the error with respect to some property of the AA model (cooler colors represent lower errors). The first row characterizes the root mean square error (rmse) with respect to the LR parameters of the AA model, while the second row quantifies the rmse with respect to the slowest 2 dynamical processes of the system (as characterized by the corresponding eigenvectors of the MSM at the $Q_{\rm h}$ level of resolution, see the Methods section for details) and also the ratio of timescales between these two processes. There is a clear correspondence between models that reproduce the structural and kinetic metrics. In other words, given the constraint of $\langle N_{\rm h}\rangle$, approximately fixing the LR parameters is enough to also accurately re-

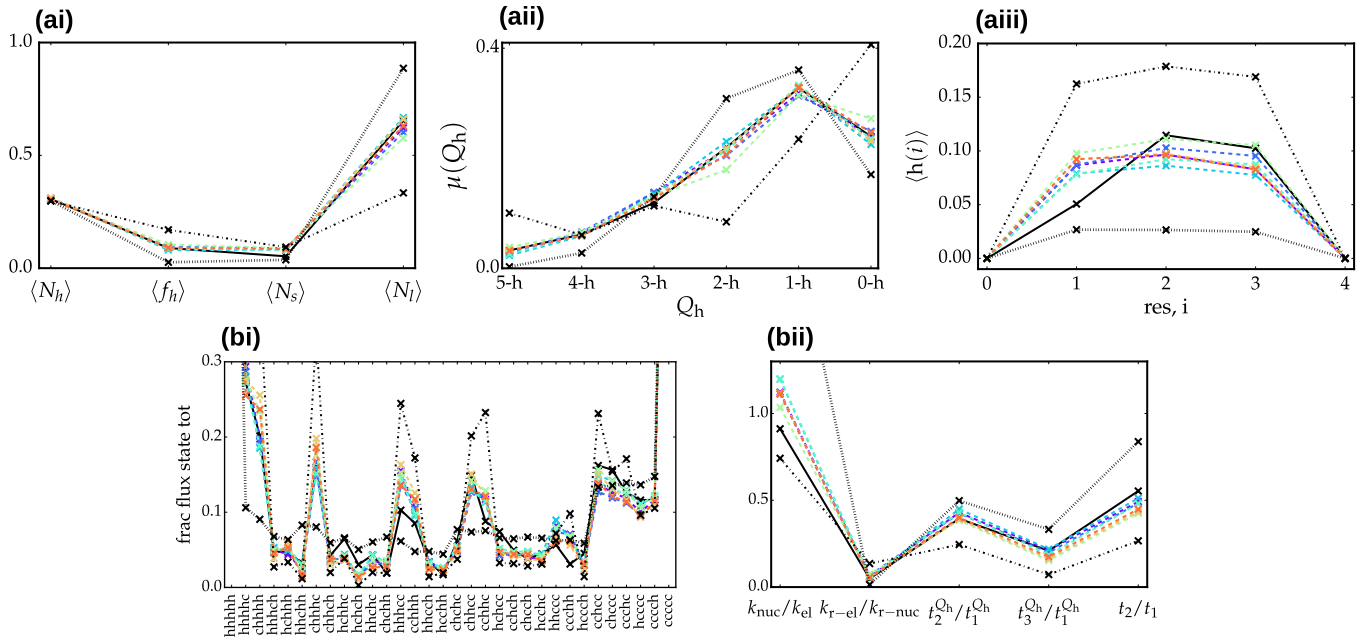

FIG. 4. Static (row *a*) and kinetic (row *b*) properties for the Ala_7 helix-coil transition generated by an all-atom model (AA; solid, black curves) and various Hy-Gō models. 7 AA-like Hy-Gō models (colored curves) were identified as optimally reproducing the properties of the AA model. The dashed and dashed-dotted black curves denote two Hy-Gō models that reproduce $\langle N_h^{AA} \rangle$ at the highest and lowest temperatures, respectively, of all models considered. *(ai)* The average propensity of the h state, $\langle N_h \rangle$, average fraction of helical segments, $\langle f_h \rangle$, average fraction of sequential pairs of helical residues, $\langle N_s \rangle$, and average fraction of isolated helical residues, $\langle N_l \rangle$. *(aii)* The equilibrium distribution along the number of helical residues in the sequence, Q_h. *(aiii)* The average fraction of helical segments per residue, $\langle h(i) \rangle$. *(bi)* The fractional flux passing through each state when considering pathways which begin in the coil state and end in the helix state. *(bii)* Ratios of important timescales. k_{nuc}/k_{el} ($k_{r\text{-}nuc}/k_{r\text{-}el}$) is the ratio of (reverse) nucleation to (reverse) elongation rate. t_i/t_1 denotes the ratio of the ith slowest kinetic process to the slowest kinetic process according to the MSM. The Q_h superscript denotes timescales calculated at the resolution of the number of h residues in the sequence.

produce the overarching hierarchy of kinetic processes of the underlying system at T^*. In particular, 7 "AA-like" Hy-Gō models were identified as reproducing the properties of the AA transition with the highest accuracy. It is important to note that while previous studies have connected the LR parameters to long-timescale features of the helix-coil transition,[58,59] the generation of consistent helix-coil properties from a microscopic simulation model is non-trivial. This is perhaps best demonstrated by the difficulties of AA models to reproduce the melting curve for relatively simple peptides.[5,56]

Fig. 4 presents both static and kinetic properties generated from the AA model (solid, black curves), the 7 AA-like Hy-Gō models (colored curves), and for two additional Hy-Gō models that also reproduce $\langle N_h^{AA} \rangle$ (while still sampling all 32 states), but do so at the largest and smallest simulated temperatures (dashed and dashed-dotted, black curves, respectively). The properties of the latter 2 models represent the range of attainable values of each observable, given the weak structural constraint of $\langle N_h^{AA} \rangle$ and also the implicit constraints enforced by the details of the model, but allowing for distinct LR parameters and, thus, $\langle f_h \rangle$ values. Panel *(ai)* presents $\langle N_h \rangle$ and $\langle f_h \rangle$ (described above), as well as the average fraction of neighboring pairs of helical residues, $\langle N_s \rangle$, and the average fraction of isolated helical residues, $\langle N_l \rangle$. Panel *(aii)* presents the equilibrium distribution along the number of helical residues in the sequence, Q_h. Panel *(aiii)* presents the average fraction of helical segments per residue, $\langle h(i) \rangle$. Panels *(ai)* and *(aii)* demonstrate that the AA-like Hy-Gō models reproduce the static properties of the AA model quite well. As expected, panel *(aiii)* shows that these models are incapable of perfectly reproducing the strong end effects of the underlying ensemble.

