Published in final edited form as:

Proteins. 2008 February 15; 70(3): 626–638. doi:10.1002/prot.21515.

A coarse-grained α -carbon protein model with anisotropic hydrogen-bonding

Eng-Hui Yap¹, Nicolas Lux Fawzi¹, and Teresa Head-Gordon^{1,2,*}

¹UCSF/UCB Joint Graduate Group in Bioengineering, Berkeley, California 94720

²Department of Bioengineering, University of California, Berkeley, California 94720

Abstract

We develop a sequence based α -carbon model to incorporate a mean field estimate of the orientation dependence of the polypeptide chain that gives rise to specific hydrogen bond pairing to stabilize α -helices and β -sheets. We illustrate the success of the new protein model in capturing thermodynamic measures and folding mechanism of proteins L and G. Compared to our previous coarse-grained model, the new model shows greater folding cooperativity and improvements in designability of protein sequences, as well as predicting correct trends for kinetic rates and mechanism for proteins L and G. We believe the model is broadly applicable to other protein folding and protein–protein co-assembly processes, and does not require experimental input beyond the topology description of the native state. Even without tertiary topology information, it can also serve as a mid-resolution protein model for more exhaustive conformational search strategies that can bridge back down to atomic descriptions of the polypeptide chain.

Keywords

Coarse-grained protein models; anisotropic hydrogen-bonding; protein folding; simulation; kinetics; multi-scale models

INTRODUCTION

Understanding the general energetic principles of protein self-assembly is a long-standing problem in biophysical chemistry. Recently, the framework of energy landscape theory has provided direction in the design of protein folding models that should exhibit correct folding thermodynamics by optimization of a funneled free energy surface.^{1–3} The spatial resolution of the models do not have to be at full atomic detail since it is well known that models with topological features that correctly reproduce the spatial distribution of local and non-local contacts are sufficient for reproducing trends in thermodynamic and kinetic folding data.⁴

Inspired by early efforts of Thirumalai and co-workers,^{5–8} we have developed a “minimalist” protein bead model that uses an α -carbon (C_α) trace to represent the protein backbone, in which structural details of the amino acids and aqueous solvent are integrated out and replaced with effective bead–bead interactions. These physics-based potentials are formulated so that there is still a connection between bead type and amino acid sequence in a reduced letter code, and hence stand distinct from Go-based potentials.⁹ We have successfully used the coarse-grained protein model to study the folding mechanism and

kinetics of several proteins of the ubiquitin α/β topology,^{10–13} to analyze folding simulation protocols,¹⁴ for competition between folding and aggregation in which we correlate differences in aggregation kinetic rates to differences in structural populations of unfolded ensembles,¹⁵ and most recently in aggregation processes relevant for the A β peptide indicted in Alzheimer's disease.¹⁶

When the experimental folding and aggregation data to be understood is of higher spatial or timescale resolution, then isotropic interactions used in protein bead models may break down. One example is the study of early molecular origins of amyloid fiber formation for the A β peptide, in which the mature amyloid aggregate has a precise morphology of unbranched fibers composed of parallel intermolecular β -sheets.¹⁷ To understand these more complex protein assembly or co-assembly problems, it is important to both retain the efficiency of a single bead C_α model while incorporating some of the orientation-dependent properties of amino acids in protein structures. Several models formulated in this spirit include the extension of bead Go-potentials with orientation-dependent statistical potentials,¹⁸ or amino acid specific residue–residue distances.¹⁹

More closely related to this work are formulation of backbone hydrogen bond potentials in the context of off-lattice bead models.^{3,20–22} Onuchic and Cheung incorporated an implicit hydrogen bond in terms of a pseudo-dihedral angle between four C_α centers straddling two separate beta-strands potential within their Go model that uses two centers per residue.³ However, their formulation incorrectly assumes that the strands' C_α centers and hydrogen bonds lie in the same plane, when in fact hydrogen bonds are roughly perpendicular to the planes described by the C_α centers. Brooks and co-workers (private communication) use a three bead per residue model in which the C_α centers are straddled by additional centers embedded with a point dipole to represent the carbonyl and amide peptide linker. The work by Klimov and Thirumalai^{21,22} approximates virtual positions of CO and NH moieties based on C_α positions, which are then used to determine whether the strands are well oriented to form hydrogen bonds. However, their implementation only takes into account hydrogen bond directionality and not hydrogen bond distance, and as a result the folding transition does not exhibit great cooperativity, with folding transitions occurring over a broad temperature range. Furthermore, their model is only effective for α -helical and anti-parallel β -sheet structures, but could not adequately describe parallel β -sheets. The protein model of Smith and Hall²⁰ uses a four center amino acid in which hydrogen-bonds are described as pseudo-bonds between residues to restrict both distance and orientation to realize α -helical and β -sheet structure. In all of these coarse-grained models, the additional centers per residue for a N residue chain scales up the computational cost by $\sim(cN)^2$, where c is the number of centers per residue.

In this article we propose a reformulation of a one-site α -carbon model to introduce a potential of mean force hydrogen bonding term that encourages the cooperative formation of protein-like secondary structures. The orientation-dependent hydrogen bonding term is based on a similar *functional form* developed by Marcus and Ben-Naim²³ and later adopted by Silverstein *et al.*²⁴ to characterize hydrogen-bonding in a model of bulk water. Our protein model now incorporates a mean field estimate of the orientation dependence of the polypeptide chain that give rise to specific hydrogen bond pairing to stabilize α -helices and β -sheets. The model is first parameterized for protein G (PDB code: 2GB1),²⁵ and then validated using folding studies of protein L (PDB code: 2PTL).²⁶ As we show in the Results, the model shows improvements in designability and greater folding cooperativity, and kinetic rates and mechanistic outcomes consistent with experiment.

MODELS AND METHODS

Energy function

The modified minimalist model potential energy function is given by

$$\begin{aligned}
 E = & \sum_{\text{angles}} \frac{1}{2} k_{\theta} (\theta - \theta_0)^2 \\
 & + \sum_{\text{dihedrals}} \left[A [1 + \cos(\phi + \phi_0)] + B [1 - \cos(\phi + \phi_0)] + \right. \\
 & \left. C [1 + \cos 3(\phi + \phi_0)] + D [1 + \cos(\phi + \phi_0 + \frac{\pi}{4})] \right] \quad (1) \\
 & + \sum_{i,j \geq i+3} 4\epsilon_H S_1 \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - S_2 \left(\frac{\sigma}{r_{ij}} \right)^6 \right] + \sum_{\text{Hbonds}} U_{\text{HB}}
 \end{aligned}$$

where θ is the bond angle defined by three consecutive C_{α} beads, ϕ is the dihedral angle defined by four consecutive C_{α} beads, and r_{ij} is the distance between beads i and j . The hydrophobic strength ϵ_H sets the energy scale. The bond angle term is a stiff harmonic potential with force constant $k_{\theta} = 20 \epsilon_H/\text{rad}^2$. The optimal bond angle θ_0 for bead i is set to 95° if bead $i-1$ has helical dihedral propensity, and 105° otherwise.

