doi:10.1016/j.jmb.2004.04.073 J. Mol. Biol. (2004) 340, 571–585

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

# Gatekeepers in the Ribosomal Protein S6: Thermodynamics, Kinetics, and Folding Pathways Revealed by a Minimalist Protein Model

**Antitsa D. Stoycheva[1,2,3], Charles L. Brooks III[1,2]* and José N. Onuchic[2,3]**

[1]*Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037 USA*

[2]*Center for Theoretical Biological Physics, La Jolla CA 92093-0374, USA*

[3]*Department of Physics University of California at San Diego, La Jolla CA 92093-0319, USA*

**Corresponding author*

We investigate the effect of structural gatekeepers on the folding of the ribosomal protein S6. Folding thermodynamics and early refolding kinetics are studied for this system utilizing computer simulations of a minimalist protein model. When gatekeepers are eliminated, the thermodynamic signature of a folding intermediate emerges, and a marked decrease in folding efficiency is observed. We explain the prerequisites that determine the "strength" of a given gatekeeper. The investigated gatekeepers are found to have distinct functions, and to guide the folding and time-dependent packing of non-overlapping secondary structure elements in the protein. Gatekeepers avoid kinetic traps during folding by favoring the formation of "productive topologies" on the way to the native state. The trends in folding rates in the presence/absence of gatekeepers observed for our minimalist model of S6 are in very good agreement with experimental data on this protein.



*Keywords:* protein folding; gatekeepers; S6; computer simulations

## Introduction

Protein folding has remained a topic of intensive research for several decades, becoming a truly interdisciplinary subject, the investigation of which has brought together efforts in biology, chemistry and physics.[1] A wide variety of experimental techniques, ranging from nuclear magnetic resonance to spectroscopic measurements and protein engineering[2–16] have been employed to elucidate the structure of proteins, as well as the dynamical processes that lead a protein molecule from its unfolded state to its biologically active native conformation. The combination of the wealth of experimental data with theoretical treatments of the folding problem[17–30] has resulted in classifying proteins as a unique set of heteropolymers and in the emergence of the energy landscape theory. Energy landscape ideas applied to folding suggest that a protein molecule diffuses along a funnel-shaped energy surface on its way from a denatured conformation to the folded state. Besides being characterized by a global minimum corresponding to the native state (or, to be more precise, an ensemble of nearly identical native conformations), the folding funnel is also riddled by local minima, manifest as kinetic traps. No single predetermined pathway along the folding funnel defines the folding process of a protein molecule. Rather, an ensemble of routes leads from the denatured state to the native state.[24] Despite the fact that the folding funnel presents the protein with a "choice" of pathways, some of which pass through a number of kinetic traps on the way to the native state and are thus slower,[31] proteins manage to fold quickly on a biologically relevant time-scale, the longest folding times being of the order of milliseconds to seconds.

It is of interest to ask what governs the choice of efficient folding pathways over unproductive ones. Experiments on the ribosomal protein S6 have explored the existence of "gatekeepers" in protein folding.[32,33] Some of these "special" amino acid residues appear to steer the protein away from unproductive folding routes, including slow-folding intermediates;[32] other gatekeepers prevent intermolecular aggregation.[33] We recently investigated these issues from a theoretical point of view utilizing a model β-barrel protein.[34] The

Abbreviations used: GK, gatekeeper; MC, mutant control; WT, wild-type.

E-mail address of the corresponding author: brooks@scripps.edu

0022-2836/$ - see front matter © 2004 Published by Elsevier Ltd.

introduction of gatekeepers enhanced the early folding kinetics of the β-barrel, whose wild-type (WT) has a folding landscape that is characterized by a multitude of long-lived kinetic traps.[35–37]

The present work investigates the mechanisms of gatekeeper action in the ribosomal protein S6 and is directly inspired by the experimental investigation by Otzen *et al*. mentioned above.[32] In that study, an off-pathway intermediate of S6 was observed in the presence of inorganic salt. Additionally, protein engineering analysis demonstrated that mutation of either one of the three gatekeeper amino acid residues in S6 (V85, E22, or A35) stabilized the off-pathway intermediate. Kinetic data were presented suggesting that the intermediate species in the gatekeeper-lacking mutants did eventually reach the native state *via* a direct folding channel; however, the process was several orders of magnitude slower compared to the folding of WT S6.

Here, we address these interesting issues with the help of computer simulations. We study a $C^{\alpha}$ Gō-like[38,39] model of S6. Model variants that include the experimentally suggested gatekeepers[32] and without gatekeepers are explored. We report that mutation of the experimentally suggested gatekeepers leads to the emergence of the thermodynamic signature of an intermediate. The elimination of gatekeepers slows early refolding kinetics considerably. Mutant-control species, which were isoenergetic to the gatekeeper-lacking mutants, but still had the gatekeepers intact, folded as fast as WT S6. The effect of the gatekeepers appears to be mainly topological in nature, as has been suggested by experiment.[32] Our findings regarding gatekeeper location, as well as the effect of gatekeepers on refolding rates, are in very good agreement with the relevant experimental observations of the ribosomal protein S6.[32]

## Results and Discussion

We report our findings regarding the thermodynamic and kinetic properties of the WT S6 sequence, as well as three single mutants in which the experimentally determined gatekeepers E22, A35, and V85 were disabled.[32] The gatekeepers were "mutated" in the $C^{\alpha}$ Gō-like models by removing all favorable long-range native interactions, associated with the particular gatekeeper, from the summation of $V_{LR}(r)$ in equation (4).

### Thermodynamics

Thermodynamic data were collected during constant-temperature simulations as described in Materials and Methods. WT S6 undergoes a folding phase transition at $T_f = 0.336\epsilon/k_B$. The dimensionless heat capacity, $C_v/k_B$, is shown as a function of temperature, in units of $T_f$, in Figure 2(a). The behavior of the heat capacity is indicative

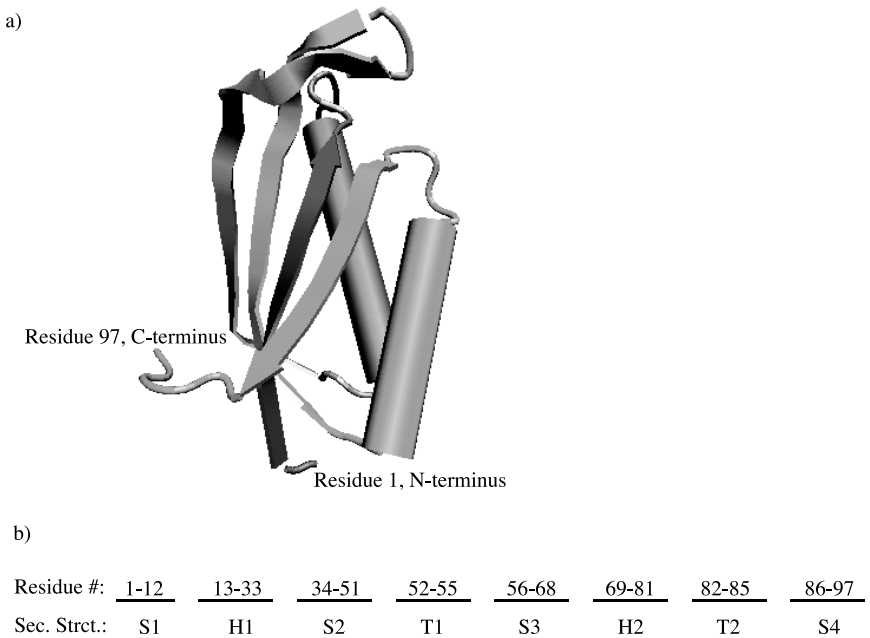


**Figure 1**. The ribosomal protein S6 (PDB accession code 1RIS.pdb[44]). (a) Representation of S6 in which helices are indicated by cylinders, and β-strands by ribbons. Image generated by VMD.[45] (b) Association of a bead's number in the protein sequence of our reduced model of S6 with a secondary structure element. We define strands 1 (S1), 2 (S2), 3 (S3), and 4 (S4); helices 1 (H1), and 2 (H2); and turns 1 (T1), and 2 (T2). In our model, amino acid residues that were experimentally found to lie between subsequent secondary structure elements, but were not explicitly defined as turns, were incorporated in either helices or strands, depending on the dihedral angle potential (see Materials and Methods) that was used to model that particular stretch of the protein sequence.

