Changes of Protein Folding Pathways by Circular Permutation

OVERLAPPING NUCLEI PROMOTE GLOBAL COOPERATIVITY

Received for publication, March 5, 2008, and in revised form, June 5, 2008. Published, JBC Papers in Press, June 18, 2008, DOI 10.1074/jbc.M801776200

Ellinor Haglund1, Magnus O. Lindberg1, and Mikael Oliveberg2

From the Department of Biochemistry and Biophysics, Arrhenius Laboratories of Natural Sciences, Stockholm University, S-106 91 Stockholm, Sweden

The evolved properties of proteins are not limited to structure and stability but also include their propensity to undergo local conformational changes. The latter, dynamic property is related to structural cooperativity and is controlled by the folding-energy landscape. Here we demonstrate that the structural cooperativity of the ribosomal protein S6 is optimized by geometric overlap of two competing folding nuclei: they both include the central β-strand 1. In this way, folding of one nucleus catalyzes the formation of the other, contributing to making the folding transition more concerted overall. The experimental evidence is provided by an extended set of circular permutations of S6 that allows quantitative analysis of pathway plasticity at the level of individual side chains. Because similar overlap between competing nuclei also has been discerned in other proteins, we hypothesize that the coupling of several small nuclei into extended “supernuclei” represents a general principle for propagating folding cooperativity across large structural distances.

Proteins are highly evolved molecules that are selected and optimized by the functional requirements of the cell. The optimization, however, works not only on the native structures but also involves their physical properties (1). A well known example is protein stability, i.e., the equilibrium between folded states and the denatured counterparts, that needs to be maintained within certain limits to assure functionality and yet allow degradation. Protein sequences are moreover tuned to resist aggregation and erroneous interactions with other biomolecules. In part such tuning is achieved by electrostatic repulsion (2, 3) and negative design, i.e., side chains in the form of gatekeepers that specifically reduce sequence stickiness (4) or otherwise obstruct unwanted interactions between or within proteins (4–8). At a more generic level, protein aggregation seems also to be prevented by the folding potential itself. The factors that govern structural specificity disfavor automatically non-native contacts between molecules and bias the proteins to aggregate by native-like interactions (4, 9, 10). The archetypal example of such aggregates is the domain swap (11, 12). A seemingly different type of aggregation is the ordered assembly into fibrils driven by sequence signatures that allow linear propagation of non-native β-structure (13). In either case, aggregation relies on the exposure of amino acid segments that are normally hidden within the native structure, for example by local unfolding. There is growing evidence that proteins suppress such detrimental unfolding events by increasing the unfolding cooperativity (14–16). It is interesting to note that, unlike native structure and stability, cooperativity is a property that is primarily determined by the folding-energy landscape (14). The principle can be illustrated as follows. If the folding-energy landscape allows only one folding route, thermal motions of the native state would repeatedly lead to excursions along the same unfolding trajectory. Such narrow sampling of the conformational space could make the affected parts of the protein loose and susceptible to local unfolding by even modest fluctuations (14). If in contrast the energy landscape is tuned to have a broad, diffuse folding progression there is no preferred way of unfolding the protein: each microscopic unfolding attempt involves different parts of the structure (17). In effect, this uniform probability of unfolding will produce a native structure that is relatively “rigid” because the simultaneous occupancy of partly unfolded states that are matched for aggregation is minimized. From experimental studies of small two-state proteins it is apparent that most natural proteins belong to the latter category (18). Folding seems optimized for high cooperativity manifested in transition-state structures, i.e., folding nuclei, that resemble diffuse versions of the native structures (14). Consistent with the idea that the diffuse transition-state structure is a biological adaptation, it can readily be changed by circular permutation into polarized, less cooperative counterparts where half of the protein is fully structured and the other half is disordered (19, 20). Also the transition-state structure responds readily to extensive changes in the side-chain contacts as observed in comparative studies of sequence-divergent homologs (21–23). Even so, the factors governing folding cooperativity are yet poorly understood, and cooperativity is also the molecular trait of naturally evolved proteins that is most difficult to reproduce in silico and by de novo design. One possibility is that cooperativity is achieved by balancing long sequence separation between interacting residues with strong contacts. Such equalization of the folding probability across the protein structure has been seen to promote diffuse nuclei in both computational (17, 24, 25) and experimental studies (14). Consistently circular permutations of S6 and SH3 show that the contribution of individual side chains to the folding nucleus is directly related to changes in sequence separation of the interacting side chains, implicating that chain entropy is indeed a
responsive factor for tuning the folding trajectory. But there is more to it. The plasticity of the S6 folding reaction seems restrained to two competing pathways corresponding to nucleation in either half of the protein structure (19, 20). It is further apparent that the competing nuclei of S6 are to some degree overlapping, providing a clue to how the folding cooperativity can be extended from one part of the structure to another (20): strand 1 (β1) in the center of the β-sheet is part of both nuclei (see Fig. 1). To shed further light on how this putative coupling of cooperativity works, here we specifically investigate the behavior of the nuclei overlap. The study is based on ϕ-value analysis of a new, expanded set of S6 permutants that includes incisions between all six secondary structure elements. The fate of each part of the S6 structure is thus monitored and compared on six different settings of the folding-energy landscape, yielding the number of interactions with each of the two S6 nuclei. One side of β1 is part the α1 nucleus, whereas the other is part of the α2 nucleus. Thus, when folding shifts between the α1 and α2 nuclei, the β1 contacts simply shift from one side to the other with little effect on the macroscopic ϕ-values. From a general perspective, this balanced shift of nucleating contacts around β1 provides a concrete example of how structural order can be propagated from one nucleus to the other without the accumulation of partly structured intermediates. On this basis, we conclude that folding cooperativity is a modular property that can be extended through protein structures by multiple, competing, folding nuclei that share structural overlap.