Our model has been extended to now include new dihedral types in the turn region. As C_{α} -only models lack chirality, we introduce $-/+90^\circ$ turns (designated Q and P, respectively) to distinguish the native topology from its mirror image decoys, and 0° dihedral (designated U) to impose some rigidity in hairpin turns, beyond the original model turn T parameters. In accordance with the flexible nature of turn regions, these new dihedral types have lower barriers than their helical and extended counterparts. Each dihedral angle in the protein chain is then designated to be one of the following types: helical (H), extended (E), or one of the turns (T, P, U, or Q). The parameters A, B, C, D, and ϕ_0 in Eq. (1) are chosen to produce the desired minima (Table I). While all dihedral types encourage formation of the assigned secondary structures, they also allow access to other competing local secondary minima through manageable ($\sim 1-2.8 \epsilon_H$) barriers.

We have also increased the number of bead flavors from three of our original model to four in our new model. The third term in Eq. (1) represents nonlocal interactions between these four bead flavors: strong attraction (B), weak attraction (V), weak repulsion (N), and strong repulsion (L). The amino acid sequence of a protein can be mapped to its four-flavor sequence using the mapping rule shown in Table II, and the bead types determine the type of non-bonded interaction between two beads (Fig. 1). The parameters in Eq. (1) for attractive interactions B-B, B-V, and V-V all have $S_2 = -1$, while $S_1 = 1.4, 0.7$, and 0.35 , respectively. For repulsive interactions, $S_1 = 1/3$ and $S_2 = -1$ for L-L, L-V, and L-B interactions; and $S_1 = 1$ and $S_2 = 0$ for all N-X interactions. The sum of van de Waals radii σ is set at 1.16 to mimic the large exclusion volume due to side chains.

The last term in Eq. (1) represents a new distance and orientation-dependent potential that models backbone hydrogen bond explicitly, to describe a pair-wise mean force hydrogen bond interaction U_{HB} , which is inspired by the Mercedes Benz (MB) model of water first introduced by Marcus and Ben-Naim²³ and further developed by Silverstein and co-workers²⁴. In the original MB model, water molecules are represented as two-dimensional discs with three symmetrically arranged arms, separated by an angle of 120° . Water molecules interact through a standard Lennard-Jones term and an explicit hydrogen-bonding (HB) interaction that is favorable when the arm of one molecule aligns with the arm of another. We have adapted the *functional form* of the hydrogen bonding interaction to our three-dimensional minimalist protein model. The hydrogen bond potential between two beads i and j is given by:

$$U_{\text{HB}} = \varepsilon_{\text{HB}} F \left(r_{ij} - r_{\text{HB}} \right) \times G \left(|\mathbf{t}_{\text{HB},i} \cdot \hat{\mathbf{r}}_{ij}| - 1 \right) H \left(|\mathbf{t}_{\text{HB},j} \cdot \hat{\mathbf{r}}_{ij}| - 1 \right) \quad (2)$$

where

$$\begin{aligned} F \left(r_{ij} - r_{\text{HB}} \right) &= \exp \left[- \left(r_{ij} - r_{\text{HB}} \right)^2 / \sigma_{\text{HBdist}}^2 \right], \\ G \left(|\mathbf{t}_{\text{HB},i} \cdot \hat{\mathbf{r}}_{ij}| - 1 \right) &= \exp \left[\left(|\mathbf{t}_{\text{HB},i} \cdot \hat{\mathbf{r}}_{ij}| - 1 \right) / \sigma_{\text{HB}}^2 \right], \\ H \left(|\mathbf{t}_{\text{HB},j} \cdot \hat{\mathbf{r}}_{ij}| - 1 \right) &= \exp \left[\left(|\mathbf{t}_{\text{HB},j} \cdot \hat{\mathbf{r}}_{ij}| - 1 \right) / \sigma_{\text{HB}}^2 \right] \end{aligned} \quad (3)$$

where r_{ij} is the distance and $\hat{\mathbf{r}}_{ij}$ the unit vector between beads i and j , respectively. The distance dependent term F is a Gaussian function centered at the ideal hydrogen bond distance r_{HB} . For the direction dependent terms G and H , we use an exponential instead of a Gaussian function to ensure a smoother potential energy surface. The vectors $\mathbf{t}_{\text{HB},i}$ and $\mathbf{t}_{\text{HB},j}$ are unit vectors normal to the planes described by bead centers $(i-1, i, i+1)$ and $(j-1, j, j+1)$, respectively. The ideal hydrogen bond distance r_{HB} is set to 1.35 length units for α -helices and 1.25 length units for β -sheets in accordance with a survey of secondary structures in the PDB database. All other hydrogen bond parameters are identical for α -helices and β -sheets, with the width of functions F , G , and H set by $\sigma_{\text{HBdist}} = \sigma_{\text{HB}} = 0.5$.

The hydrogen bond potential is evaluated for all i - j bead-pairs capable of forming hydrogen bonds. Depending on its dihedral propensity, each bead is assigned a hydrogen bond forming capability from three possible types: helical (designated A), sheet (designated B), or none (designated C). For a bead assigned B, the hydrogen bond potential is evaluated between itself and all B-beads situated within a cutoff distance of 3.0 length units. For a bead assigned A, helical hydrogen bond potential is evaluated if its +3 neighbor is similarly assigned A. We find that the helical hydrogen bond is better modeled in a C_{α} -only model as an interaction between $(i, i+3)$ bead pairs, rather than $(i, i+4)$. From a survey of helices in the PDB, the distribution of $r_{i,i+3}$ has both a smaller mean and variance than $r_{i,i+4}$. Hence a potential using $(i, i+3)$ bead pairs is more stringent in discriminating between helical and non-helical geometry. The strength of the hydrogen bond is modulated by ε_{HB} , which is set to $0.7\varepsilon_{\text{H}}$ if the bead pair is B-B, B-V, or V-V. For L-X and N-X pairs, a higher ε_{HB} of $0.98\varepsilon_{\text{H}}$ is required to compensate for the non-bonded repulsion. This provides anisotropy in our C_{α} -only model: L and N residues can maintain closer contact with their hydrogen bonding partners, while remaining repulsive to beads in all other directions.