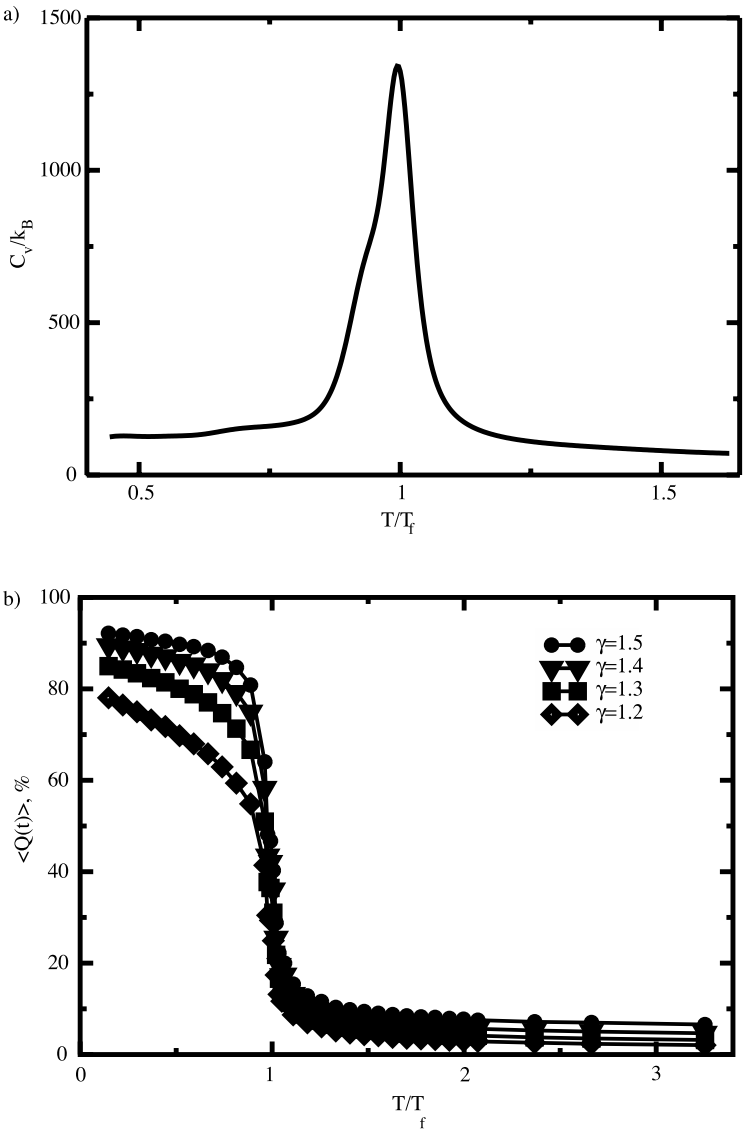


**Figure 2**. (a) The dimensionless heat capacity, $C_v/k_B$, for WT S6 as a function of temperature, in units of $T_f$. The behavior of the heat capacity is indicative of a first-order phase transition in a finite system. (b) Average number of native contacts $\langle Q(T)\rangle$ for WT S6 as a function of $T/T_f$. Curves for native distance cutoff coefficients $\gamma = 1.2$, 1.3, 1.4, and 1.5 are shown. Our results were not altered by small variations in the choice of $\gamma$. Larger values for $\gamma$ result in an effectively sharper folding transition, since structures corresponding to low-temperature vibrational states of the protein are considered at a lower resolution and only significant changes in the protein conformation, which start to appear near the folding transition, change the slope of the $\langle Q(T)\rangle$ curve.

of a first-order phase transition in a finite system. Figure 2(b) shows $\langle Q(T)\rangle$, the average number of native contacts formed during a thermodynamic sampling simulation at temperature $T$, as a function of $T/T_f$. A pair of beads $(i, j)$, found at a distance $r_{ij}^0$ in the native structure of S6, was considered to form a native contact at step $k$ of a simulation if the distance between them was $d_{ij} \leq \gamma r_{ij}^0$, with $\gamma = 1.5$. Our results were not altered by small variations in the choice of $\gamma$. Figure 2(b) shows $\langle Q(T)\rangle$ for $\gamma = 1.2$, 1.3, 1.4, and 1.5. Larger values for $\gamma$ result in an effectively sharper folding transition, as illustrated by $\langle Q(T)\rangle$, since structures corresponding to low-temperature vibrational states of the protein are considered at a lower resolution and only significant changes in the protein conformation, which start to appear near the folding transition, change the slope of the curve. In mathematical terms, since the relation between $Q(T)$ and the internal energy $U(T)$ for our off-lattice model is essentially linear (data not shown), then for a two-state protein the width of the transition[40] in $Q(T)$ would be proportional to $1/(S_u - S_f)^2$, where $S_u$ is the entropy of the unfolded state, and $S_f$ is the entropy of the folded state. Increasing the value of $\gamma$ effectively decreases $S_f$ and thus reduces the width of the transition. We also compared the results of refolding kinetics (see Early folding kinetics) as a function of $\gamma$ for 20% of WT S6 refolding trajectories, and observed identical folding yields for $\gamma = 1.2$ and $\gamma = 1.5$.

The Helmholtz free energy ($F = U - TS$) of WT S6 is shown as a function of the internal energy $U$ in Figure 3(a), and as a function of the folding order parameter $Q$ (percentage of native contacts) in Figure 3(b) for several temperatures close to $T_f$. At the folding temperature $T = T_f$, $F$ exhibits two wells of equal depth, corroborating the first-order phase transition in a finite system already indicated for our model by the peak in the heat capacity. The free energy minimum at $-11.5$ $U/\epsilon$ corresponds to a basin of folded structures and the minimum at 9.5 $U/\epsilon$ corresponds to a basin of unfolded conformations. The two basins are

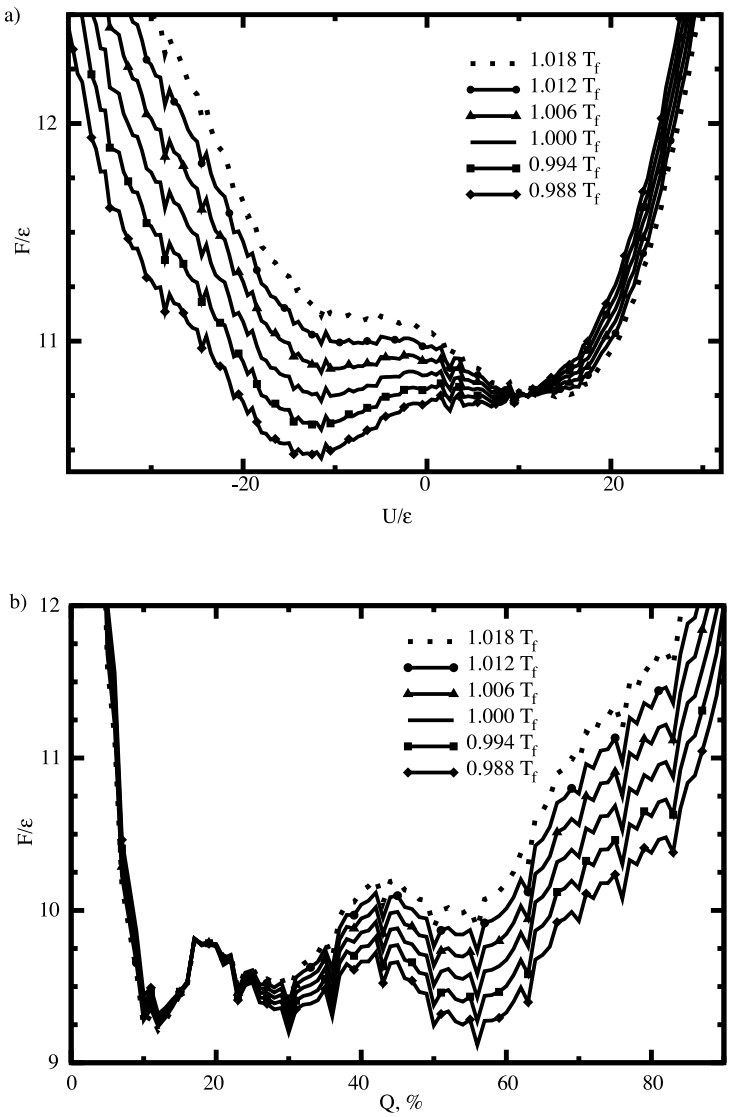


**Figure 3**. (a) The Helmholtz free energy ($F = U - TS$) of WT S6 as a function of the internal energy $U$ for several temperatures close to $T_f$. All energies are given in units of ϵ and temperatures in units of $T_f$. At the folding temperature $T = 1.000T_f$, $F$ exhibits two wells of equal depth, corroborating the first-order phase transition in a finite system already indicated for our model by the peak in the heat capacity (Figure 2). (b) *F versus Q*. The high-$Q$ values, centered around $Q = 60\%$, correspond to the folded-structures, low-energy basin in the $F-U$ plane. The structures high in internal energy, corresponding to the unfolded basin in (a), cluster around $Q = 10\%$ and $Q = 30\%$. This is a result of the gradual dissolution of the central part of the protein hydrophobic core, formed by S2, S3, and T1 (Figure 1).

separated by a small barrier of $0.1\epsilon = 0.3k_BT_f$. In Figure 3(b) the high-$Q$ values centered around $Q = 60\%$ correspond to the folded-conformations, low-energy basin in the $F-U$ plane (Figure 3(a)). These are separated from the unfolded structures ($Q \approx 30\%$ or less) by a barrier of 0.45 ϵ ($1.3k_BT_f$). Structures high in internal energy (unfolded basin in Figure 3(a)), cluster around $Q = 10\%$ and $Q = 30\%$. The two groups of unfolded structures are separated by a free energy barrier of 0.35 ϵ ($1.0k_BT_f$). This additional unfolded ensemble barrier results from the gradual dissolution of the central part of the protein hydrophobic core, formed by S2, S3, and T1 (Figure 1). Contacts between these three secondary structure elements contribute 27% of all native interactions. Visual inspection of trajectories that yield $Q = 30\%$ or $Q = 10\%$ also shows that the loop formed by S2, S3 and T1 is present to a varying extent in these instances, with other secondary structure elements or local tertiary structure either completely random, or only partly formed.

Slightly below $T_f$ the C-terminal strand S4 becomes highly mobile and part or all of the 40 native contacts associated with it (18.6% of all native contacts) are easily broken. That is why, in Figure 3(b) the basin of low-energy structures is centered around $Q \approx 60\%$. For temperatures $T \leq 0.91T_f$ the protein populates a state with $Q \approx 80\%$, i.e. the native structure (data not shown).