Panel *(bi)* presents the fractional flux passing through each state for trajectories that begin in the coil state and end in the helix state. Panel *(bii)* presents ratios of important timescales in the underlying simulations. It is important to note that, since the CG unit of time does not correspond to a physical time, there is always an arbitrary speed-up associated with dynamical properties generated by CG models. For this reason, we only consider ratios of timescales, to effectively account for this speed-up. k_{nuc} ($k_{r\text{-}nuc}$) and k_{el} ($k_{r\text{-}el}$) characterize the (reverse) rates of nucleation and elongation, respectively (see Supporting Information for details about the calculation of

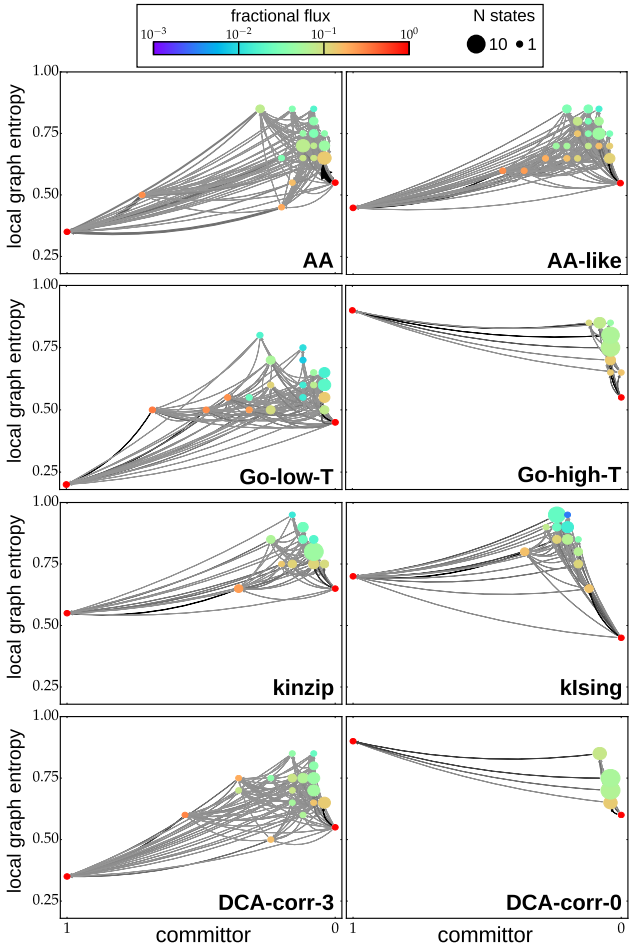

FIG. 5. Network representations of the Ala_7 helix-coil transition generated by an AA model, several CG models, and various approximations constructed from the AA simulation. (See main text for descriptions of the CG models and approximations). Each network is presented along the committor (i.e., the probability of reaching the helix state before returning to the coil state) on the horizontal axis and the local graph entropy, $S^{\rm loc}$, on the vertical axis. The local graph entropy is normalized by the maximum possible value (i.e., for a uniform distribution of transition probabilities). For visual clarity, we discretized the networks along these two coordinates and grouped states sharing the same grid into a single node. The relative size of each node denotes the number of underlying states. The color of each node corresponds to the average fractional flux passing through the node for trajectories from coil to helix.

these rates). t_i denotes the timescale of the $i^{\rm th}$ slowest kinetic process according to the MSM. The $Q_{\rm h}$ superscript denotes timescales calculated from a reduced MSM with microstates corresponding the number of h residues in the sequence. Despite the detailed structural deficiencies of these models, both the fractional flux per state and the various timescale ratios are reproduced with surprisingly high accuracy. Interestingly, the approximate bounds on the timescale ratios, represented by the black dashed and dashed-dotted curves, demonstrate that there is a rather small amount of freedom in these quantities (other than for $k_{\rm nuc}/k_{\rm el}$). We suggest that this is due to the implicit constraints on the kinetic network, enforced by the underlying physics of the model.

Fig. 5 presents network representations of the AA model and a representative AA-like Hy-Gō model (top panel) as well as the low- and high-temperature Hy-Gō models introduced above. The networks are presented in a different way than Fig. 1, with the horizontal axis corresponding to the value of the committor for each state and the vertical axis corresponding to the local graph entropy: $S^{\rm loc}(i) = \sum_j T_{ij} \ln T_{ij}$, where T_{ij} is the transition probability from state i to state j. $S^{\rm loc}(i)$ quantifies the degree of randomness for pair transitions originating in state i. For visual clarity, we discretized the networks along these two coordinates and grouped states sharing the same grid into a single node. The relative size of each node denotes the number of underlying states. The color of each node corresponds to the average fractional flux passing through the node for trajectories starting from the coil state and ending in the helix state. The networks in Fig. 5 provide an illustration of the network topologies which generate the various properties presented in Fig. 4. The AA-like model appears to mimic the AA network in many ways, although the spread of states along each axis is not quite reproduced. On the other hand, the low- and high-temperature networks display drastically different topologies, albeit while retaining the average helicity of the AA model, $\langle N_{\rm h}^{\rm AA} \rangle$. Thus, $\langle N_{\rm h} \rangle$ is indeed a weak structural constraint for this system.