Protein model

The structural, thermodynamic, and kinetic properties of protein L and G have been well characterized experimentally.^{27–35} Both proteins consist of an N-terminus hairpin, made up by β -strands 1 and 2, followed by a helix, and lastly a C-terminus hairpin made up by β -strands 3 and 4. Despite their similar topologies, L and G share only 15% sequence identity, and fold via different mechanisms. Experimental studies have shown that while the transition state of protein L consists of partially formed β -hairpin 1,^{35,36} that of protein G comprises of partially formed β -hairpin 2.^{30,37} Our existing sequence-based model has been shown capable of predicting the mechanistic differences in L and G folding,¹³ something not possible with Go potentials.

Here we show that our new model preserves this sequence-based feature, and can thus replicate the different folding mechanisms of L and G. In developing the model we optimized the potential energy parameters for protein G in order to reliably reach a global minimum corresponding to the native state topology using simulated annealing, as well as to

yield reasonable thermodynamics, such as sharp cooperative melting curves and heat capacities. We then fixed those parameters to validate the model by characterizing the kinetic mechanism of protein G, as well as the thermodynamics and kinetic mechanism of protein L.

The amino acid sequences of proteins L and G were mapped to reduced minimalist code as per Table II. The dihedral angle propensities were assigned according to their respective PDB structures, with the hairpin turns described using P, U, and Q to encourage the correct chirality. Since we wish to focus on whether differences in the folding behaviors are due to sequence, we assign identical dihedral propensities to hairpins in both L and G. However, the first hairpin turn in protein L (Phe, Ala, Asn, Gly, Ser) is one residue longer than that of protein G (Gly, Lys, Thr, Leu). To address this we use a modified sequence for protein L in which the 11th residue (Asn) is omitted. Dihedral propensities in the hairpins in both proteins can now be similarly assigned for fair comparison, although the model can be reformulated with this extra bead if desired. The hydrogen bond forming capability (A, B, or C) follows the dihedral specification above. The mapped sequence, dihedral propensity and hydrogen bond assignments are listed in Table III.

The initial mapping of the primary sequence from the 20-amino acid code to the 4-letter minimalist code contains some ambiguity. For instance, lysine has both a long hydrocarbon chain and a charged amine group, and could be treated as either hydrophilic or hydrophobic. The initial energy landscape contains many competing local minima due in part to such ambiguity. Sequence design based on the minimal frustration principle is done to smooth the potential energy surface and improve foldability. Our sequence design strategy is based on the theoretical criterion^{1,2} that a foldable heteropolymer sequence has a significant energy gap ΔE between its native-state energy E_{native} and average misfold energy $\langle E_{\text{misfold}} \rangle$. Using our initial mapping sequence, we generate a library of misfolded (non-native) structures from simulated annealing. To obtain a better folding sequence, we generated sequences with various single mutations, threaded them to structures in the misfold library, and select the mutant sequence that maximizes the energy gap ΔE . To minimize drift from the original sequence, we allow only single mutations of types $B \leftrightarrow V$, $V \leftrightarrow N$, or $N \leftrightarrow L$, or dihedral mutations. The mutation process is repeated until we obtain a foldable sequence that finds the native state reliably 50% of the time using simulated annealing.

Simulation protocol

All simulations are performed in reduced units with mass m , energy ϵ_H , length σ_0 , and k_B set to unity. The bond length between adjacent C_α beads serves as the unit of length σ_0 , and is held rigid by using the RATTLE algorithm.³⁸ Reduced temperature and time are given

by $T^* = \epsilon_H/k_B$ and $\tau = (m\sigma_0^2/\epsilon_H)^{1/2}$, respectively. We use constant-temperature Langevin dynamics with a friction coefficient of $0.05\tau^{-1}$, and a timestep of 0.005τ to perform simulations for characterizing the thermodynamics and kinetics of folding.

For each simulated annealing run we launch 50 trajectories at a high temperature ($T^* = 1.6$) and evolve them for 1250τ to generate uncorrelated, unfolded conformations, then gradually cool these trajectories to $T^* = 0.1$ for 7500τ . The trajectories are then annealed at $T^* = 0.45$ for 50τ , and cooled for 5000τ to $T^* = 0.1$, and the anneal-cool cycle repeated once more before the resulting structure is quenched from $T^* = 0.1$ to $T^* = 0$.

The free energy landscape is characterized with the multidimensional histogram technique.^{39,40} We collect multiple nine-dimensional histograms over energy E , radius of gyration R_g , number of native contacts formed Q , number of native contacts formed between

strand 1 and strand 2 ($Q_{\beta 1}$), number of native contacts formed between strand 3 and 4 ($Q_{\beta 2}$), and native-state similarity parameters χ , χ_{α} , $\chi_{\beta 1}$, and $\chi_{\beta 2}$, where χ is given by

$$\chi = \frac{1}{M} \sum_{i,j \geq i+4}^N h\left(\varepsilon - |r_{ij} - r_{ij}^{\text{native}}|\right) \quad (4)$$

The double sum is over beads on the chain, and r_{ij} and r_{ij}^{native} are the distances between beads i and j in the state of interest and the native state, respectively; h is the Heaviside step function, with $\varepsilon = 0.2$ to account for thermal fluctuations away from the native-state structure. M is a normalizing constant to ensure that $\chi = 1$ when the chain is identical to the native state and $\chi \approx 0$ in the random coil state. The remaining χ parameters are specific to their respective elements of secondary structure. That is, χ_{α} involves summation over beads in the helix, and $\chi_{\beta 1}$ and $\chi_{\beta 2}$ involve summation over beads in the first and second β -sheet regions, respectively.