We proceed to discuss the thermodynamics of the three mutants in which gatekeeper-associated native contacts for residues E22 (mutant M-E22), A35 (mutant M-A35), or V85 (mutant M-V85) were removed. We once again point out that determining which contacts should be disabled for the gatekeeper-lacking mutants was based purely on an analysis of the contact map of WT S6 (extracted form the PDB file for this species as described in Materials and Methods) without any additional bias. This analysis allowed us to classify the native contacts of the gatekeepers according to two criteria: (1) which secondary structures of S6 (defined in Figure 1) were involved in a given contact; and (2) the interaction range of the contact, based on the sequence separation of the two residues

**Table 1.** The nature of gatekeeper and control contacts mutated in this study

| Gatekeeper (GK) | Native contacts for GK | Mutant controls–removed contacts |
|---|---|---|
| V85 | S1–T2; long-range<br>H2–T2; short-range | S1–S4; longer-range than GK<br>H2–S4; longer-range than GK |
| E22 | H1–H1; $i$, $i+4$ and $i$, $i+8$<br>H1–T2; long-range | H1–H1; $i$, $i+5$<br>H1–T2; long-range |
| A35 | H1–S2; short-range<br>S2–S3; longest-range | H1–S2; long-range<br>S2–S3; medium long-range |

Gatekeeper native contacts are characterized in column 2 of the Table; contributions due to these contacts were excluded from the summation of favorable native interactions in our studies of gatekeeper-lacking mutants (M-A35, M-E22 and M-V85). The types of native contacts that were removed in the mutant controls for which gatekeepers were left intact (MC-A35, MC-E22 and MC-V85) are summarized in column 3 of the Table. All contacts are classified according to secondary structure involved and interaction range.

involved; the range of a gatekeeper native contact associated with a given pair of protein secondary structure elements was defined as "long", "short", etc. relative to the separation of all other native contacts between the same two regions of S6. Column 2 of Table 1 summarizes our findings. According to the WT contact map, the gatekeeper residue A35 has a total of six native contacts: two short-range interactions between helix H1 and strand S2, and four long-range contacts between S2 and S3. The gatekeeper E22 is involved in three short to medium-range intrahelical H1 interactions and two long-range H1–T2 interactions. V85 is associated with four long-range interactions between S1 and T2, and three H2–T2 short-range interactions. As we mentioned at the start of the Results and Discussion section, "mutating" a given gatekeeper consisted of omitting all contributions to equation (4) due to native interactions of that gatekeeper. This alters the mutant Hamiltonians as follows:

The mutants were destabilized by 5.5 $\epsilon$ (M-V85, seven native contacts removed), 5.0 $\epsilon$ (M-E22, five native contacts removed), 4.0 $\epsilon$ (M-A35, six native contacts removed), compared to WT S6. (In our kinetic studies of these mutants, discussed in the Early folding kinetics and subsequent sections we took special care to ensure that the differences between WT and mutant Hamiltonians did not introduce biases favoring one species over the others in the refolding simulations. As an additional impartial test, refolding was also preformed for mutant control species as described later on in the text.) Figure 4(a), (c), and (e) show the dimensionless heat capacity, $C_v/k_B$, as a function of temperature, in units of $T_f$, for the three mutants. All three mutants exhibit a "main" folding transition around $T = T_f$, the same temperature as WT S6. Additionally, the $C_v$ curves for the mutants show extra "bumps", indicating the presence of folding intermediates. Evidence for intermediates is also presented by the "bumps" and "steps" in the transition regions of the $\langle Q(T)\rangle$ curves for the three systems (Figure 4(b), (d), and (f)). Experimentally,[32] the most pronounced gatekeeper was found to be V85, closely followed by E22 and, to a lesser extent, by A35. Our minimalist models for M-V85, M-E22, and M-A35 reflect this fact.

Figures 5–7 show the free energy profiles for M-V85, M-E22, and M-A35, respectively. In all three Figures, part (a) refers to the corresponding folding intermediate transitions observed in the $C_v$ curves at 0.740 $T_f$ for M-V85, 0.920 $T_f$ for M-E22, and 0.914 $T_f$ for M-A35. Part (b) illustrates the behavior of the systems around $T_f$, the "main" folding transition. For all three mutants, the transition associated with the intermediate is characterized by a very small barrier in both the $F-U$ and $F-Q$ planes, and the protein populates an "intermediate" structure in which about 60% of the native contacts are formed (the native state has about 80% of the PDB contacts formed, as shown in Figures 2(b) and 4(c)–(e)).

The structure of the intermediate is essentially identical for all three mutants, as judged by a comparison of the respective contact maps (data not shown). This "thermodynamic" intermediate, which gets significantly populated in the absence of gatekeepers, also corresponds to the protein conformations with $Q \approx 60\%$ that are only transiently observed for the WT species around $T_f$. In all cases, the C-terminal strand S4 becomes entropically rich and mobile, and "detaches" itself from the rest of the protein, which remains in a near-native configuration. Some native contacts associated with S4's neighboring turn T2 and helix H2 are also broken.

Our simplified models of the gatekeeper-lacking mutants of S6 clearly identify the importance of reinforcing certain local secondary structure elements, such as turns, as well as long-range interactions that facilitate native packing. These findings broaden our understanding of the role of the structural gatekeepers of S6. However, we point out that in order to directly compare the nature of a model misfolded "intermediate" to the experimentally suggested one,[32] a more detailed model including side-chains would likely offer further insights.

## Early folding kinetics

In this section we discuss the effect that

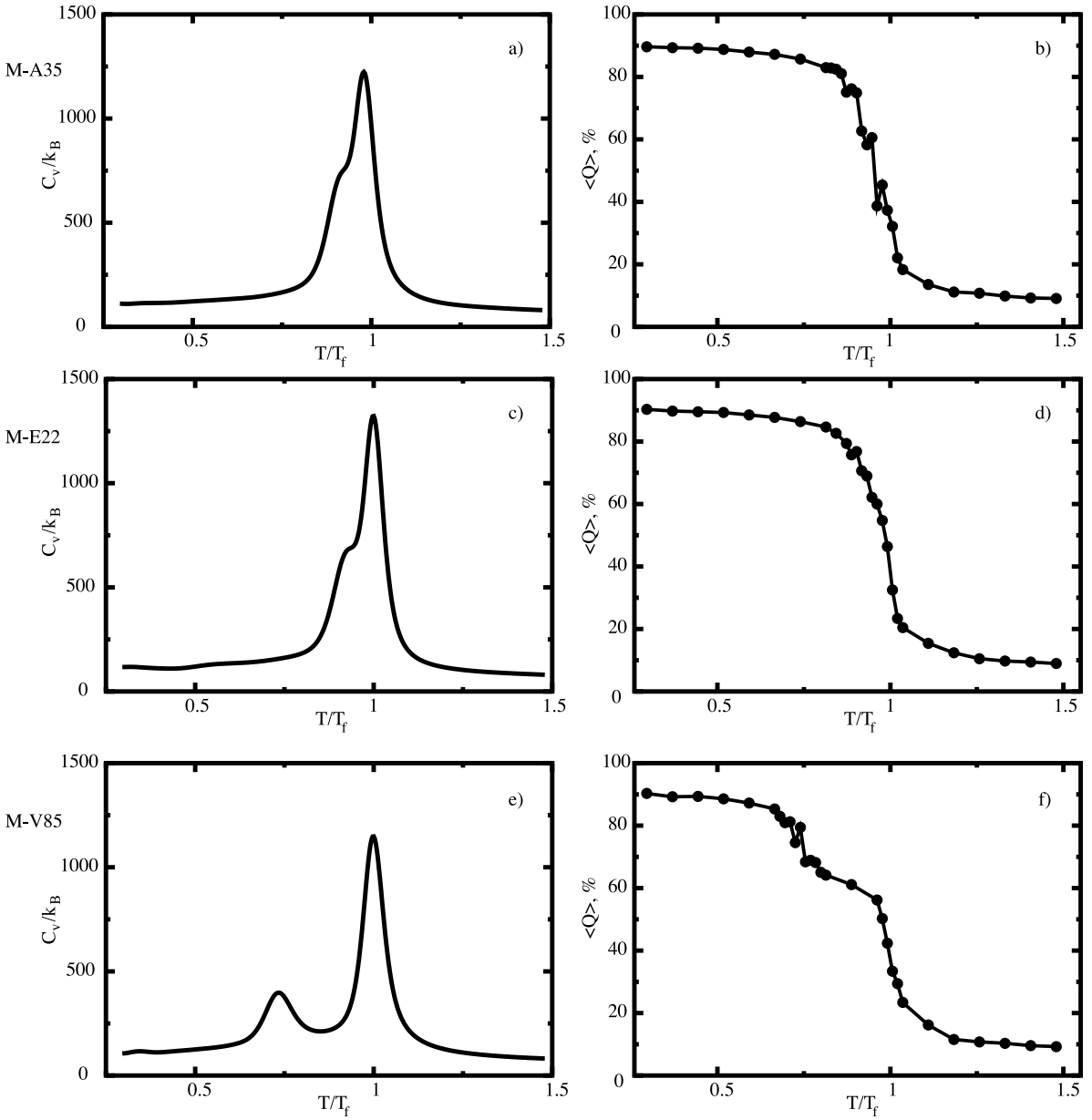


**Figure 4**. Left column of panels: the dimensionless heat capacity, $C_v/k_B$, as a function of temperature, in units of $T_f$, for the gatekeeper-lacking mutants M-A35, M-E22, and M-V85. All three mutants exhibit a "main" folding transition around $T = T_f$, the same temperature as WT S6. Additionally, the $C_v$ curves for the mutants show extra "bumps", indicating the presence of folding intermediates. Right column of panels: $\langle Q(T) \rangle$ curves for the three systems. The evidence of intermediates is corroborated by the "bumps" and "steps" in the transition regions of $\langle Q(T) \rangle$. Experimentally, the most pronounced gatekeeper was found to be V85, closely followed by E22, and to a lesser extent by A35.[32] Our minimalist models for M-V85, M-E22, and M-A35 reflect this fact.

gatekeepers have on the early folding kinetics of S6. We have investigated the refolding behavior of WT S6, M-V85, M-E22, and M-A35 according to the simulation procedure outlined in Materials and Methods. The systems were refolded at the following $T_{ref}$ temperatures: 0.888 $T_f$ (WT S6), 0.666 $T_f$ (M-V85), 0.843 $T_f$ (M-E22), and 0.858 $T_f$ (M-A35). $T_{ref}$ values were chosen to be lower than and approximately equally offset from the corresponding folding transition temperatures indicated by the heat capacities of the four systems (Figures 2 and 4) which reflect the thermodynamic stabilities of the different species (see Thermodynamics). Our choice of refolding temperatures ensured that the WT and the mutants were refolded "on an equal footing" from an energetic point of view: since the mutant species were destabilized compared to WT S6 (as discussed in the Thermodynamics section), the native protein was refolded at a higher temperature.