**MATERIALS AND METHODS**

**Protein Engineering**—All circular permutants of S6 are labeled according to the position of the incisions, and their amino acids are numbered according to the wild-type sequence, Protein Data Bank code 1RIS (26). The wild-type numeration was used for all point mutations in this study. The circular permutants P13–14 and P68–69 were designed and constructed as described previously (27), and the circular permutants P33–34, P81–82, P54–55, P33–34, and P81–82 were designed and constructed according to the procedures in Ref. 19. The genes designed for P33–34 and P81–82 were purchased from Entelechon. By these five constructs, all possible incisions between the secondary structure elements of S6 have been covered. The point mutations used for mapping out the interactions in the transition-state ensembles of the permuted proteins were chosen to cover all parts of the hydrophobic core.

**Protein Expression and Purification**—Mutations were performed with the QuikChange site-directed mutagenesis kit (Stratagene), oligonucleotides were purchased from DNA Technology, and all mutations were confirmed by sequencing (eurofins mug operon). The mutant proteins were transformed and overexpressed in competent *Escherichia coli* strain BL21 or C41 DE3 and then purified by two-step precipitation followed by cation-exchange chromatography (CM-Sepharose) and gel filtration (Sephacryl S-100) as described previously (28). The buffer was 50 mM Tris, pH 7.5. The identity of the purified protein was confirmed by mass spectroscopy.

**Kinetic Measurements**—Stopped-flow measurements and curve fitting were performed on SX-18MV and PiStar instruments (Applied Photophysics, Leatherhead, UK). The excitation wavelength was 280 nm, and the emission was collected with a 305-nm-cutoff filter. Mixing was 1 + 10, and the final protein concentration was 0.8 μM. All measurements were conducted at 25 °C in 50 mM MES3 at pH 6.3 (Sigma) using GdmCl as denaturant (UltraPure, Invitrogen).

**Chevron Analysis**—To minimize the effect of chevron curves at low and high GdmCl concentrations, the kinetic m-values were derived from the linear regime at the bottom of the chevron plots, i.e. midpoint ± 2 M, using the standard equation,

\[
\log k_{\text{obs}} = \log k_f + k_u = \log(10^{(\log k_f^{\text{obs}} + m_i[GdmCl])} + 10^{(\log k_u^{\text{obs}} + m_i[GdmCl])})
\]

where \(k_f^{\text{obs}}\) and \(k_u^{\text{obs}}\) are the refolding and unfolding rate constants at \([GdmCl] = 0 \text{ M}\), and \(m_i\) and \(m_u\) are the slopes of the refolding and unfolding chevron limbs, respectively. Data analysis was performed with the software KaleidaGraph 4.0 and GraphPad Prism 4.0.

**ϕ-Value Analysis**—By systematically truncating side chains while measuring the effects on the folding and unfolding kinetics it is possible to map out the interaction patterns in the transition-state ensemble (30). In essence, mutations that slow down the refolding reaction are considered to target stabilizing contacts in its structure. The strengths of these interactions are measured by the ϕ-values. A ϕ-value of 0 indicates that the targeted side chain experiences a denatured-like environment in the transition-state ensemble, whereas a ϕ-value of 1 indicates that it has a fully native-like environment. Fractional values of ϕ are thus taken to indicate the degree of native-like interactions, i.e. ϕ = 0.5 indicates that the truncated side-chain moieties form half of the native contacts in the folding transition state. To reduce the effects of extrapolation errors and transition-state shifts occurring at high [GdmCl] (29), the ϕ-values (30) were calculated from chevron fits (Equation 1) according to Equation 2,

\[
\phi = \frac{(\log k_f^{\text{obs}} - \log k_f^{\text{mut, A}})}{(\log k_f^{\text{obs}} - \log k_f^{\text{mut, A}}) - (\log k_f^{\text{mut, B}} - \log k_f^{\text{mut, E}})}
\]

where \(A = 1 \text{ M}\) and \(B = 4 \text{ M}\), referring to the GdmCl concentrations at which the refolding and unfolding rate constants were taken. The change of ϕ upon circular permutation was then calculated according to Equation 3.