B. Simple kinetic approximations

As already mentioned above, the emergent kinetics (i.e., long-timescale features) of the helix-coil transition are well characterized by simple kinetic models constructed directly from the LR parameters.[57,58] Here, we briefly assess the resulting properties of these approximate models for comparison with the CG simulation models. We consider two models built directly from the LR parameters, denoted kinzip and kIsing. The kinzip model corresponds to the kinetic zipper model,[58] which assumes that the rate of transition from the c to h state for each residue is limited by the nucleation parameter, v. The kIsing model is based on the formulation of Schwarz,[57] which uses the LR parameters to define the reaction rates for individual triplets of residues transitioning between conformational states. We also employed a different approach for constructing approximate models for the simulation kinetics, following the n-m approximation.[62] In this procedure, we start with a "local" kinetic model determined from the simulation data, i.e., by ignoring the state of residues which lie beyond some number of peptide bonds away from a given residue, and then construct a full MSM, assuming that the residue dynamics beyond the chosen separation are decoupled. In this way, a systematic investigation into the coupling

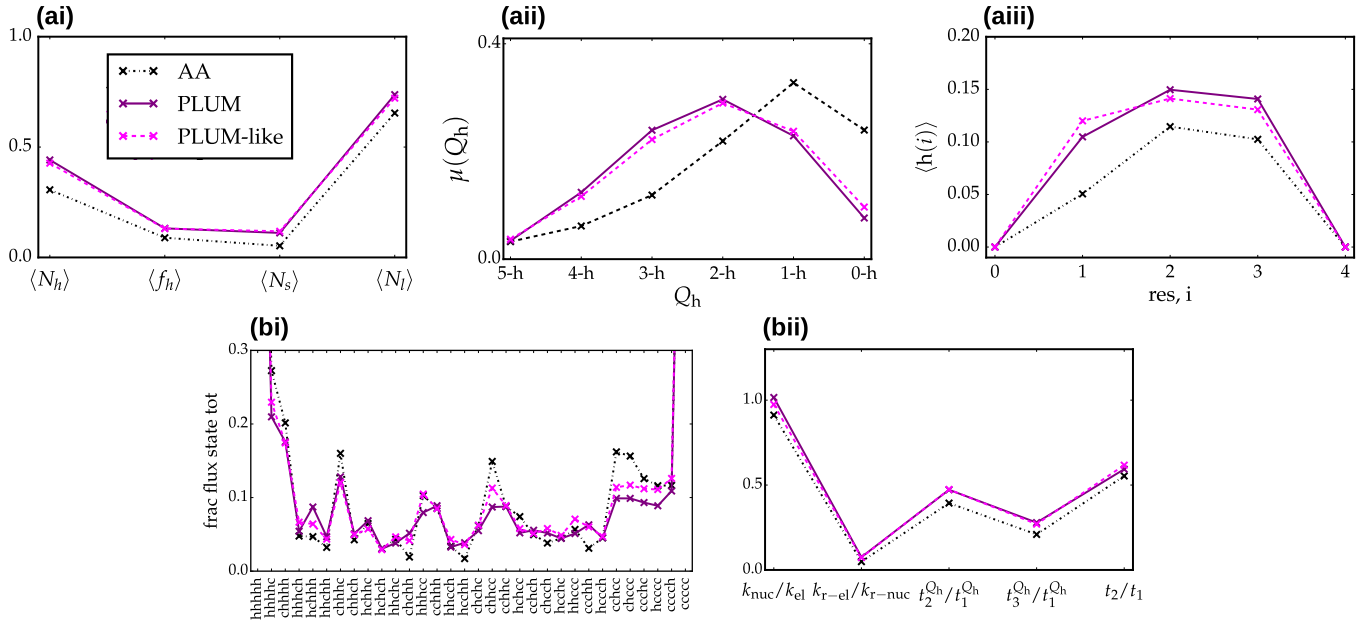

FIG. 6. Static (row *a*) and dynamical (row *b*) properties for the Ala_7 helix-coil transition generated by an all-atom model (AA; solid, black curves), the original PLUM model (PLUM; solid, violet curves), and a PLUM-like Hy-Gō model (PLUM-like; dashed, magenta curves). *(ai)* The average propensity of the h state, $\langle N_h \rangle$, average fraction of helical segments, $\langle f_h \rangle$, average fraction of pairs of helical residues, $\langle N_s \rangle$, and average fraction of isolated helical residues, $\langle N_l \rangle$. *(aii)* The equilibrium distribution along the number of helical residues in the sequence, Q_h. *(aiii)* The average fraction of helical segments per residue, $\langle h(i) \rangle$. *(bi)* The fractional flux passing through each state when considering pathways which begin in the coil state and end in the helix state. *(bii)* Ratios of important timescales. k_{nuc}/k_{el} ($k_{r\text{-}nuc}/k_{r\text{-}el}$) is the ratio of (reverse) nucleation to (reverse) elongation rate. t_i/t_1 denotes the ratio of the ith slowest kinetic process to the slowest kinetic process according to the MSM. The Q_h superscript denotes timescales calculated at the resolution of the number of h residues in the sequence.

of residue dynamics along the peptide backbone can be performed. We refer to this procedure as the dynamic coupling analysis (DCA), and the corresponding models are denoted DCA-corr-x, where x is the chosen number of peptide bonds to be correlated in the model. See the Supporting Information section for details about the construction of each of these models.