Folding progress was measured by tracking $P_v(t)$, the fraction of unfolded trajectories as a function of time.[31,34] The percentage of native contacts $Q$ was used as the measure for nativeness during

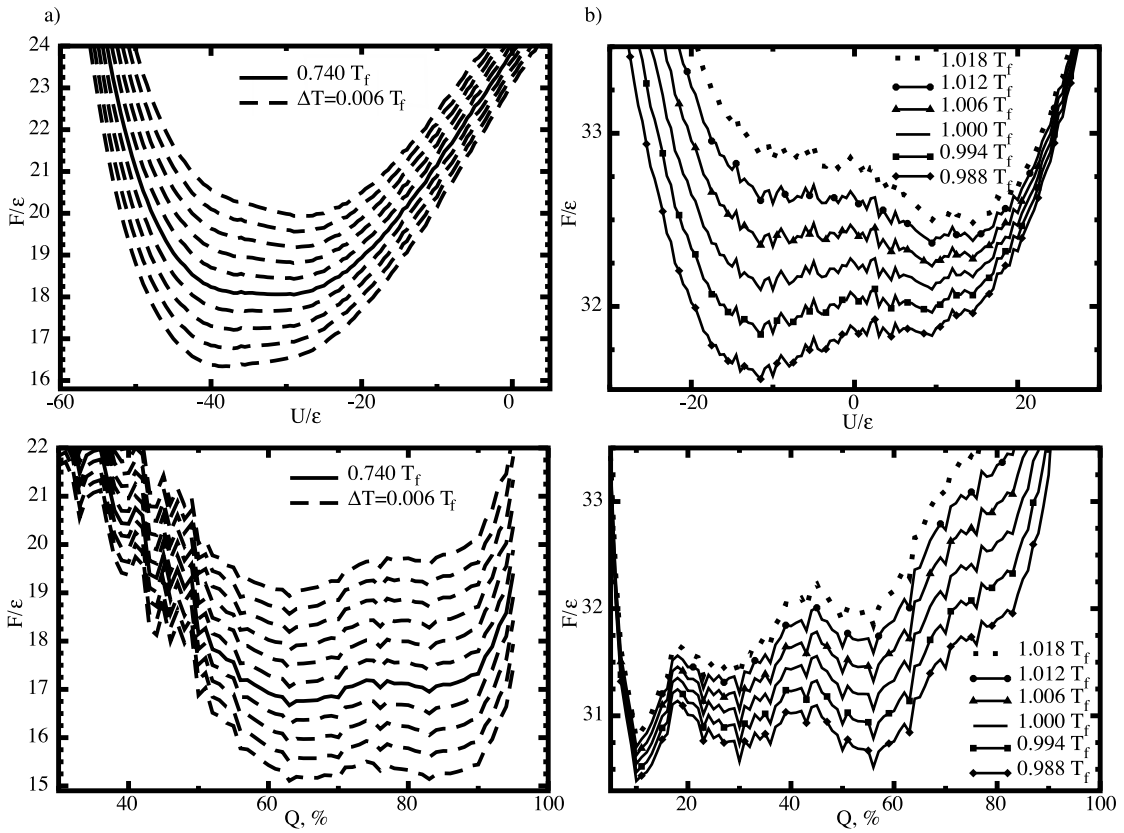


**Figure 5**. Free energy profiles for M-V85. All energies are given in units of ϵ and temperatures in units of $T_f$. (a) Native-to-intermediate transition observed at 0.740 $T_f$. Curves are drawn in intervals of $\Delta T = 0.006T_f$. The continuous lines correspond to $T = 0.740T_f$. (b) "Main" folding transition (i.e. intermediate-to-unfolded) at $T = 1.000T_f$.

refolding, and the cutoff value $Q_c$ separating folded from unfolded states was determined separately for WT S6, M-V85, M-E22, and M-A35 from the $\langle Q(T)\rangle$ thermodynamic data (Figures 2 and 4) at the appropriate temperature. (Exact $Q$ values were available for all $T_{ref}$, so no extrapolation was done for $Q_c$.)

Figure 8 characterizes the early refolding kinetics of WT S6 and its three gatekeeper-lacking mutants. WT-species trajectories (75%) folded within $2.0 \times 10^6$ time steps (13.46 ns). An equivalent folding yield (78%) is observed for M-A35, the mutant that lacks the least significant experimentally suggested gatekeeper.[32] On the other hand, M-V85 and M-E22, for which efficient gatekeepers[32] have been removed, show a marked decrease in productive early folding, with only 54% and 49% of trajectories leading to the folded state, respectively. The refolding trends we observe are in very good agreement with the experimentally measured kinetics of S6 and its mutants.[32]

We have also investigated the refolding kinetics of three mutant control (MC) species: MC-V85, MC-E22, and MC-A35. The purpose of the control studies was to clearly demonstrate that gatekeepers are located in crucial positions in the protein.[32,33] As with M-V85, M-E22, and M-A35, certain native interactions were excluded from the Hamiltonians of the mutant controls. The type of native contacts that were removed in the mutant control for each gatekeeper (V85, E22, or A35) are summarized in the third column of Table 1. In all three MC cases, the actual gatekeepers were left in place, and other native interactions were eliminated such that the mutant control and its corresponding gatekeeper-lacking mutant were destabilized by the same amount of native interaction energy, compared to WT S6, namely: 5.5 ϵ for M-V85 and MC-V85, 5.0 ϵ for M-E22 and MC-E22, and 4.0 ϵ for M-A35 and MC-A35. In the mutant controls, native contacts between the same secondary structure elements, as in the original gatekeeper-lacking mutant being tested, were removed. (V85 is the only residue that contributes native contacts between H2 and T2 in WT S6, as determined by the procedure described in Materials and Methods. Therefore, for MC-V85, native contacts between H2 and S4 were removed. S4 follows immediately after T2, as shown in Figure 1.) Additionally, all native interactions between the secondary structures of interest were examined and native contacts that involved V85, E22, and A35 were classified as either short-range or long-range, based on sequence separation of the gatekeeper and its contact (column 2 of Table 1). Most native interactions that were removed in MC-V85, MC-E22, and MC-A35 were deliberately of different range than those eliminated in M-V85, M-E22, and M-A35 (column three of Table 1). Figure 9 compares the early folding of WT S6,

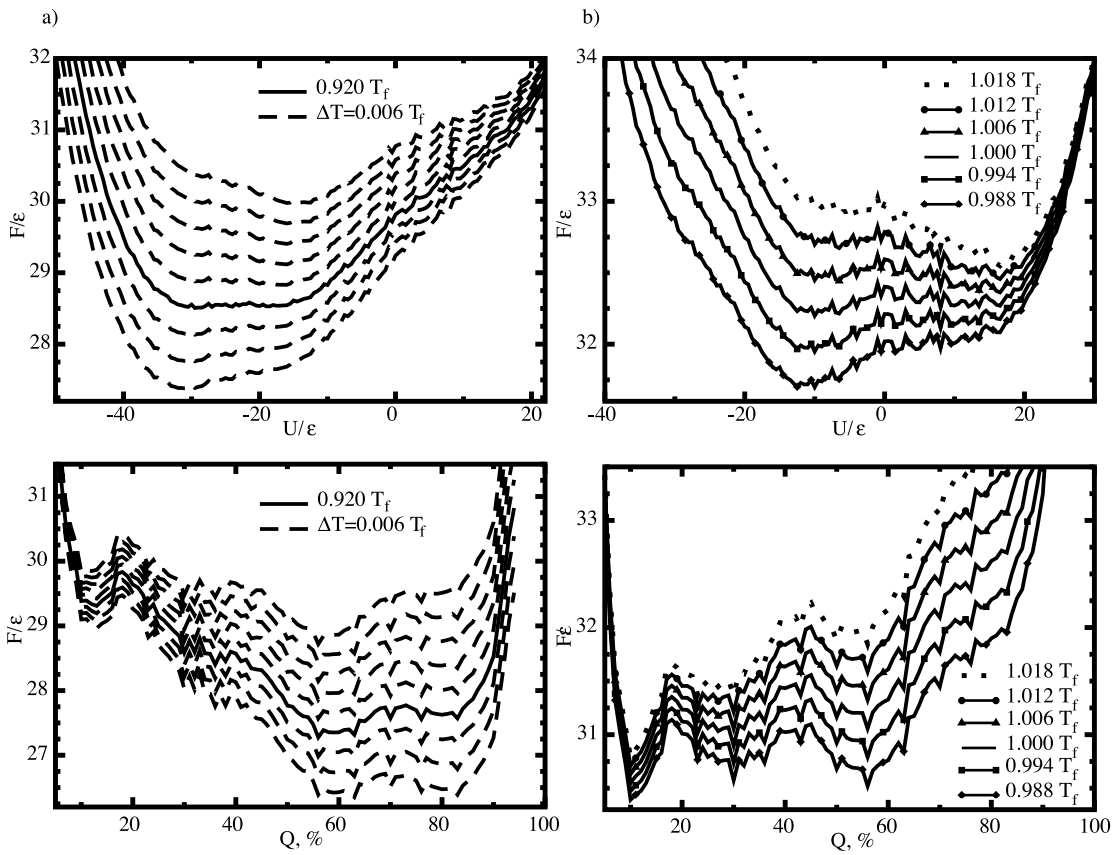


**Figure 6**. Free energy profiles for M-E22. All energies are given in units of ϵ and temperatures in units of $T_f$. (a) Native-to-intermediate transition observed at 0.920 $T_f$. Curves are drawn in intervals of $\Delta T = 0.006T_f$. The continuous lines correspond to $T = 0.920T_f$. (b) "Main" folding transition (i.e. intermediate-to-unfolded) at $T = 1.000T_f$.