From the histogram method, we get the density of states as a function of a set $[O]$ of nine order parameters, $\Omega([O]) = \Omega(E, R_g, Q, Q_{\beta 1}, Q_{\beta 2}, \chi, \chi_{\alpha}, \chi_{\beta 1}, \chi_{\beta 2})$, which can be used to calculate thermodynamic quantities. In constructing the free energy surfaces, we collect histograms at 14 different temperatures: 1.30, 1.00, 0.80, 0.60, 0.50, 0.40, 0.38, 0.36, 0.34, 0.32, 0.30, 0.25, 0.20, and 0.15. We run five to eight independent trajectories at each temperature and collect 4,000 data points per trajectory. The potential of mean force W along reaction coordinate Q is given by

$$W(Q') = -kT \ln \left[\int d[O] \delta(Q - Q') \Omega([O]) e^{-E/k_B T} \right] \quad (5)$$

The folding kinetics is studied using mean first passage time (MFPT) based on a native state cut-off. With the MFPT method, we decorrelate 2000 independent trajectories at $T^* = 1.6$ for 1250τ , jump to the temperature of interest, and continue evolving the trajectories. We recorded the time τ_i that each trajectory took to enter the native basin of attraction, defined as $Q > 0.8$. The fraction of trajectories folded at time t is then calculated by $P_{\text{nat}}(t) = (\text{no. of trajectories with } \tau_i < t) / N$. Analysis of the $P_{\text{Nat}}(t)$ kinetic data are detailed in Results and Discussions.

Studies of transition state (TS) ensembles are performed using the P_{fold} analysis method.⁴¹ We first identify putative transition states from various projections of order parameters onto the free energy surface. Because we are vetting the new model against a known mechanism, we focused our free energy projections for protein L and G along the order parameters Q and/or $\chi_{\beta 1}$ and Q and/or $\chi_{\beta 2}$, respectively, in order to collect putative TS structures. P_{fold} analysis is then performed: for each putative TS structure, we launch 100 trajectories at the folding temperature, evolve them for 1000τ , and evaluate the probability (P_{fold}) that these trajectories fall into the folded basin (defined as $Q > 0.8$). Structures with $0.4 \leq P_{\text{fold}} \leq 0.6$ are considered to be part of the TS ensemble.

RESULTS AND DISCUSSIONS

Sequence design and native structures

We obtained an optimized sequence for protein L after 12 sequence mutations and three dihedral mutations, while the optimized sequence for protein G consists of 11 sequence mutations and one dihedral mutation. Table III compares the optimized sequences to their original mapping. We find that the original mapping is robust since 50% of the sequence mutations involved ambiguous definitions of valine (B or V) or alanine (V or N), and thus could be explained by these amino acids being “borderline” on the hydrophobic scale. We

find a trend that valines and alanines in the core tend to be retained as B and V (more strongly hydrophobic), while those on the periphery are mutated to V and N (less hydrophobic).

We performed simulated annealing using these optimized sequences to obtain the lowest energy structures (Fig. 2). First we compare the structural similarity of the native state of our protein L and G models with the experimental structures using the Combinatorial Extension (CE) method.⁴² The CE algorithm excludes loop α -carbon positions to align the model and solution structures despite the different lengths of the loop regions. Using the CE method the new model gave RMSDs of 2.6 Å for Protein L and 3.0 Å for protein G, compared to the old model RMSDs of 4.4 Å for Protein L and 5.3 Å for protein G.¹³ We also calculated the root mean square distance (RMSD) of C_{α} atoms between these simulated native structures and their NMR counterparts using the *rms.pl* script from the MMTSB toolbox.⁴³ To ensure a stringent comparison, this time we do not allow gaps or deletions in our alignments, although we modified the 2PTL coordinate file to omit Asn-11 to allow a bead-to-bead comparison with our 60-bead model of protein L. The calculated RMSDs of our simulated native structures are 4.4 Å for Protein L and 3.0 Å for protein G using the alignments with no gaps.

Thermodynamics

Figure 3 plots the thermodynamic averages of percentage folded P_{Nat} [Fig. 3(a)], heat capacity C_v [Fig. 3(b)], and radius of gyration R_g [Fig. 3(c)] against temperature for Protein L and G. Compared to results from our old model with fewer flavors and without the hydrogen bond potential,¹³ the new model demonstrates improved folding cooperativity. The folding temperature T_f , defined as the temperature at which $P_{\text{Nat}} = 0.5$, is 0.36 for protein L and 0.325 for protein G. The thermal stability plots show sharp transitions about T_f , a sign of greater folding cooperativity. The heat capacity and radius of gyration plots likewise show distinct transitions. The collapse temperatures are $T_{\theta} = 0.36$ for protein L and $T_{\theta} = 0.335$ for protein G, indicating that folding (T_f) is concomitant with collapse (T_{θ}).

The thermal stability P_{Nat} plot suggests that Protein L is more stable than protein G at any given temperature. This disagrees with experimental findings that protein G is marginally more stable than protein L under various denaturant conditions.^{30,35} It has been suggested that protein L's instability arises in part from torsional strain in the second hairpin.³⁶ Since we have adopted identical dihedral propensities for hairpins in our model L and G to focus on sequence effects, our models do not take into account this torsional destabilization and could explain why our model protein L appears more stable than protein G. The heat capacity peak for protein L has a larger magnitude than that of protein G, which could be explained by protein L forming more hydrophobic contacts and hydrogen bonds in its native state than protein G.

To examine the free energy landscape, we project the potential mean force W along various order parameters. Figure 4(a,b) show the projections along Q for protein L and G at different temperatures. At their respective folding temperatures, proteins L and G each have two minima (denatured and native), suggesting a two-state folding mechanism. Figure 4(c,d) show the two-dimensional (2D) projections along $\chi_{\beta 1}$ and $\chi_{\beta 2}$ for L and G at their folding temperatures. For Protein L, the minimum-energy path proceeds through a transition state in which β -hairpin 1 is partially formed while β -hairpin 2 is structureless, before reaching the native state. Protein G, on the other hand, has a minimum energy path that involves formation of a native-like β -hairpin 2, before crossing the transition state to reach the native state. The 2D projections are in agreement with experimental evidence that the denatured state ensemble (DSE) and transition state ensemble (TSE) of protein L consist of partially formed β -hairpin 1,^{35,36} while those of protein G involve a partially buried β -hairpin 2.^{30,37}

However, P -fold analysis is needed to determine whether transition state ensembles obtained from the free energy projections are meaningful with respect to folding mechanism.