Overall, the approximate kinetic models quite accurately reproduce both the static and kinetic properties of the underlying, AA model (Fig. S4). A full decoupling of residue dynamics (DCA-corr-0) is required to introduce large errors, yielding a spread of flux throughout the network and the systematic oversampling of intermediate states, as previously reported for bottom-up CG models for the helix-coil transition.[63] Smaller discrepancies in the other models can be understood in terms of the strong end effects in this short, uncapped peptide. Fig. 5 presents network representations of the kinzip and kIsing models, and two approximate models constructed by systematic decoupling of the residue dynamics, assuming no correlations and $i/i+3$ correlations (DCA-corr-0 and DCA-corr-3, respectively). The kinzip and kIsing networks seem to reproduce the AA properties via a simplified network topology, while DCA-corr-3 retains a closer resemblance to the AA network. The DCA-corr-0 network displays drastically different topological features, as is apparent in the resulting properties of this model. Decoupling of residue dynamics leads to a decrease in committor values and an increase in local entropies, especially for the largely helical states. Interestingly, many properties of the underlying network (as demonstrated in Figs. 4 and S4), may be reproduced by quite distinct network topologies. Further investigation into the properties generated from these simple kinetic approximations may provide insight into how the details of the kinetic network are affected by the particular approximations made within the model construction. This is beyond the scope of the present work.

C. Further validation of structural-kinetic relationships using a transferable CG model

We also considered the transferable CG PLUM model with three distinct parametrizations. Results for the original model are presented in the main text, while equivalent results for the two reparametrizations are presented in the Supporting Information. The static and kinetic properties of the PLUM model for Ala_7 are presented in Fig. 6 as solid violet curves. The PLUM model strongly stabilizes helical structures for Ala_7, and is incapable of achieving the low helical content of the AA

model. The significant discrepancies between the AA (black, dashed-dotted curves) and PLUM model are not surprising since *(i)* neither the PLUM nor the AA model was parametrized to reproduce properties of small peptides and *(ii)* various AA models yield widely varying structural properties for peptide systems.[56] Rather, in this study the AA and PLUM models represent two distinct reference ensembles for investigating the interplay between structure and kinetics. We extended the search in Hy-Gō parameter space to find a set of models which more closely reproduce the structural and kinetic properties of the PLUM model. Fig. 6 demonstrates that by approximately matching $\langle N_{\mathrm{h}} \rangle$ and $\langle f_{\mathrm{h}} \rangle$, the "PLUM-like" Hy-Gō model nearly quantitatively reproduces both the equilibrium and kinetic properties of the PLUM model. The network topology of the PLUM-like Hy-Gō model also closely resembles that of the PLUM model (Fig. S7).

D. Thermodynamics of Ala_7 helix-coil transitions

The temperature dependence of the helicity, in particular $\langle f_{\mathrm{h}} \rangle(T)$, is a fundamental quantity for characterizing secondary structure formation in proteins. Fig. 7 presents the T-dependence of w (top panel) and $\langle f_{\mathrm{h}} \rangle$ (bottom panel) for the AA-like Hy-Gō models (purple, circle markers), other Hy-Gō models which reproduce $\langle N_{\mathrm{h}}^{\mathrm{AA}} \rangle$ at some temperature T^* (blue, upward triangle markers), Hy-Gō models incapable of reproducing $\langle N_{\mathrm{h}}^{\mathrm{AA}} \rangle$ (green, sideways triangle markers), the PLUM-like Hy-Gō models (red, square markers), and the PLUM model (black, X markers). The energy scales of the different models were aligned by shifting the temperature such that all models achieve $\langle N_{\mathrm{h}}^{\mathrm{AA}} \rangle$ at T^*. For models not simulated at this temperature, a linear extrapolation is employed.

From the temperature dependence of w, we fit a thermodynamic model to quantify the T-independent enthalpy and entropy of helix extension. In particular, we assume $k_{\mathrm{B}}T \ln w(T) \sim \Delta H_{\mathrm{hb}} - T\Delta S_{\mathrm{hb}}$, neglecting the heat capacity contributions to the free energy and considering a relatively small range in T. ΔH_{hb} corresponds to the slope of the curves in Fig. 7*(a)* and is a simple measure of the cooperativity of the transition. The resulting thermodynamic values for each model are given in Table S1 and are consistent (in terms of order of magnitude) with those determined from various AA models for a longer peptide,[56] although the entropic values are somewhat larger. It is straightforward to correlate these thermodynamic quantities with the parameters of the Hy-Gō model. This analysis confirms general intuition about their behavior with changes in the model parameters and provides useful insight for tuning the Hy-Gō parameters to reproduce particular features of a reference model. For example, the ϵ_{nc} and ϵ_{hp} interaction nearly exclusively determine ΔH_{hb}, while ΔS_{nuc} (obtained from building a thermodynamic model for v) depends strongly on both ϵ_{db} and the effective impact of the backbone interactions