M-V85, M-E22, and M-A35 (already shown in Figure 8) to the behavior of the mutant controls. Folding progress is measured by $P_v(t)$, the fraction of unfolded trajectories as a function of time. All three mutant controls exhibit folding yields that are essentially the same as those of the WT protein (i.e. the best anticipated folder of all considered systems) and the mutant of the "weakest" gatekeeper, M-A35. This demonstrates that the gatekeepers' locations, as well as the particular combination of long-range and short-range interactions in which they are involved, are essential to the protein's ability to choose productive folding pathways early on in the folding process.

Analysis of trajectories that did not fold within the investigated time-scale (data not shown) has demonstrated that the protein conformations associated with the kinetic traps correspond to the "thermodynamic" intermediate discussed in the Thermodynamics section.

## Folding pathways

To gain further insight into how exactly the folding mechanism of S6 is influenced by the gatekeepers, we have examined the refolding kinetics data presented above using ensemble-averaged local native contact values as a function of time $\langle Q_{ij}\rangle(t)$.[34] Here, $i$ and $j$ are any of the secondary structure elements of S6 given in Figure 1: S1, S2, S3, S4, H1, H2, T1, or T2. There are 36 unique $(i, j)$ pairs, but only 22 of these actually share native contacts, and $\langle Q_{ij}\rangle(t)$ were calculated just for them. The local native contact values monitor the dynamics of folding of individual parts of the protein and are thus capable of distinguishing differences in folding events that arise due to the absence of gatekeepers.

The results of our local contact order investigations are graphically illustrated in Figure 10. Regardless of the type of species (WT or mutants), folding occurs at five distinct stages. (Using raw data, we distinguished the different stages from each other by observing the time-dependent, steep increase in native contact formation between the secondary structure elements of the above-discussed $(i, j)$ pairs, which maps the sequence of folding events.) At each stage, more than one "discrete" event of secondary structure formation or tertiary packing occurs. Each stage in the folding process is represented in Figure 10 by a backbone trace and a projection diagram (top view), in which residues along the protein sequence were colored depending on whether they had formed native contacts at that particular point during refolding. The coloring scheme used in Figure 10

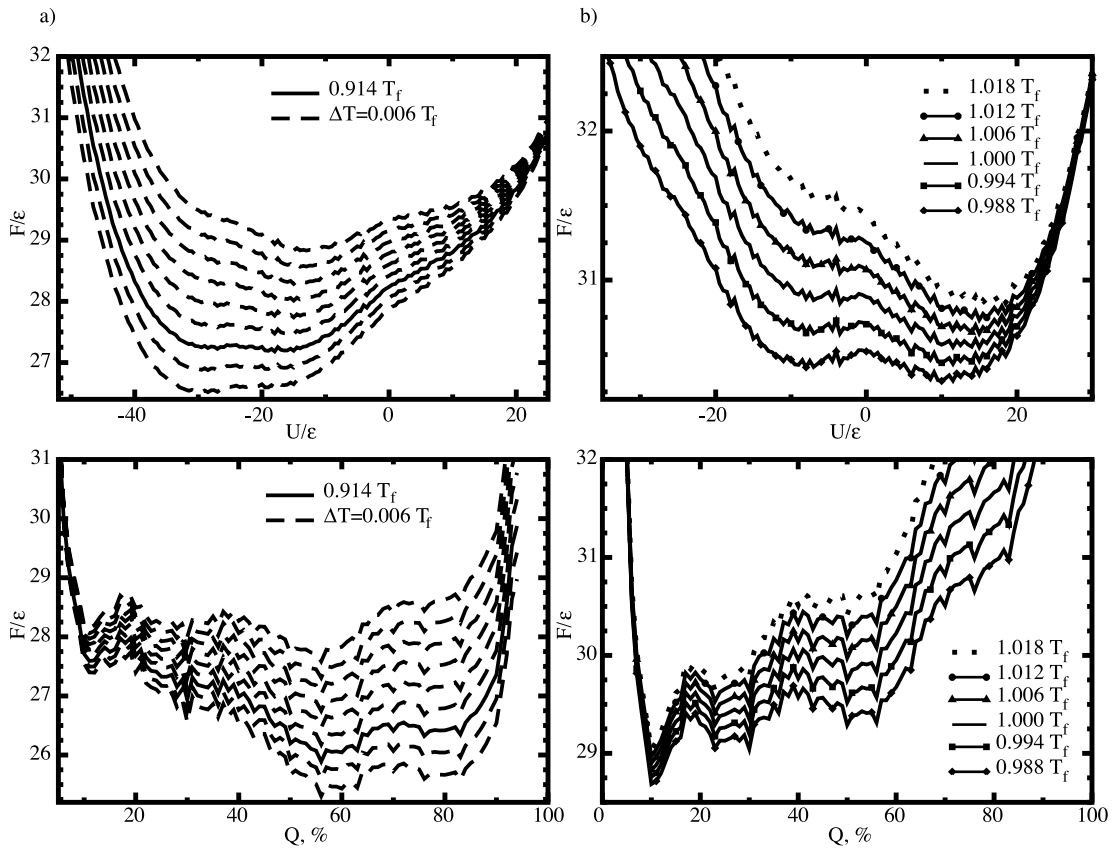


**Figure 7**. Free energy profiles for M-A35. All energies are given in units of ϵ and temperatures in units of $T_f$. (a) Native-to-intermediate transition observed at 0.914 $T_f$. Curves are drawn in intervals of $\Delta T = 0.006T_f$. The continuous lines correspond to $T = 0.914T_f$. (b) "Main" folding transition (i.e. intermediate-to-unfolded) at $T = 1.000T_f$.

is as follows: navy, non-native conformation; red, native contacts associated with helix H1; green, native contacts associated with helix H2; golden, native contacts associated with the β-sheet (consisting of strands S1, S2, S3, and S4); purple, native contacts associated with turn T1; and ice-blue, native contacts associated with turn T2. (Certain residues participate in the formation of different native contacts at different stages during folding. The "folded color" of these residues in the backbone trace representations changes to reflect that fact. For instance, V85 forms H2–S4 contacts at stage 4 of folding and is colored green in the $C^{\alpha}$ trace at that stage. At stage 5, V85 participates in

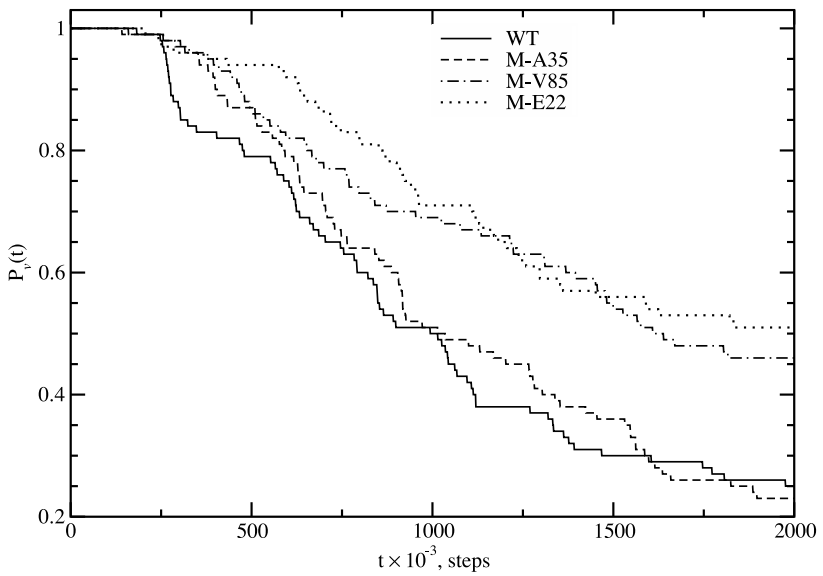


**Figure 8**. Early refolding kinetics of WT S6 and its three gatekeeper-lacking mutants, characterized by the fraction of unfolded trajectories as a function of time $P_v(t)$. For the WT, 75% of the refolding trajectories folded within $2.0 \times 10^6$ time steps (13.46 ns). An equivalent folding yield (78%) is observed for M-A35, the least significant experimentally suggested gatekeeper.[32] On the other hand, M-V85 and M-E22, for which efficient gatekeepers[32] have been removed, show a marked decrease in productive early folding, with only 54% and 49% of trajectories leading to the folded state, respectively. The refolding trends we observe are in very good agreement with the experimentally measured kinetics for S6 and its mutants.[32]

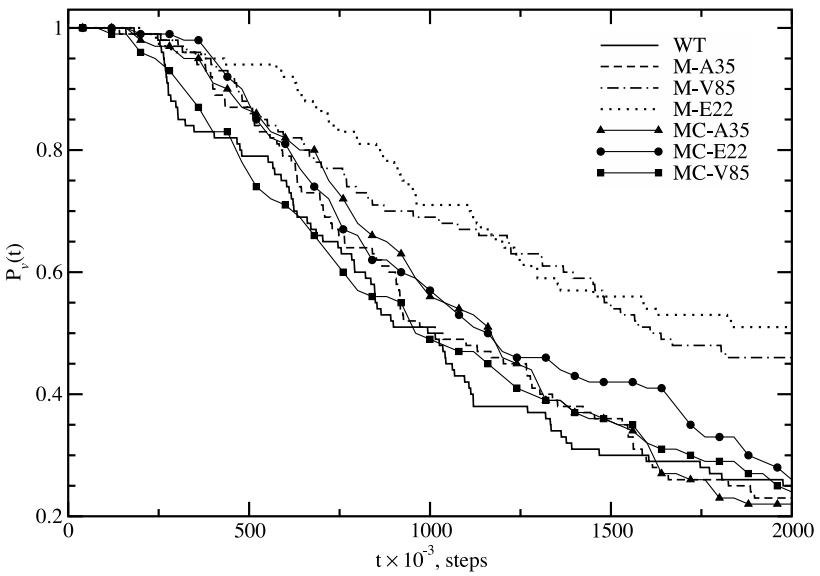


**Figure 9**. Comparison of the early folding of WT S6, M-V85, M-E22, and M-A35 (already shown in Figure 8) to the mutant controls MC-V85, MC-E22, and MC-A35. Folding progress is measured by the fraction of unfolded trajectories as a function of time $P_v(t)$. All three mutant controls exhibit folding yields that are essentially the same as those of the WT protein (i.e. the best anticipated folder of all considered systems) and the mutant of the "weakest" gatekeeper, M-A35. This demonstrates that the gatekeepers' locations, as well as the particular combination of long-range and short-range interactions in which they are involved, are essential to the protein's ability to choose productive folding pathways early on in the folding process.

additional native contacts between T2 and S4 and therefore its color changes to ice-blue. "Old folded colors" are retained at all times in the projection diagrams, e.g. T2, which contains residue V85, has both green and ice-blue sections at stage 5 of folding.)