Transition states analysis

The 2D free energy projections along $\chi_{\beta 1}$ and $\chi_{\beta 2}$ [Fig. 4(c,d)] suggest different minimum free energy paths for the folding of L and G. From these projections, highest energy state for protein L appears to have a partially formed β -hairpin 1, while that of protein G has a partially formed β -hairpin 2, although the relevant transition state ensemble (TSE) may be of higher dimension than suggested by simpler reaction coordinates $\chi_{\beta 1}$ or $\chi_{\beta 2}$. In fact these simpler reaction coordinates proved not to be saddle points on the multi-dimensional energy landscape according to P_{fold} , and therefore we needed to collect putative transition states for more complicated reaction coordinates. We found that the collective Q coordinate combined with $\chi_{\beta 1}$ and $\chi_{\beta 2}$ for proteins L and G respectively were sufficient to determine the TSE. According to the Q - $\chi_{\beta 1}$ projection for protein L, the putative TSE structures are collected for structures with $0.4 < Q < 0.6$ and $0.5 < \chi_{\beta 1} < 0.7$ [Fig. 5(a)]. According to the Q - $\chi_{\beta 2}$ projection for protein G, putative TSE structures are collected for structures with $0.6 < Q < 0.8$ and $0.35 < \chi_{\beta 2} < 0.8$ [Fig. 5(b)]. P_{fold} analysis was performed (see Methods) and we identified the true TSEs for proteins L and G [Fig. 5(c,d) respectively]. Comparing the transition state contacts (red contours) for protein L and G, it is evident that the TSE of protein L consists of more native-like contacts in β -hairpin 1, while the TSE of protein G has more native-like contacts in β -hairpin 2. This is consistent with experimental studies using ϕ -value analysis.^{30,36} The contact maps also show some contacts between strand 1 and 4, which are consistent with experiments. However both TSE contours indicate well-formed helices for L and G, while mutagenesis studies have suggested helices are relatively disrupted in TSEs.

To explore how our model TSE correlates with mutagenesis experiments at a residue level, we perform single mutations on the optimized sequence of protein L and monitor how its transition state is perturbed by each mutation. From the mutations done by Kim *et al.*,³⁶ we performed 16 single-site mutations which can easily be represented by our model (Table IV). Instead of a full free energy calculation for each mutant, we instead evaluated the importance of the mutation for perturbing the N_{TSE} conformational members of the TSE of the WT sequence. For each conformation of the TSE we performed the relevant mutation and ran a P_{fold} calculation in order to compute $1 - N_{\text{TSE(MUT.,i)}}/N_{\text{TSE}}$, where $N_{\text{TSE(MUT.,i)}}$ refers to the number of conformations collected with $0.4 < P_{\text{fold}} < 0.6$.

To compare against ϕ -values, we define a parameter R_i

$$R_i = \begin{cases} 0.2; & 0 \leq 1 - N_{\text{TSE(MUT.,i)}}/N_{\text{TSE}} \leq 0.33 \\ 0.4; & 0.33 < 1 - N_{\text{TSE(MUT.,i)}}/N_{\text{TSE}} \leq 0.5 \\ 0.8; & 0.5 < 1 - N_{\text{TSE(MUT.,i)}}/N_{\text{TSE}} \leq 1.0 \end{cases} \quad (6)$$

to simplify outcomes into low, medium, and high perturbations to the WT TSE. Figure 6 shows the correlation between the experimental ϕ -values and R_i . While there are three outliers (N11V, N26V, and N41V) that deserve to be investigated further, the general trend is consistent with the experimental findings that residues in β -hairpin 1 are more important in the transition state than those in β -hairpin 2. This encouraging result suggests that we can pursue more rigorous free energy calculations of ϕ -values to build on the approximate approach used here.

Kinetics

To rule out the possibility of glassiness, we evaluate the glass transition temperature, T_g , for our model. Wolynes and co-workers⁴⁴ have shown that a foldable, minimally frustrated heteropolymer has a folding temperature well above its glass transition, so that a ratio of T_f/T_g should be greater than one. A working definition of the kinetic glass temperature T_g is the temperature at which average folding time $\langle \tau_f \rangle$ is midway between $\tau_{\min}$, the fastest (minimum) folding time achievable, and $\tau_{\max}$ the simulation cutoff time chosen to greatly exceed the observable folding times⁴⁵ (set to $100,000\tau$ in this work). In Figure 7 we show that this occurs at $T_g = 0.14$, so that $T_f/T_g \sim 2.3$ for our model of Protein G, indicating that the energy landscapes is sufficiently smooth down to fairly low temperatures.

The protein L and G models were next analyzed for the kinetic rates and mechanism of folding at their folding temperatures $T_f = 0.36$ and $T_f = 0.325$, respectively. During folding simulations, there is a finite equilibration time during which trajectories equilibrate from the initial free energy surfaces at $T = 1.6$ to those at their target temperatures. The conventional treatment is to include a fitting parameter for dead time τ_D when fitting $P_{\text{Nat}}(t)$

$$P_{\text{Nat}}(t) = 1 - \sum_i A_i \exp[-(t - \tau_D)/\tau_i] \quad (7a)$$

where A_i is the population for average timescale process τ_i . The parameters used to fit the kinetic data for proteins L and G using Eq. (7a) are listed in Table V.

We have shown in previous work⁴⁶ that, instead of using a constant deadtime, the initial equilibration to the new folding conditions could be better modelled as a relaxation process with Gaussian distributed probability. The overall kinetic data could hence be modeled as a sequential process with (a) initial Gaussian relaxation followed by (b) subsequent (multi)exponential kinetics

$$P_{\text{Nat}}(t) = \int_{u=0}^t \int_{s=0}^{t-u} \frac{1}{\sigma \sqrt{2\pi}} e^{-\left(\frac{(u-\mu)^2}{2\sigma^2}\right)} \cdot \alpha e^{-\alpha s} ds du \quad (7b)$$

Integration of Equation 7b leads to

$$P_{\text{Nat}}(t) = \frac{1}{2} \left[1 + \text{erf}\left(\frac{t-\mu}{\sigma\sqrt{2}}\right) \right] - \sum_i \frac{A_i}{2} \left[1 + \text{erf}\left(\frac{t-B_i}{\sigma\sqrt{2}}\right) \right] e^{-\frac{t}{\tau_i}} e^{-\frac{D_i}{2\sigma^2}} \quad (7c)$$

where $B_i = (\mu + \alpha_i \sigma^2)$ and $D_i = \mu^2 - (\mu + \alpha_i \sigma^2)^2$, with mean μ and variance σ , and α_i is the kinetic folding rate for average timescale process τ_i . The fitting parameters using the sequential fit are also listed in Table V. Comparing the fit quality, it is evident that the sequential mechanism provides a better fit than the dead time treatment, and Figure 8 shows the quality of fit for $P_{\text{Nat}}(t)$ for Protein L and G at their respective folding temperatures. Beyond the equilibration phase, the $P_{\text{Nat}}(t)$ data of Protein L [Figure 8(a)] fits to a single exponential, in agreement with experimental data³⁵. The $P_{\text{Nat}}(t)$ data of Protein G at $T_f = 0.325$ also fits a single exponential [Figure 8(b)], agreeing with single exponential kinetics reported for protein G at its denaturant midpoint.³² The folding time constants for L and G are $11,895\tau$ and 3963τ , respectively. This is in qualitative agreement with experimental data^{30,35} that protein G folds faster than L.