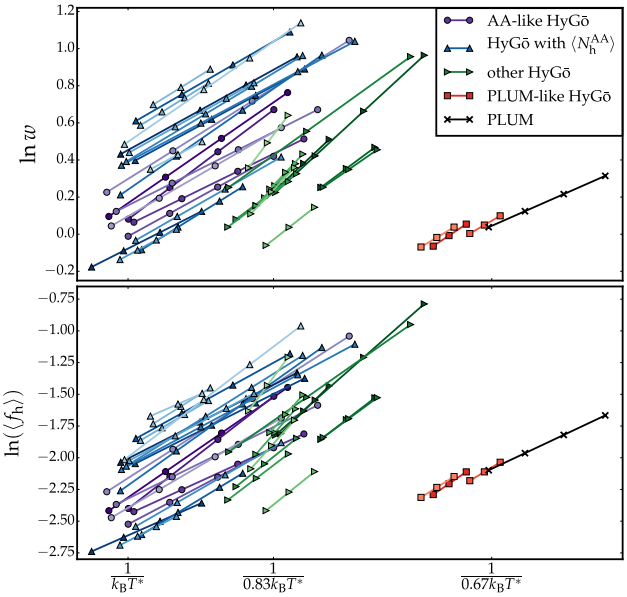

FIG. 7. Temperature dependence of the Lifson-Roig helix propagation parameter, w (top panel), and the average fraction of helical segments, $\langle f_{\mathrm{h}} \rangle$ (bottom panel). The curves presented are a linear fit of the raw data. Purple curves with circle markers denote the 7 AA-like Hy-Gō models. Blue curves with upward triangle markers denote other Hy-Gō models capable of reproducing $\langle N_{\mathrm{h}}^{\mathrm{AA}} \rangle$. Green curves with sideways triangle markers denote Hy-Gō models incapable of reproducing $\langle N_{\mathrm{h}}^{\mathrm{AA}} \rangle$. Red curves with square markers denote the PLUM-like Hy-Gō models. Black curves with X markers denote the PLUM models. Note that the models are aligned by shifting the temperature such that all models achieve $\langle N_{\mathrm{h}}^{\mathrm{AA}} \rangle$ at T^*. For models not simulated at this temperature, a linear extrapolation is employed.

(Figs. S5 and S6).

Although Fig. 4 demonstrates that the AA-like Hy-Gō models display very similar structural and kinetic properties at T^*, Fig. 7 and Table S1 show that their cooperativities (i.e., ΔH_{hb}) vary by more than 25% of the average value for these models. In other words, not surprisingly, reproducing simple structural properties at a single temperature is not enough to determine the thermodynamics of the model. To further validate the structural-kinetic relationships discussed above, we examined these relationships for each of the Hy-Gō models and PLUM models considered, regardless of their particular structural features. There is a 1-to-1 correspondence between the ratio of nucleation and elongation rates and the average propensity of helical segments, as well as the T-dependence of these quantities (Fig. S9). These particular relationships are unlikely to generally hold for much longer, heterogeneous sequences, where complex nucleation/elongation kinetics may arise.[60] However, these results indicate that, when Arrhenius behavior holds, there may be simple relationships between the T-dependence of structural and kinetic properties. Moreover, these findings suggest that reproducing particular structural fea-

tures over multiple temperatures may be an avenue for improving structural, thermodynamic and kinetic consistency for CG models.

E. Relationships between thermodynamics and network topology

The Markov state model characterization of the system kinetics provides a convenient and powerful framework for investigating the details of the hierarchy of kinetic processes sampled by particular helix-coil transitions. Indeed, graph-theoretic observables have been previously employed for understanding complex phenomena, e.g., kinetic frustration,[64] arising within the dynamical processes of proteins. We applied discrete transition path theory[65] to investigate the ensemble of pathways between the helix and coil states and identified simple relationships between the network topology and typical physical observables, e.g., $\langle f_{\mathrm{h}} \rangle$. We found that $\langle f_{\mathrm{h}} \rangle$, at a particular temperature, is largely determined by the number of paths needed to account for half of the fractional flux of probability for trajectories traveling from the coil to the helix (Fig. S12). This somewhat surprising relation is likely due to the relatively strong constraints on the network topology enforced implicitly through the underlying physics of the models.

As the temperature of each system is decreased, the committor value of each state tends to increase while the local entropy tends to decrease (Fig. S11). Because of these generic changes in the network as a function of temperature, there may exist a feature of the network topology, at a particular reference temperature T^{ref}, which determines the thermodynamics, i.e., cooperativity, of the corresponding transition. In the following, T^{ref} corresponds to the highest temperature at which each model was simulated (see Fig. 7). We consider the path entropy, i.e., Shannon entropy of the probability distribution of pathways, which characterizes the degree of randomness of paths traveling from a starting state, s, to an ending state, d. Similarly, the conditional path entropy,[66] $H_{sd|u}$, characterizes the average degree of randomness for paths from s to d passing through a particular intermediate state, u.

Naively, one may hypothesize that networks with more directed flux from coil to helix at T^{ref} would display a faster change of population into the helical state as the temperature is reduced. In other words, in order to maximize the cooperativity, the conditional path entropy, averaged over all intermediate states, should be minimized. To investigate the relationship between directed flux in the network and cooperativity of the underlying transition, we performed a correlation analysis by constructing a linear combination of input features, $\{\mathbf{f}\}$, to characterize the target observable, O: $O \approx \mathcal{F}[\mathbf{f}] = \sum_i c_i f_i$, where the coefficients, $\{c_i\}$, are determined as a best fit over all models. In this case, O corresponds to the slope of $\ln\langle f_{\mathrm{h}} \rangle (1/T)$, denoted simply Δf_{h}, or equivalently ΔH_{hb}.