We first discuss Figure 10 in terms of WT S6, and subsequently point out the differences in the folding mechanism of the gatekeeper-lacking mutants.

Short-range contacts are the ones that form most rapidly during refolding, as is shown in the first stage of events in Figure 10. These contacts involve the establishment of intrahelical structure within H1 and H2, and the formation of the turn between S3 and H2. The E22 gatekeeper is involved in 16.7% of H1–H1 interactions.

During the second stage of folding the hydrophobic core begins to form, starting at the turn between strands S2 and S3 then followed by the packing of H1 against S2 and S3. The gatekeeper A35 is involved in 14.3% of H1–S2 interactions and 7.5% of S2–S3 interactions. The slowest process at the second stage of folding is the formation of native contacts between H2 and T2 (i.e. the turn region between helix H2 and strand S4 in the native structure). Notably, all these contacts involve the strongest gatekeeper V85.

At the third stage of the folding process, helix H1 and strand S1 are completely associated into the core, after the correct formation of H1–T1 contacts and S1–H2, S1–T1, and S1–S3 interactions, respectively. H1 and H2 adopt native mutual orientation.

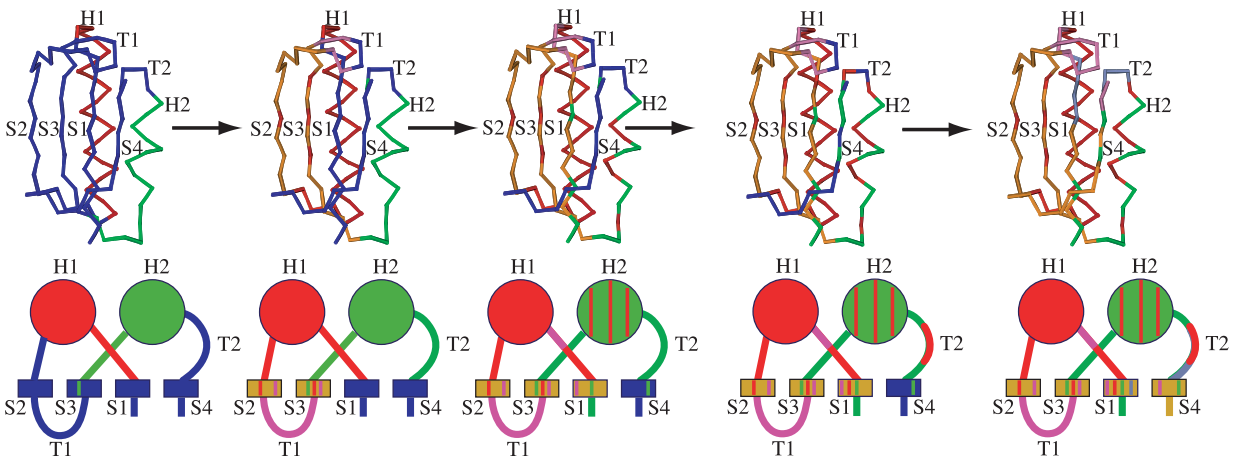


**Figure 10**. Folding mechanism of S6. Each stage in the folding process is represented by a backbone trace and a projection diagram (top view), in which residues along the protein sequence were colored depending on whether they had formed native contacts at that particular point during refolding as follows: navy, non-native conformation; red, native contacts associated with helix H1; green, native contacts associated with helix H2; golden, native contacts associated with the β-sheet (consisting of strands S1, S2, S3, and S4); purple, native contacts associated with turn T1; and ice-blue, native contacts associated with turn T2.

During the remaining two slowest stages of native contact formation, the two helices fall into native arrangement with respect to the β-strands (H1–T2 and H2–S4 contacts), and the incorporation of the entropically rich strands S1 and S4 into the hydrophobic core is completed (S1–H1, S1–S2, S1–S4, S1–T2, T1–S4, and S3–S4 contacts). The gatekeeper E22 is involved in 50.0% of the contacts that bring H1 and T2 together. V85 participates in 66.7% of S1–T2 contacts.

The above discussion elucidates why V85 is the most significant gatekeeper observed experimentally:[32] it exclusively controls the interactions between H2 and T2, reinforcing correct turn formation between helix H2 and strand S4, and is responsible for a large fraction of the native interactions between the entropically rich strand S1 and turn T2. V85 is followed in significance by E22,[32] which largely determines proper interaction between helix H1 and turn T2. V85 and E22 have significant impact on the proper packing of the protein. The guiding function of A35, on the other hand, is diminished in importance[32] because the absence of this gatekeeper is potentially compensated by a large number of other "A35 redundant" native hydrophobic core contacts between H1 and S2, and S2 and S3. Our analysis shows that the three "structural gatekeepers"[32] that guide the folding of S6 and are at the focus of the present theoretical investigation are not mutually redundant, i.e. they are not involved in native interactions between the same elements of secondary structure. (In terms of amino acid properties, their sidechains are either hydrophilic, or hydrophobic. On the other hand, "conformational gatekeepers",[33] which are thought to prevent aggregation by interrupting contiguous stretches of hydrophobic residues, have been experimentally found to reside close to each other in protein sequence and are all charged.[33])

Gatekeepers have a pronounced effect on the folding of S6. When they are eliminated, new dynamic bottlenecks, untypical of the WT, emerge. Removal of V85 alters the packing of the hydrophobic core. It drastically slows the formation of H2–T2 contacts at stage 2 of folding. This leads to faster and more compact packing of S2, S3, T1, and H1, creating a kinetic trap. The elimination of V85 also inhibits the interactions of S1 with H2 and S3 at stage 3 of folding. Instead, S1 is seen to associate faster with H1 than in WT S6, even though the gatekeeper is not involved directly in native contacts between these secondary structure elements. S1–T2 interactions are slower than WT for M-V85.

Mutation of E22 has negligible effect on H1–H1 intrahelical interactions at stage 1 of folding. The absence of this gatekeeper does not impact hydrophobic core assembly at stages 2 or 3; however, it somewhat slows the formation of H1–H2 native contacts at stage 3, even though the native list for those does not include E22. This is likely due to the slight disruption of H1 intrahelical native structure around the E22 position (about halfway through helix H1). At stage 4 of folding, the interactions of helix H1 with the core (i.e. T2) is hampered, and instead H1 packs against S1 prematurely, compared to the WT.

Elimination of A35 does not markedly change the folding mechanism of S6, and M-A35 behaves rather similarly to the WT species. The mutant controls MC-V85, MC-E22, and MC-A35 generally exhibit the same folding characteristics as WT S6, as well.

The above observations shed light on why and how gatekeepers steer S6 along productive folding pathways. The gatekeepers ensure that a "proper" sequence of folding events (secondary structure formation and tertiary packing) is followed. The gatekeepers' role appears to be closely related to topology: these special amino acid residues increase the protein's chances of sampling "productive" conformations as it moves along its energy landscape. Establishing a correlation between protein topology and folding rates has been the focus of a number of experimental and theoretical works to date. A concise summary of these efforts,[41] as well as a new variant of the topomer search model[41,42] and an application to systems with smooth energy landscapes[43] have been reported recently. These works suggest that the search for unfolded conformations with roughly native-like topology largely determines the height of the folding barrier, and hence the folding rate, for simple single-domain proteins. The gatekeeper perspective explains a mechanism that proteins have apparently "devised" to successfully find their native topomers.

The above discussion is based on observations regarding the ensembles of folded trajectories generated for each investigated species during kinetic refolding. Performing a similar analysis on the corresponding ensembles of unfolded trajectories provides additional insights into the role of gatekeepers. We have already pointed out that, according to our minimalist model of S6, the long-lived kinetic traps we observe are very similar in nature for both good folders (with gatekeepers) and bad folders (without gatekeepers). The important difference is that the species with gatekeepers fall into the kinetic traps much less frequently. Comparison of the WT S6 ensemble of unfolded trajectories to the unfolded ensembles of the two mutants that lack strong gatekeepers, M-V85 and M-E22, has revealed that gatekeepers prevent native but premature tertiary structure packing and thus ensure a smooth passage to the folded state. The unfolded ensembles of M-A35 and the mutant control species exhibit characteristics similar to those of the WT.

We conclude by summarizing our findings and pointing to future directions. We have designed and studied a minimalist Gō-like[38,39] model of S6 that successfully reproduces the experimentally observed folding characteristics[32] of this ribosomal protein. Our work sheds light on the mechanisms

by which the experimentally suggested structural gatekeepers of S6[32] ensure fast folding kinetics and guide the protein along productive folding pathways to the native state. We have demonstrated that the three gatekeepers V85, E22, and A35[32] have non-overlapping functions in ensuring proper folding by acting in different regions of S6. The gatekeepers regulate local secondary structure formation, as well as the overall gradual assembly of the protein during folding. By reinforcing productive topologies on the way to the native state, the gatekeepers prevent native but premature tertiary packing that can cause kinetic traps and hamper refolding. The strong correlation between efficient gatekeepers and protein topology is emphasized by the fact that a simplified Gō-type model is capable of reproducing gatekeeper-related effects. Additionally, our analysis elucidates that "gatekeeping strength" depends on the relative number of native contacts between two secondary structure elements that are controlled by the gatekeeper. The weakest gatekeeper, A35, is found to have the largest number of native contact alternatives, while the strongest gatekeeper, V85, entirely controls native turn formation between the second helix and the C-terminal strand of S6.