CONCLUSIONS

We have presented an improved coarse-grained model capable of modeling directional hydrogen bonding. The model retains a strong connection between sequence and folding mechanism for proteins L and G, and shows increased folding cooperativity. The model native states also exhibit a greater structural faithfulness to experimentally solved structures. The addition of a fourth bead flavor (V) also provides an improvement over the old model by providing a more graded spectrum of attractive interaction energies (Fig. 1). Overall the improvements to the original model, without introducing greater computational cost, translate to a smoother energy landscape and improved T_f/T_g ratios. The thermodynamic data presented demonstrate that our model assembles more cooperatively and preserves the sequence information that result in different free energy pathways for proteins L and G. This finding is further reinforced by kinetic P_{fold} analysis of their respective TSEs, which show good agreement with experimental mechanisms of protein L and G folding, and decent correspondence with ϕ -value mutation study. The kinetics performed at their melting point ($T = T_f$) showed that both L and G fold via two-state mechanisms, consistent with experimental consensus under these midpoint denaturant conditions.^{32,35}

We believe the model shows promise in application to other protein folding studies. One interesting outcome of the new model is our observation of kinetic complexity and burst phase kinetics under more strongly folding conditions for protein G that we hope to report in a future paper. The computational efficiency of the model has also permitted us to develop molecular models of the Alzheimer's $A\beta_{1-40}$ fibril in order to determine the critical nucleus, stability with chain size, and fibril elongation,¹⁶ opening opportunities for other protein-protein co-assembly processes.

Acknowledgments

THG gratefully acknowledges a Schlumberger Fellowship while on sabbatical at Cambridge University. Molecular graphics for this paper were created in PyMOL⁴⁷. NLF thanks the Whitaker Foundation for its graduate research fellowship.

Grant sponsor: NIH.

REFERENCES

1. Bryngelson JD, Wolynes PG. Intermediates and barrier crossing in a random energy-model (with applications to protein folding). *J Phys Chem.* 1989; 93:6902–6915.
2. Onuchic JN, LutheySchulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. *Annual Rev Phys Chem.* 1997; 48:545–600. [PubMed: 9348663]
3. Cheung MS, Finke JM, Callahan B, Onuchic JN. Exploring the interplay between topology and secondary formation in the protein folding problem. *J Phys Chem B.* 2003; 107:11193–11200.
4. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol.* 1998; 277:985–994. [PubMed: 9545386]
5. Honeycutt JD, Thirumalai D. Metastability of the folded states of globular-proteins. *Proc Natl Acad Sci USA.* 1990; 87:3526–3529. [PubMed: 2333297]
6. Guo ZY, Thirumalai D, Honeycutt JD. Folding kinetics of proteins—a model study. *J Chem Phys.* 1992; 97:525–535.
7. Guo ZY, Thirumalai D. Kinetics of protein-folding—nucleation mechanism, time scales, and pathways. *Biopolymers.* 1995; 36:83–102.
8. Guo Z, Thirumalai D. Kinetics and thermodynamics of folding of a de Novo designed four-helix bundle protein. *J Mol Biol.* 1996; 263:323–343. [PubMed: 8913310]
9. Go N. Theoretical-studies of protein folding. *Ann Rev Biophys Bioeng.* 1983; 12:183–210. [PubMed: 6347038]

10. Sorenson JM, Head-Gordon T. Protein engineering study of protein L by simulation. *J Computat Biol.* 2002; 9:35–54.
11. Sorenson JM, Head-Gordon T. Matching simulation and experiment: a new simplified model for simulating protein folding. *J Computat Biol.* 2000; 7(3/4):469–481.
12. Brown S, Head-Gordon T. Intermediates and the folding of proteins L and G. *Protein Sci.* 2004; 13:958–970. [PubMed: 15044729]
13. Brown S, Fawzi NJ, Head-Gordon T. Coarse-grained sequences for protein folding and design. *Proc Natl Acad Sci USA.* 2003; 100:10712–10717. [PubMed: 12963815]
14. Sorenson JM, Head-Gordon T. Redesigning the hydrophobic core of a model beta-sheet protein: destabilizing traps through a threading approach. *Proteins.* 1999; 37:582–591. [PubMed: 10651274]
15. Fawzi NL, Chubukov V, Clark LA, Brown S, Head-Gordon T. Influence of denatured and intermediate states of folding on protein aggregation. *Protein Sci.* 2005; 14:993–1003. [PubMed: 15772307]
16. Fawzi N, Okabe Y, Yap E, Head-Gordon T. Determining the critical nucleus and mechanism of fibril elongation of the alzheimer's A β 1-40 peptide. *J Mol Biol.* 2006; 365:535–550. [PubMed: 17070840]
17. Dobson CM. Principles of protein folding, misfolding and aggregation. *Semin Cell Dev Biol.* 2004; 15:3–16. [PubMed: 15036202]
18. Buchete NV, Straub JE, Thirumalai D. Orientational potentials extracted from protein structures improve native fold recognition. *Protein Sci.* 2004; 13:862–874. [PubMed: 15044723]
19. Das P, Matysiak S, Clementi C. Balancing energy and entropy: a minimalist model for the characterization of protein folding landscapes. *Proc Natl Acad Sci USA.* 2005; 102:10141–10146. [PubMed: 16006532]
20. Smith AV, Hall CK. Alpha-helix formation: discontinuous molecular dynamics on an intermediate-resolution protein model. *Proteins.* 2001; 44:344–360. [PubMed: 11455608]
21. Klimov DK, Betancourt MR, Thirumalai D. Virtual atom representation of hydrogen bonds in minimal off-lattice models of alpha helices: effect on stability, cooperativity and kinetics. *Folding Design.* 1998; 3:481–496. [PubMed: 9889160]
22. Klimov DK, Thirumalai D. Mechanisms and kinetics of beta-hairpin formation. *Proc Natl Acad Sci USA.* 2000; 97:2544–2549. [PubMed: 10716988]
23. Marcus Y, Ben-Naim A. A study of the structure of water and its dependence on solutes, based on the isotope effects on solvation thermodynamics in water. *J Chem Phys.* 1985; 83:4744–4759.
24. Silverstein KAT, Haymet ADJ, Dill KA. A simple model of water and the hydrophobic effect. *J Am Chem Soc.* 1998; 120:3166–3175.
25. Gronenborn AM, Filpula DR, Essig NZ, Achari A, Whitlow M, Wingfield PT, Clore GM. A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein-G. *Science.* 1991; 253:657–661. [PubMed: 1871600]
26. Wikstrom M, Drakenberg T, Forsen S, Sjobring U, Bjorck L. 3-dimensional solution structure of an immunoglobulin light chain-binding domain of protein-I—comparison with the IgG-binding domains of protein-G. *Biochemistry.* 1994; 33:14011–14017. [PubMed: 7947810]
27. Alexander P, Fahnestock S, Lee T, Orban J, Bryan P. Thermodynamic analysis of the folding of the streptococcal protein-G IgG-binding domains B1 and B2—why small proteins tend to have high denaturation temperatures. *Biochemistry.* 1992; 31:3597–3603. [PubMed: 1567818]
28. Alexander P, Orban J, Bryan P. Kinetic-analysis of folding and unfolding the 56-amino acid IgG-binding domain of streptococcal protein-G. *Biochemistry.* 1992; 31:7243–7248. [PubMed: 1510916]
29. Krantz BA, Mayne L, Rumbley J, Englander SW, Sosnick TR. Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. *J Mol Biol.* 2002; 324:359–371. [PubMed: 12441113]
30. McCallister EL, Alm E, Baker D. Critical role of beta-hairpin formation in protein G folding. *Nat Struct Biol.* 2000; 7:669–673. [PubMed: 10932252]
31. Park SH, ONeil KT, Roder H. An early intermediate in the folding reaction of the B1 domain of protein G contains a native-like core. *Biochemistry.* 1997; 36:14277–14283. [PubMed: 9400366]