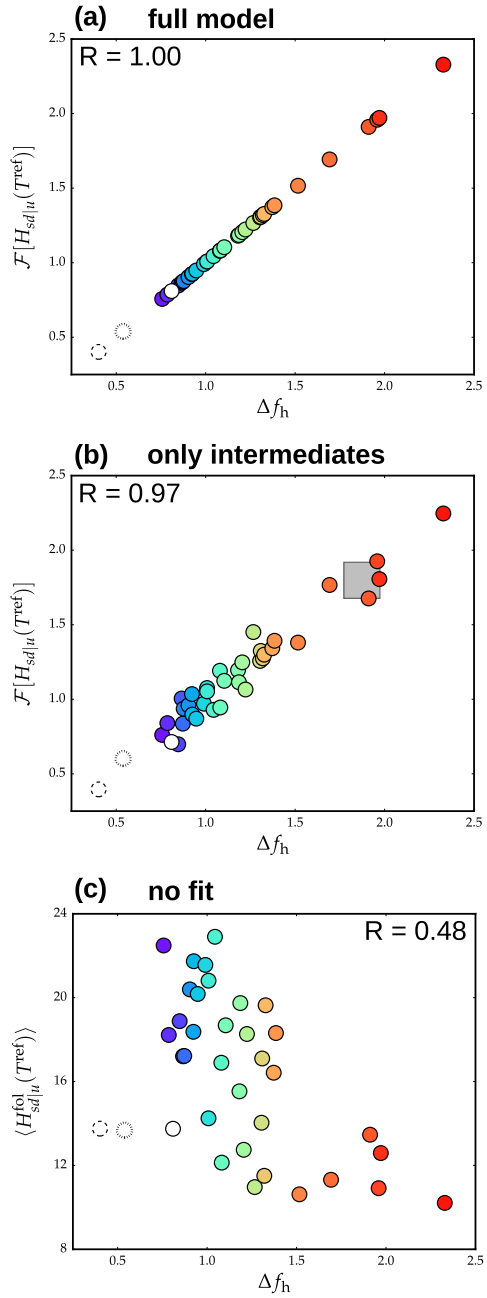

FIG. 8. Relationships between thermodynamic and network properties. Correlation analysis using *(a)* 60 features corresponding to the conditional path entropy, $H_{sd|u}$, for each state in either the folding or unfolding direction and *(b)* a reduced set of 16 features corresponding $H_{sd|u}$, in either direction, for 8 particular states of intermediate helical content. *(c)* Average conditional path entropy in the folding direction versus the slope of a linear fit of $\ln\langle f_{\mathrm{h}} \rangle (1/T)$, denoted Δf_{h}, which represents a simple measure for cooperativity of the transition. Hy-Gō models are denoted by the colored circles, while the PLUM models are denoted by the empty circles. The gray box in panel *(b)* represents the predicted cooperativity of the AA model, based on the topological features of the network at T^*.

The set of 60 conditional path entropies for each intermediate state between coil and helix and for both the folding and unfolding directions were initially considered as input features. For numerical convenience, we weighted each of the features by the fractional flux passing through each state (as introduced above). Panel *(a)* of Fig. 8 demonstrates the resulting correlation when all 60 features are employed. Clearly, there is more than sufficient information in these features to distinguish between the cooperativities of the different models. However, we believe that this correlation is non-trivial since *(i)* the network topology metric, $\mathcal{F}[H_{sd|u}]$, was determined at different temperatures and, therefore, for distinct structural ensembles for each model, *(ii)* we are considering two types of models, namely, Hy-Gō and PLUM, with various parametrizations, and *(iii)* the reference ensemble for each model corresponds to an unfolded ensemble. Moreover, the removal of input features corresponding to the unfolding direction results in significant deviations in the correlation, indicating the necessity of particular features in order to faithfully describe cooperativity.

By systematically disregarding features which played a minimal role in the resulting correlation, we determined a reduced set of 16 features necessary to retain a reasonably accurate correlation (panel *(b)* of Fig. 8). Interestingly, the most important features corresponded to the conditional path entropies for a set of 8 states with intermediate helical content: {cchhh,hcchh,chhcc,cchhc,hchcc,chcch,hcchc,hhccc}. This correlation analysis revealed that, unlike our naive assumption, it is not a simple minimization of the average conditional path entropies which determines a maximally cooperative model. Rather, there is a balance between achieving small conditional path entropies (i.e., directed flux) for some intermediate states and higher conditional path entropies (i.e., nondirected flux) for others. Additionally, there is a strong anticorrelation between features corresponding to the folding and unfolding directions. Apparently, it is a maximization of directed flux in the folding direction for particular states, while maintaining sufficient connections in the network to facilitate a relatively high degree of randomness in the unfolding direction which yields a network topology that, upon cooling, results in the most cooperative transition.

To further demonstrate this concept, panel *(c)* of Fig. 8 presents the conditional path entropy in the folding direction averaged over all intermediate states (i.e., the naive picture of cooperativity described above) versus Δf_{h}. Clearly, the naive interpretation has some merit, indicated by the weak correlation between these quantities, with no adjustable parameters. However, the discrepancies in the correlation, especially for the PLUM models which are significant outliers to the trend, indicate that an essential feature is missing from this simple description of cooperativity. Interestingly, if the anticorrelation between folding and unfolding directions is incorporated by subtracting the average conditional path entropy in the unfolding direction, the correlation for the Hy-Gō models is somewhat improved (see Fig. S13). Further analysis is required to better understand the connection between network topology at a single reference temperature and cooperativity of the model and to test the applicability of this relationship for distinct systems.

IV. CONCLUSIONS

In this study, we have demonstrated that a relatively simple coarse-grained model is capable of nearly quantitatively reproducing the detailed kinetic properties of a complex helix-coil transition. Additionally, we identified a robust relationship between the Lifson-Roig parameters, or similarly the average fraction of helical segments $\langle f_{\mathrm{h}} \rangle$, and the resulting hierarchy of kinetic processes (as determined by a Markov state model of the transition). Importantly, this relationship is likely dependent upon the implicit constraints enforced by the physics of the simulation model. Here, we employed near-atomistic steric and intramolecular backbone interactions, along with simple Gō-type interactions, in order to accurately sample the underlying structural ensemble. This Hybrid Gō (Hy-Gō) model proved extremely useful for the kinetic investigation since it *(i)* employs few, physically-motivated parameters, allowing an extensive and easily interpretable search through parameter space, while *(ii)* allowing a direct, microscopic comparison with other models employing similar backbone resolution and *(iii)* retaining accurate sampling of metastable states along the Ramachandran plot. Similar to other Gō-type models, the Hy-Gō model provides a transparent approach for elucidating the essential interactions necessary to accurately model kinetic properties and may be particularly useful for future investigations of kinetic protein experiments.