Our experience with designing structural gatekeepers into an all-β model protein,[34] combined with the findings of the present study of the ribosomal protein S6, have revealed some persistent trends that are beginning to outline the essential prerequisites for a good gatekeeper. This puts us in a position to theoretically predict possible gatekeeper sites for an arbitrary protein. As far as local interactions are concerned, we expect to find gatekeepers in turn regions, especially if the secondary structure elements that precede/follow the turn do not share numerous native contacts between residues located in the turn's immediate vicinity. Predictions regarding tertiary packing guided by long-range contacts are less obvious. In those instances, an analysis of the protein native contact map, combined with *in silico* observations of productive refolding of the WT is likely to yield additional clues regarding the location of effective gatekeepers. We are currently starting to test this approach on S6, and anticipate providing new gatekeeper candidates for experimental validation.

## Materials and Methods

### Model system

The 101 residue long ribosomal protein S6 (PDB accession code 1RIS.pdb[44]) is a mixed α–β protein that consists of two helices packed against a four-stranded β-sheet. Figure 1(a) shows a representation of S6 in which the helices are indicated by cylinders, and the β-strands by ribbons. (Image generated by VMD.[45]) We have constructed a minimalist off-lattice model representation of S6. The positions of residues 98–101 in the protein sequence were not experimentally observable in the electron density[44] and were not reported in 1RIS.pdb. Therefore, we included only residues 1–97 in our model. We first outline the general characteristics of the model and subsequently discuss each term in the Hamiltonian in detail.

As is typical of Gō-like[38,39] models, amino acid residues in our reduced model of S6 are represented by identical beads of mass $m$, connected by bonds of uniform length σ. Figure 1(b) identifies the association of a bead's number in the protein sequence with a secondary structure element, as revealed by 1RIS.pdb.[44] We define strands 1 (S1), 2 (S2), 3 (S3), and 4 (S4); helices 1 (H1), and 2 (H2); and turns 1 (T1), and 2 (T2), as shown in Figure 1(b). In 1RIS.pdb,[44] amino acid residues that were experimentally found to lie between subsequent secondary structure elements were not assigned to belong to either one of them. In our model, these "transitional" residues were incorporated in either helices or strands, depending on the dihedral angle potential (see below) that was used to model that particular stretch of the protein sequence.

In our reduced model, beads interact *via* bond, angle, and dihedral angle potentials, determined by chain connectivity. Additionally, each bead interacts with every other bead in the system *via* excluded volume, preventing the superposition of protein residues. A favorable long-range interaction is assigned to "native" pairs of beads, i.e. beads corresponding to residues that interact in the folded state of the protein. The list of native contacts (native bead pairs) was derived from the PDB crystal structure of S6,[44] based on either one of the following cutoff distance criteria:

$$r(C^{\alpha i}, C^{\alpha j}) \leq 7.5\ \text{Å} \tag{1}$$

where $r(C^{\alpha i}, C^{\alpha j})$ is the distance between $C^{\alpha}$ backbone atoms of residues $i$ and $j$, or:

$$r(\text{sch}^i, \text{sch}^j) \leq 4.0\ \text{Å} \tag{2}$$

where $r(\text{sch}^i, \text{sch}^j)$ is the distance between any two atoms that belong to the side-chains of residues $i$ and $j$, respectively. For both cutoffs above, the condition $j \geq i + 4$ was also imposed. A total of 215 native contacts for S6 were determined.

The Hamiltonian of the system is thus a sum of five terms:

$$\mathcal{H}(r, \theta, \phi) = V_{\text{LR}}(r) + V_{\text{EV}}(r) + V_{\text{B}}(r) + V_{\text{A}}(\theta) + V_{\text{D}}(\phi) \tag{3}$$

The long-range interactions between native contacts are calculated using:

$$V_{\text{LR}}(r_{ij}) = \sum_{\langle ij \rangle} \epsilon \left[ 5 \left( \frac{r_{ij}^0}{r_{ij}} \right)^{12} - 6 \left( \frac{r_{ij}^0}{r_{ij}} \right)^{10} \right] \tag{4}$$

The summation above goes over pairs of native contact residues $(i, j)$, and $r_{ij}$ is the distance between the $C^{\alpha}$ atoms of residues $i$ and $j$. The distance at which the function in equation (4) gives a minimum, $r_{ij}^0$, is specific to each native pair of beads and was extracted from 1RIS.pdb.[44] ε is the amount by which the system's energy is lowered upon the formation of one native contact, and is used to define the energy scale for the model. ε and the bond length σ define the reduced units utilized throughout this work. Energies are presented in units of ε, temperatures in units of $\epsilon/k_{\text{B}}$, lengths in units of σ, and time in units of $\tau = (m\sigma^2/\epsilon)^{1/2}$. For native contacts formed between residues that both belong to a β-strand of S6 (see Figure 1(b) for the locations of strands in the sequence), a value of $1/2\epsilon$ was used instead of ε in equation (4). This was done in order to avoid

overstabilizing of the protein hydrophobic core (formed by strands 2 and 3, and the turn between them) which concentrates a large number of native contacts (≈30%) in a relatively limited region of the protein.

Excluded volume interactions are modeled by:

$$V_{EV}(r) = \sum_{\langle ij \rangle} 1.061\epsilon \left( \frac{\sigma}{r_{ij}} \right)^{12} \quad (5)$$

where the summation goes over all pairs of atoms $(i, j)$, and $\sigma$ is the bond length between successive residues.

All bonds are held fixed at an equilibrium bond length value of $r_0 = 3.78$ Å using SHAKE.[46] Bond-angle interactions are harmonic:

$$V_A(\theta) = \sum_{\text{angles}} K_a(\theta - \theta_0)^2 \quad (6)$$

with $K_a$ set to 20 $\epsilon/\text{deg}^2$, and $\theta_0 = 105°$.

The dihedral potential describes rotations around three successive bonds involving four beads. Two different forms were used for the dihedral potential, depending on the type of secondary structure involved: α-helix or β-sheet (turns were modeled using the potential for β-sheets). The dihedral potential for helices is given by:

$$V_D^{\alpha}(\phi) = \sum_{\alpha} K_d[1 + \cos(\phi - \phi_0)] \quad (7)$$

where $K_d = 0.2\epsilon$, and $\phi_0 = 230°$. The choice of $\phi_0$ is based on the fact that the average dihedral angle from 1RIS.pdb for the two helices of S6 is approximately 50°. The summation is performed over all dihedral angles formed by beads that belong to the two helices. Dihedral interactions for β sheets (and turns) are of the form:

$$V_D^{\beta}(\phi) = \sum_{\beta} K_d[1 + \cos(3\phi)] \quad (8)$$

which gives minima of equal depth at dihedral angle values of 60° and 180°. The above summation is performed over all dihedral angles formed by beads that belong to β-strands or turns. These dihedral potentials do not specifically impose a constraint that would guarantee the natural "left-handedness" of the protein in the turn regions preceding the two helices. Both the left and right-handed isomer were considered to be "valid" outcomes of refolding in this work.

### Computer simulations

We have investigated models representing WT S6 and several mutants. Both the thermodynamics and early folding kinetics of these systems were characterized. All simulations were performed using the CHARMM package.[47]

#### *Constant-temperature thermodynamic sampling*

The thermodynamic properties of S6 and its mutants were calculated from data generated by constant temperature Nose–Hoover[48,49] molecular dynamics simulations using the Hamiltonian described above. The integration time step was set to 0.0075τ. For each system (WT or mutant), independent constant $T$ simulations were performed at 22–29 temperature values, corresponding to a temperature range of $0.149\epsilon/k_B$ to $0.547\epsilon/k_B$ for WT S6, and $0.0995\epsilon/k_B$ to $0.497\epsilon/k_B$ for the mutants. Low-$T$ and high-$T$ collection temperature points were spaced $0.0249\epsilon/k_B$ from each other, while temperature points around the folding transitions were spaced $0.00497\epsilon/k_B$ from each other. At temperatures that were far from the folding phase transition region, data were collected for $2.0 \times 10^6$ steps (13.46 ns), following $1.0 \times 10^5$ steps of equilibration, at $1.0 \times 10^3$ step intervals (to ensure the sampling of uncorrelated protein conformations). Around the folding transition, data collection involved up to $1.0 \times 10^7$ simulation time steps (67.30 ns) for WT S6, and $4.0 \times 10^6$ simulation time steps (26.92 ns) for the mutants, all at $1.0 \times 10^3$ step intervals. Thermodynamic properties were extracted from raw data using WHAM.[50–54] The simulation protocol provided good sampling, as measured by the relative error in the density of states[53] for our system:

$$\frac{\delta\Omega(X)}{\Omega(X)} = \sqrt{\frac{g}{\sum_{i=1}^{N} N_i(X)}} \quad (9)$$

where $X$ denotes either the order parameter $Q$, or the internal energy $U$. $g = 1$, $i$ runs over all constant-temperature simulations for a given system (WT or mutant), and $N_i(X)$ is the number of times the value $X$ was sampled during simulation $i$. Our simulation protocol ensured $\delta\Omega(Q)/\Omega(Q) \leq 5\%$, and $\delta\Omega(U)/\Omega(U) \leq 10\%$.