32. Park SH, Shastry MCR, Roder H. Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing. *Nat Struct Biol.* 1999; 6:943–947. [PubMed: 10504729]
33. Roder H, Maki K, Cheng H. Early events in protein folding explored by rapid mixing methods. *Chem Rev.* 2006; 106:1836–1861. [PubMed: 16683757]
34. Roder H, Maki K, Cheng H, Shastry MCR. Rapid mixing methods for exploring the kinetics of protein folding. *Methods.* 2004; 34:15–27. [PubMed: 15283912]
35. Scalley ML, Yi Q, Gu HD, McCormack A, Yates JR, Baker D. Kinetics of folding of the IgG binding domain of peptostreptococcal protein L. *Biochemistry.* 1997; 36:3373–3382. [PubMed: 9116017]
36. Kim DE, Fisher C, Baker D. A breakdown of symmetry in the folding transition state of protein L. *J Mol Biol.* 2000; 298:971–984. [PubMed: 10801362]
37. Kuszewski J, Clore GM, Gronenborn AM. Fast folding of a prototypic polypeptide—the immunoglobulin binding domain of streptococcal protein-G. *Protein Sci.* 1994; 3:1945–1952. [PubMed: 7703841]
38. Andersen HC. Rattle—a velocity version of the shake algorithm for molecular-dynamics calculations. *J Computat Phys.* 1983; 52:24–34.
39. Ferguson DM, Garrett DG. Simulated annealing—optimal histogram methods. *Monte Carlo Methods Chem Phys.* 1999; 105:311–336.
40. Ferrenberg AM, Swendsen RH. Optimized Monte-Carlo data-analysis. *Phys Rev Lett.* 1989; 63:1195–1198. [PubMed: 10040500]
41. Du R, Pande VS, Grosberg AY, Tanaka T, Shakhnovich ES. On the transition coordinate for protein folding. *J Chem Phys.* 1998; 108:334–350.
42. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* 1998; 11:739–747. [PubMed: 9796821]
43. Feig M, Karanicolas J, Brooks CL. MMTSB tool set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J Mol Graph Model.* 2004; 22:377–395. [PubMed: 15099834]
44. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein-folding—a synthesis. *Proteins.* 1995; 21:167–195. [PubMed: 7784423]
45. Socci ND, Onuchic JN. Folding kinetics of proteinlike heteropolymers. *J Chem Phys.* 1994; 101:1519–1528.
46. Marianayagam NJ, Fawzi NL, Head-Gordon T. Protein folding by distributed computing and the denatured state ensemble. *Proc Natl Acad Sci USA.* 2005; 102:16684–16689. [PubMed: 16267133]
47. DeLano, WL. The PyMOL Molecular Graphics System. DeLano Scientific; San Carlos, CA: 2002.

Figure 1. Non-bonded interaction energy as a function of pair-wise distance between bead i and j . Interactions BB, BV, and VV have attractive minima at $r_{ij} = 1.3$ while NX and BL/VL/LL interactions are purely repulsive.

Figure 2.

Simulated Annealing Results for Protein L and G. **(a)** PDB structure of Protein L (2PTL) with N-terminus loop region (residue 1–17) omitted. **(b)** Lowest energy structure from simulated annealing of 60-residue optimized sequence of Protein L. RMSD between 2PTL and our model protein L is 4.4 Å **(c)** PDB structure of Protein G (2GB1). **(d)** Lowest energy structure from simulated annealing of 56-residue optimized sequence of Protein G. RMSD between 2GB1 and our model protein G is 3.0 Å.

Figure 3. Thermodynamics averages for proteins L and G as functions of temperature. **(a)** Percentage folded P_{Nat} , **(b)** heat capacity C_v , and **(c)** radius of gyration R_g .

Figure 4.

Free energy surface projections onto different reaction coordinates. **(a)** Projection of protein L's free energy along reaction coordinate Q over temperature range of $0.32 < T < 0.39$. **(b)** Projection of protein G's free energy along reaction coordinate Q over temperature range of $0.29 < T < 0.36$. **(c)** Projection of protein L's free energy surface onto $\chi_{\beta 1}$ and $\chi_{\beta 2}$ at $T_f = 0.36$. **(d)** Projection of protein G's free energy surface onto $\chi_{\beta 1}$ and $\chi_{\beta 2}$ at $T_f = 0.325$. Contours for (c) and (d) are spaced 0.5 kT apart.

Figure 5.