Analysis of an all-atom simulation of a heptapeptide of alanines revealed that the seemingly disordered ensemble of pathways between the helix and coil states does not preclude significant coupling of residue dynamics. The Hy-Gō models not only produced consistent kinetics by matching certain structural features of the underlying model, but also demonstrated rather stringent topological constraints on the network, due to the underlying physics of the model. These constraints supposedly resulted in simple relationships between the network topology and standard physical observables. Quite remarkably, the degree of randomness of paths traveling between the coil and helix states through particular intermediate states (as measured by the conditional path entropy) *at a single reference temperature* determines the relative cooperativity of the transition. Although further analysis is required to better understand this relationship, the prediction of structural changes upon environmental or chemical perturbations from the network topology of the unfolded ensemble may be an advantageous approach for understanding the behavior of intrinsically disordered

proteins.

Our investigation successfully identified relationships between various emergent properties of a model through a systematic parameter scan which targeted the features of distinct reference models. The relationships between structural and kinetic properties and between thermodynamics and network topology found in this work were tested on a large variety of Hy-Gō models with varying interaction parameters and also for three different parametrizations of the transferable CG PLUM model. While investigation into the validity of these particular relationships for distinct peptide systems remains for future work, the results presented here contain several general take home messages for the examination of CG kinetic properties. First, the physics of a CG model determine features of the free energy landscape that may constrain the allowable kinetic properties for any particular parametrization. Our previous work demonstrated that, for very short peptides, an accurate representation of local free energy barriers is necessary for a faithful description of the hierarchy of kinetic processes.[24,25] The results of the present work demonstrate, more generally, that when the essential physics is accounted for, i.e., when the structural ensemble is reproduced with sufficient accuracy, simple relationships between structure and kinetics may emerge. These relationships may then be exploited to ensure kinetic consistency through the reproduction of structural properties. Therefore, our work motivates further investigation into structural-kinetic relationships for CG models and suggests that the matching of particular structural properties over multiple temperatures (or other thermodynamic state points) may provide a general scheme for simultaneous inclusion of structural, thermodynamic, and kinetic consistency into CG models.

SUPPLEMENTARY MATERIAL

An online supplement to this article with further methodological details as well as additional results can be found at `http://biorxiv.org`.

DATA

An online database consisting of various analysis scripts and input files for the coarse-grained simulations can be found at `https://github.com/JFRudzinski/Scripts_and_models_for_Structural-kinetic-thermodynamic_relationships.git`

ACKNOWLEDGEMENTS

We thank Fabian Knoch and Raffaello Potestio for critical reading of the manuscript. We are also very grateful to Gerhard Stock for the use of the Ala_7 AA trajectory. J.F.R. thanks Florian Sittel for many fruitful discussions and assistance with the density clustering software. J.F.R. also thanks Amedeo Caflisch and Andreas Vitalis for useful discussions concerning the Lifson-Roig models. J.F.R. is grateful to the organizers of the Hünfeld workshop for supplying a stimulating atmosphere for scientific discussions which made a positive impact to this work. This work was funded through a postdoctoral fellowship from the Alexander von Humboldt foundation (J.F.R.) and an Emmy Noether fellowship from the German Research Foundation (T.B.).

REFERENCES

[1] R. O. Dror, R. M. Dirks, J. P. Grossman, H. Xu, and D. E. Shaw, Ann. Rev. Biochem. **41**, 429 (2012).
[2] T. Ha, Nat. Methods **11**, 1015 (2014).
[3] K. A. Beauchamp, Y.-S. Lin, R. Das, and V. S. Pande, J. Chem. Theor. Comp. **8**, 1409 (2012).
[4] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and A. D. MacKerell Jr., J. Chem. Theor. Comp. **8**, 3257 (2012).
[5] R. B. Best, W. Zheng, and J. Mittal, J. Chem. Theor. Comp. **10**, 5113 (2014).
[6] F. Vitalini, A. S. J. S. Mey, F. Noé, and B. G. Keller, J. Chem. Phys. **142**, 84101 (2015).
[7] C. Schwieters, J. Kuszewski, and G. Clore, Prog. in Nucl. Magn. Reson. Spectrosc. **48**, 47 (2006).
[8] J.-r. Huang and S. Grzesiek, J. Am. Chem. Soc. **132**, 694 (2010).
[9] A. Borgia, W. Zheng, K. Buholzer, M. B. Borgia, A. Schueler, H. Hofmann, A. Soranno, D. Nettels, K. Gast, A. Grishaev, R. B. Best, and B. Schuler, J. Am. Chem. Soc. **138**, 11714 (2016).
[10] B. Schuler, A. Soranno, H. Hofmann, and D. Nettels, Ann. Rev. Biochem. **45**, 207 (2016).
[11] J. D. Honeycutt and D. Thirumalai, Proc. Natl. Acad. Sci. USA **87**, 3526 (1990).
[12] H. Taketomi, Y. Ueda, and N. Gō, Int. J. Pept. Protein Res. **7**, 445 (1975).
[13] K. A. Dill and H. S. Chan, Nat. Struct. Biol. **4**, 10 (1997).
[14] C. Peter and K. Kremer, Faraday Disc. **144**, 9 (2010).
[15] S. Riniker, J. R. Allison, and W. F. van Gunsteren, Phys. Chem. Chem. Phys. **14**, 12423 (2012).
[16] W. G. Noid, J. Chem. Phys. **139**, 090901 (2013).
[17] A. Liwo, S. Oldziej, C. Czaplewski, U. Kozlowska, and H. A. Scheraga, J. Phys. Chem. B **108**, 9421 (2004).
[18] J. Maupetit, P. Tuffery, and P. Derreumaux, Prot. Struct. Func. Bioinfo. **69**, 394 (2007).
[19] T. Bereau and M. Deserno, J. Chem. Phys. **130**, 235106 (2009).
[20] A. Davtyan, N. P. Schafer, W. Zheng, C. Clementi, P. G. Wolynes, and G. A. Papoian, J. Phys. Chem. B **116**, 8494 (2012).
[21] V. A. Harmandaris and K. Kremer, Soft Matter **5**, 3920 (2009).
[22] K. Salerno, A. Agrawal, D. Perahia, and G. Grest, Phys. Rev. Lett. **116**, 058302 (2016).
[23] M. Habibi, J. Rottler, and S. S. Plotkin, PLoS Comput. Biol. **12**, e1005211 (2016).
[24] J. F. Rudzinski, K. Kremer, and T. Bereau, J. Chem. Phys. **144**, 051102 (2016).
[25] J. F. Rudzinski and T. Bereau, Eur. Phys. J. Special Topics **225**, 1373 (2016).
[26] T. Hoang, A. Trovato, F. Seno, J. Banavar, and A. Maritan, Proc. Natl. Acad. Sci. USA **101**, 7960 (2004).
[27] I. R. Cooke, K. Kremer, and M. Deserno, Phys. Rev. E **72**, 011506 (2005).
[28] G. Bellesia and J.-E. Shea, Front. Biosci. **13**, 6957 (2008).
[29] J. Greeley, T. F. Jaramillo, J. Bonde, I. B. Chorkendorff, and J. K. Norskov, Nat. Mater. **5**, 909 (2006).