#### *Langevin dynamics*

To investigate the early folding kinetics of S6 and its mutants, Langevin molecular dynamics (MD) simulations were performed. The friction coefficient was set to 0.2 $\tau^{-1}$, and the integration time step to 0.0075 τ. Bond lengths were kept constant using the SHAKE[46] algorithm. Starting with the folded conformation, the system (WT S6 or mutant) was unfolded at high temperature $T \geq 0.597\epsilon/k_B$. The protein was subsequently quenched to a "refolding temperature", chosen slightly below the folding phase transition, as indicated by the divergence (in the context of a finite system) of the heat capacity, calculated from the thermodynamic simulations for the corresponding system. For each system (WT S6 or mutant), 100 independent refolding MD simulations of $2.0 \times 10^6$ steps (13.46 ns) each were performed. Data were collected at $1.0 \times 10^3$ step intervals (to ensure the sampling of uncorrelated protein conformations).

## Acknowledgements

A.D.S. thanks Leslie Chavez, Scott Brozell, Sandeep Patel, Wonpil Im, Mike Crowley, John Finke, and Margaret Cheung for helpful discussions. We are indebted to Mikael Oliveberg and Magnus Lindberg for stimulating conversations and insightful suggestions. Fellowship and equipment support received from the La Jolla Interfaces in Science Program sponsored by the Burroughs Wellcome Fund is gratefully acknowledged. The NSF-sponsored CTBP is supported by grants PHY-0216576 and 0225630. Additional financial support from the NIH for C.L.B. III (GM48807) and from the NSF for J.N.O. (NCB-0084797) is also gratefully acknowledged.

## References


1. Brooks, C. L., III, Onuchic, J. N. & Wales, D. J. (2001). Statistical thermodynamics—taking a walk on a landscape. *Science*, **293**, 612–615.
2. Huang, G. S. & Oas, T. G. (1995). Submillisecond folding of monomeric λ repressor. *Proc. Natl Acad. Sci. USA*, **92**, 6878–6882.
3. Feng, Y., Siglar, S. G. & Wand, A. J. (1994). Solution structure of apocytochrome $b_{562}$. *Nature Struct. Biol.* **1**, 30–35.
4. Sosnick, T. R., Hiller, R. & Englander, S. W. (1994). The barriers in protein folding. *Nature Struct. Biol.* **1**, 149–156.
5. Elöve, G. A., Bhuyan, A. K. & Roder, H. (1994). Kinetic mechanism of cytochrome *c* folding: involvement of the heme and its ligands. *Biochemistry,* **33**, 6925–6935.
6. Jennings, P. A. & Wright, P. E. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science*, **262**, 892–896.
7. Balbach, J., Forge, V., Lau, W. S., van Nuland, N. A. J., Brew, K. & Dobson, C. M. (1996). Protein folding monitored at individual residues during a two-dimensional NMR experimant. *Science,* **274**, 1161–1163.
8. Fersht, A. R., Matouschek, A. & Serrano, L. (1992). The folding of an enzyme I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224**, 771–782.
9. Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995). The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation–condensation mechanism for protein folding. *J. Mol. Biol.* **254**, 260–288.
10. López-Hernández, E. & Serrano, L. (1996). Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, CI-2. *Fold. Des.* **1**, 43–55.
11. Ballew, R. M., Sabelko, J. & Gruebele, M. (1996). Direct observation of fast protein folding: the initial collapse of apomyoglobin. *Proc. Natl Acad. Sci. USA*, 5759–5764.
12. Pascher, T., Chesick, J. P., Winkler, J. R. & Gray, H. B. (1996). Protein folding triggered by electron transfer. *Science*, **271**, 1558–1560.
13. Plaxco, K. W. & Dobson, C. M. (1996). Time-resolved biophysical methods in the study of protein folding. *Curr. Opin. Struct. Biol.* **6**, 630–636.
14. Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994.
15. Schuler, B., Lipman, E. A. & Eaton, W. A. (2002). Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature*, **419**, 743–747.
16. Eaton, W. A. *et al.* (2000). Fast kinetics and mechanisms in protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 327–359.
17. Bryngelson, J. D. & Wolynes, P. G. (1987). Spin glasses and the statistical mechanics of protein folding. *Proc. Natl Acad. Sci. USA*, **84**, 7524–7528.
18. Bryngelson, J. D. & Wolynes, P. G. (1989). Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem.* **93**, 6902–6915.
19. Leopold, P. E., Montal, M. & Onuchic, J. N. (1992). Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl Acad. Sci. USA*, **89**, 8721–8725.
20. Sali, A., Shakhnovich, E. & Karplus, M. (1994). How does a protein fold? *Nature*, **369**, 248–251.
21. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Struct. Funct. Genet.* **21**, 167–195.
22. Boczko, E. M. & Brooks, C. L., III (1995). First-principles calculation of the folding free energy of a three-helix bundle protein. *Science*, **269**, 393–396.
23. Onuchic, J. N., Wolynes, P., Luthey-Schulten, Z. & Socci, N. D. (1995). Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl Acad. Sci. USA*, **92**, 3626–3630.
24. Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. (1997). Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* **48**, 545–600.
25. Shakhnovich, E. I. (1997). Theoretical studies of protein-folding thermodynamics and kinetics. *Curr. Opin. Struct. Biol.* **7**, 29–40.
26. Chan, H. S. & Dill, K. A. (1998). Protein folding in the landscape perspective: chevron plots and non-arrhenius kinetics. *Proteins: Struct. Funct. Genet.* **30**, 2–33.
27. Thirumalai, D. & Klimov, D. K. (1999). Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models. *Curr. Opin. Struct. Biol.* **9**, 197–207.
28. Dobson, C. M. & Karplus, M. (1999). The fundamentals of protein folding: bringing together theory and experiment. *Curr. Opin. Struct. Biol.* **9**, 92–101.
29. Onuchic, J. N., Nymeyer, H., García, A. E., Chahine, J. & Socci, N. (2000). The energy landscape theory of protein folding: insights into folding mechanisms and scenarios. *Advan. Protein Chem.* **53**, 87–152.
30. Shea, J.-E. & Brooks, C. L., III (2001). From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* **52**, 499–525.
31. Guo, Z. & Thirumalai, D. (1995). Kinetics of protein folding: nucleation mechanism, time scales, and pathways. *Biopolymers*, **36**, 83–102.
32. Otzen, D. E. & Oliveberg, M. (1999). Salt-induced detour through compact regions of the protein folding landscape. *Proc. Natl. Acad. Sci. USA*, **96**, 11746–11751.
33. Otzen, D. E., Kristensen, O. & Oliveberg, M. (2000). Designed protein tetramer zipped together with a hydrophobic Alzheimer homology: a structural clue to amyloid assembly. *Proc. Natl Acad. Sci. USA*, **97**, 9907–9912.
34. Stoycheva, A. D., Onuchic, J. N. & Brooks, C. L., III (2003). Effect of gatekeepers on the early folding kinetics of a model β-barrel protein. *J. Chem. Phys.* **119**, 5722–5729.
35. Honeycutt, J. D. & Thirumalai, D. (1992). The nature of folded states of globular proteins. *Biopolymers*, **32**, 695–709.
36. Guo, Z. & Brooks, C. L., III (1997). Thermodynamics of protein folding: a statistical mechanical study of a small all-β protein. *Biopolymers*, **42**, 745–757.
37. Nymeyer, H., García, A. E. & Onuchic, J. N. (1998). Folding funnels and frustration in off-lattice minimalist protein landscapes. *Proc. Natl Acad. Sci. USA*, **95**, 5921–5928.
38. Ueda, Y., Taketomi, H. & Gō, N. (1975). Studies on

protein folding, unfolding and fluctuations by computer simulation. I. The effects of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Res.* **7**, 445–459.
39. Ueda, Y., Taketomi, H. & Gō, N. (1978). Studies on protein folding, unfolding and fluctuations by computer simulation. II. A three-dimensional lattice model of lysozyme. *Biopolymers*, **17**, 1531–1548.
40. Kussell, E., Shimada, J. & Shakhnovich, E. I. (2003). Side-chain dynamics and protein folding. *Proteins: Struct. Funct. Genet.* **52**, 303–321.
41. Makarov, D. E. & Plaxco, K. W. (2003). The topomer search model: a simple, quantitative theory of two-state protein folding kinetics. *Protein Sci.* **12**, 17–26.
42. Makarov, D. E., Keller, C. A., Plaxco, K. W. & Metiu, H. (2002). How the folding rate constant of simple, single-domain proteins depends on the number of naitve contacts. *Proc. Natl Acad. Sci. USA*, **99**, 3535–3539.
43. Jewett, A. I., Pande, V. S. & Plaxco, K. W. (2003). Cooperativity, smooth energy landscapes and the origins of topology-dependent protein folding rates. *J. Mol. Biol.* **326**, 247–253.
44. Sedelnikova, S. E., Agalarov, S. C., Eliseikina, I. A. *et al.* (1991). Crystals of protein S6 from the 30 S ribosomal subunit of *Thermus thermophilus*. *J. Mol. Biol.* **220**, 549–550.
45. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD—visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38.
46. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341.
47. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4**, 187–217.
48. Nosë, S. (1984). A unified formulation of the constant temperature molecular dynamics method. *J. Chem. Phys.* **81**, 511–519.
49. Hoover, W. G. (1985). Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. ser. A*, **31**, 1695–1697.
50. Ferrenberg, A. M. & Swendsen, R. H. (1988). New monte carlo technique for studying phase transitions. *Phys. Rev. Letters*, **61**, 2635–2638.
51. Ferrenberg, A. M. (1989). Efficient use of Monte Carlo simulation data. PhD Thesis, Carnegie Mellon University, Pittsburgh, PA.
52. Ferrenberg, A. M. & Swendsen, R. H. (1989). Optimized Monte Carlo data analysis. *Phys. Rev. Letters*, **63**, 1195–1198.
53. Kumar, S., Bouzida, D., Swendsen, R. H. *et al.* (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021.
54. Roux, B. (1995). The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* **91**, 275–282.

***Edited by C. R. Matthews***

*(Received 10 February 2004; received in revised form 20 April 2004; accepted 29 April 2004)*