P_{fold} analysis of proteins L and G. Putative transition state ensembles are identified from free energy projections along (a) Q - $\chi_{\beta 1}$ and (b) Q - $\chi_{\beta 2}$ for proteins L and G, respectively. Contact maps of transition state ensembles from P_{fold} for (c) Protein L and (d) Protein G. Black contours denote native contacts. Red contours denote contacts made by 90% of structures in the transition state ensembles.

Figure 6. Correlation between experimental ϕ -values and perturbation to transition state R for protein L. There is general agreement between experiment and R, with some outliers (N11V, N26V, N41V). Both experiment and our simulation indicate residues in β -hairpin 1 are more important for the transition state than those of β -hairpin 2.

Figure 7.

Determining the kinetic glass temperature T_g of protein G. The temperature at which average folding time $\langle \tau_f \rangle$ is midway between $\tau_{\min}$, the fastest (minimum) folding time achievable, and $\tau_{\max}$ the simulation cutoff time chosen to greatly exceed the observable folding times⁴⁵ (set to 100,000τ in this work). We determine that $T_g = 0.14$, so that $T_f/T_g \sim 2.2$ for our model of Protein G, indicating that the energy landscapes is sufficiently smooth down to fairly low temperatures.

Figure 8.

Kinetics data with fits for L and G at their respective folding temperatures using mean first passage time (MFPT) data. **(a)** Percentage of trajectories folded (P_{Nat}) as a function of time for protein L at $T_f = 0.36$. **(b)** Percentage of trajectories folded (P_{Nat}) as a function of time for protein G at $T_f = 0.325$. Both set of data are fitted to both a sequential and dead time model (see text). Fit parameters are listed in Table V. The sequential process is seen to give a better fit to the kinetic data.

Table 1

Parameters for Various Dihedral Types

Dihedral type	A (e_H)	B (e_H)	C (e_H)	D (e_H)	ϕ_0 (rad)	Local minima (global minima in bold)
H (Helical)	0	1.2	1.2	1.2	+0.17	-65°, +50°, 165°
E (Extended)	0.45	0	0.6	0	-0.35	-160°, -45°, +85°
T (Turn)	0.2	0.2	0.2	0.2	0	-60°, -60°, +180°
P (+90°)	0.36	0	0.48	0	+1.57	-155°, -25°, +90°
Q (-90°)	0.36	0	0.48	0	-1.57	-90°, +25°, +155°
U (0°)	0.36	0	0.48	0	+3.14	-115°, +0°, 115°

Table II
Mapping 20-Letter (20) Amino Acid Code to Coarse-Grained Four-Letter Code (4)

20	4	20	4	20	4	20	4
Trp	B	Met	B	Gly	N	Asn	L
Cys	B	Val	B	Ser	N	His	L
Leu	B	Ala	V	Thr	N	Gln	L
Ile	B	Tyr	V	Glu	L	Lys	L
Phe	B	Pro	N	Asp	L	Arg	L

Table III

Sequence, Dihedral, and Hydrogen Bond Assignments for Proteins L and G

Protein L	
1° 2PTL	VTIKANLIFANGSTQTAEFKGTFEKATSEAYAYADTLKKDNGEYTVDVADKGTYLNKIFAG
1° 2PTL (without <i>Asn-14</i>)	VTIKANLIFAGSTQTAEFKGTFEKATSEAYAYADTLKKDGEYTVDVADKGYTILNIKIFAG
1° model L (mapped):	BNBLVLBBBVNNNLNVLBLLNNBLVNNLVVVVVLNBLLLLNLVNB�VLLNVNBLBVBVN
1° model L (optimized):	NNBLVNBVNNNNNNLNVLNBLVNNLVVV VBN NVLNLLNLVNLVVLLNVNVLBLLBN
2° model L:	EEEEEEEQPUEEEEETTPTHHHHHHHHHHHHTPUEEEEEEEEEEPUQEEEEEEE
Hbond model L:	BBBBBBBCBBB BBBBCCAAAAA AAAAAAAAAACCCCB BB BB BB CC BB BB B BB B
Protein G	
1° 2GB1	MTYKLILNGKLTGETTTAEVDAAETA EKVKQYANDNGVDGEWTDYDDATKTFTVTE
1° model G (mapped):	BNVLBBBLNLNBLNLLNNNLVBLVVNVLLBBLLVVLLLNLBLLNBLNVLNLVNLBNBNL
1° model G (optimized):	VNVLBNBLNLNVLNLLNNNLVBLVNNNLLVBLVVLLLVNLVNLVNLVNNNNBNBN
2° model G:	EEEEEEEQPUEEEEETTTQH HHHHHHHHHHHHHTTTTTTEEEEEEPUQEEEEE
Hbond model G:	BBBBBBBCBBB BBBBCCAAAAA AAAAAAAAAACCCCB BB BB BB CC BB BB B BB B

Mutations made are indicated in bold.

Table IV

Mutations Performed on Protein L

Experimental mutation ³⁶	Model mutation	Experimental ϕ -values ³⁶	$1 - N_{\text{TSE}(\text{MUT}_i)}/N_{\text{TSE}}$	<i>R</i>
K7A	L4V	0.70	0.61	0.80
A8G	V5N, E5T ^a	0.43	0.39	0.40
G15A	N11V	0.86	0.24	0.20
T17A	N13V	0.42	0.36	0.40
T19A	N15V	0.17	0.27	0.20
E21A	L17V	1.08	0.61	0.80
K23A	L19V	0.57	0.39	0.40
G24A	N20V	0.20	0.33	0.20
T30A	N26V	0.14	0.88	0.80
N44A	L40V	0.08	0.27	0.20
G45A	N41V	-0.10	0.39	0.40
T48A	N44V	0.44	0.30	0.20
G55A	N51V	0.18	0.33	0.20
T57A	N53V	0.07	0.42	0.40
N59A	L55V	0.12	0.39	0.40
K61A	L57V	0.18	0.33	0.20

Note that residue indices of model L differ from experiment.

^aMutation in dihedral sequence.

Table V

Kinetic Fit Parameters

Conditions		Sequential fit ⁴⁶ (Gaussian relaxation followed by single exponential) [*]			
		μ	σ	τ_0	χ^2
1.	L, $T^* = 0.36$	712	305	11,895	0.0408
2.	G, $T^* = 0.325$	741	450	3,963	0.0935
Single exponential fit with dead time [†]					
Conditions		τ_D	τ_0	χ^2	
3.	L, $T^* = 0.36$	694	11,928	0.0506	
4.	G, $T^* = 0.325$	641	4,142	0.3436	

^{*} Fitted using Eq. 7c.

[†] Fitted using Eq. 7a.