[30]S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, Nat. Mater. **12**, 191 (2013).
[31]R. Menichetti, K. H. Kanekal, K. Kremer, and T. Bereau, J. Chem. Phys. **147**, 125101 (2017).
[32]T. Bereau, C. Globisch, M. Deserno, and C. Peter, J. Chem. Theor. Comp. **8**, 3750 (2012).
[33]H. S. Chan, Z. Zhang, S. Wallin, and Z. Liu, Ann. Rev. Phys. Chem. **62**, 301 (2011).
[34]V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Prot. Struct. Func. Bioinfo. **65**, 712 (2006).
[35]T. Bereau, M. Bachmann, and M. Deserno, J. Am. Chem. Soc. **132**, 13129 (2010).
[36]T. Bereau, M. Deserno, and M. Bachmann, Biophys. J. **100**, 2764 (2011).
[37]T. Bereau, Z.-J. Wang, and M. Deserno, J. Chem. Phys. **140**, 115101 (2014).
[38]T. Bereau, W. F. D. Bennett, J. Pfaendtner, M. Deserno, and M. Karttunen, J. Chem. Phys. **143**, 243127 (2015).
[39]K. L. Osborne, M. Bachmann, and B. Strodel, (Proceedings, 20-22 July 2011, Julich, Germany) (CBSB11) , 151 (2012).
[40]K. L. Osborne, M. Bachmann, and B. Strodel, Prot. Struct. Func. Bioinfo. **81**, 1141 (2013).
[41]K. L. Osborne, B. Barz, M. Bachmann, and B. Strodel, Phys. Proc. **53**, 90 (2014).
[42]G. O. Rutter, A. H. Brown, D. Quigley, T. R. Walsh, and M. P. Allen, Phys. Chem. Chem. Phys. **17**, 31741 (2015).
[43]A. Altis, M. Otten, P. H. Nguyen, R. Hegger, and G. Stock, J. Chem. Phys. **128**, 245102 (2008).
[44]L. Schuler, X. Daura, and W. Van Gunsteren, J. Comp. Chem. **22**, 1205 (2001).
[45]H. Berendsen, J. Postma, W. van Gunsteren, and J. Hermans, in *Intermolecular Forces*, edited by Pullman, B. (Reidel, 1981) p. 331.
[46]B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, J. Chem. Theor. Comp. **4**, 435 (2008).
[47]Z. J. Wang and M. Deserno, J. Phys. Chem. B **114**, 11207 (2010).
[48]H. Limbach, A. Arnold, B. Mann, and C. Holm, Comp. Phys. Comm. **174**, 704 (2006).
[49]J. D. Chodera, W. C. Swope, J. W. Pitera, and K. A. Dill, Multiscale Model. Simul. **5**, 1214 (2006).
[50]F. Noé, J. Chem. Phys. **28**, 244103 (2008).
[51]Bowman, Gregory R. and Pande, Vijay S. and Noé, Frank, *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation* (Springer Science and Business Media, Dordrecht, Netherlands, 2014).
[52]F. Sittel and G. Stock, J. Chem. Theor. Comp. **12**, 2426 (2016).
[53]M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Perez-Hernandez, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz, and F. Noe, J. Chem. Theor. Comp. **11**, 5525 (2015).
[54]A. Vitalis and A. Caflisch, J. Chem. Theor. Comp. **8**, 363 (2012).
[55]S. Lifson and A. Roig, J. Chem. Phys. **34**, 1963 (1961).
[56]R. B. Best and G. Hummer, J. Phys. Chem. B **113**, 9004 (2009).
[57]G. Schwarz, J. Mol. Biol. **11**, 64 (1965).
[58]P. Thompson, W. Eaton, and J. Hofrichter, Biochemistry **36**, 9200 (1997).
[59]B. Fierz, A. Reiner, and T. Kiefhaber, Proc. Natl. Acad. Sci. USA **106**, 1057 (2009).
[60]M. M. Lin, D. Shorokhov, and A. H. Zewail, Proc. Natl. Acad. Sci. USA **111**, 14424 (2014).
[61]A. Jain and G. Stock, J. Chem. Theor. Comp. **8**, 3810 (2012).
[62]D. Ben-Avraham and J. Köhler, Phys. Rev. A **45**, 8358 (1992).
[63]J. F. Rudzinski and W. G. Noid, J. Chem. Theor. Comp. **11**, 1278 (2015).
[64]A. J. Savol and C. S. Chennubhotla, J. Chem. Theor. Comp. **10**, 2964 (2014).
[65]P. Metzner, C. Schuette, and E. Vanden-Eijnden, Multiscale Model. Simul. **7**, 1192 (2009).
[66]M. Kafsi, M. Grossglauser, and P. Thiran, IEEE Transactions on Information Theory **59**, 5577 (2013).