\[
\Delta \phi = \phi^{\text{wild type}} - \phi^{\text{permutant}}
\]

In the cases where the transition midpoints were too low to allow precise fits of the refolding limb, i.e. midpoints <2 M GdmCl, \(m_f\) was locked to the average value of the permutant in

\[^{3}\text{The abbreviations used are: MES, 4-morpholineethanesulfonic acid; GdmCl, guanidinium chloride; wt, wild type.}\]
Circular Permutants of S6

The weighted parameter $L_w$ encompasses the influence from sequence neighbors within a given secondary structure element. That is, amino acid $X$ is linked to the “actions” of its sequence neighbors $X - 1$ and $X + 1$ according to Equation 6.

$$L_w^X = \frac{L^{(X-1)} + (2 \times L^X) + L^{(X+1)}}{4}$$

(Eq. 6)

For mutations at the edges of the secondary structure elements either $L^{(X-1)}$ or $L^{(X+1)}$ was set to 0.

“Bootstrap” Analysis of the Correlation between $\Delta \phi$ and $\Delta L_w$—To investigate the contribution of individual data points to the $\Delta L_w$ versus $\Delta \phi$ correlation we performed a “bootstrap-type” analysis. In the analysis, a sliding window of three consecutive mutations along the S6 sequence was omitted, and for each step the correlation ($r$) between $\Delta \phi$ and $\Delta L_w$ was recalculated.

RESULTS

The Effect of Circular Permutation on Folding Kinetics and Stability—The stopped-flow data show that the stability alterations following circular permutation of S6 are mainly exerted through increased unfolding rate constants ($k_u$) (Fig. 1 and Tables 1–6). The accompanying changes in the refolding rate constants ($k_f$) are smaller overall and indicate for several of the permutants a stabilization of the transition-state ensemble (‡). A minimalist interpretation of these results is that the circular permutation involves two energetic components that are partly compensating: first, a structural stabilization due to the linkage of the N and C termini by a well designed hairpin (27), and second, an energetic penalty from cutting up the backbone in a new position. In the case of P$^{54-55}$, where the incision is done in a flexible loop, the effect of the new N and C termini linkage dominates and is manifested in an overall stabilization (Fig. 1 and Tables 1–6). Notably the kinetic $m$-values and transition-state placement ($\beta^3$) remain relatively unaffected by circular permutation with the exception of P$^{33-34}$ that shows decreased $m_w$ (Fig. 1 and Tables 1–6). The folding nuclei of the different S6 permutants seem thus to maintain a similar solvent-accessible surface area despite considerable redistributions of the transition-state contacts as indicated by the $\phi$-value data below (Figs. 2 and 3). An interesting detail of P$^{81-82}$ is that this permutant seems to dimerize in its native state at low GdmCl concentrations, giving rise to an addition unfolding phase (data not shown). For the purpose of this study, however, the complex kinetics of P$^{81-82}$ was resolved by unfolding the permutant at 0.8 M GdmCl where it remains monomeric.

Calculation of Changes in Sequence Separation, $\Delta L$—The average sequence separation between contacts lost ($L$) upon point mutation (14) was calculated from the Protein Data Bank structures of S6$^{wt}$ and the circular permutants according to Equation 4,

$$L = \frac{\sum L'}{n}$$

(Eq. 4)

where $L'$ is the sequence separation (loop length) between the individual carbon-carbon contacts (within a radius of 6 Å) lost upon mutation, and $n$ is the total number of contacts lost. To emphasize the contribution of tertiary contacts, four residues on either side of the target side chain were excluded from the calculation. The change of sequence separation upon circular permutation was then calculated as follows.

$$\Delta L = L^{\text{wild type}} - L^{\text{permutant}}$$

(Eq. 5)
**Circular Permutants of S6**

$\phi$-Values Reveal Systematic Alterations of the Folding Process—To facilitate the comparison of mutant data of the different S6 constructs, the amino acid positions of the circular permutuations have been numbered according to the wild-type sequence. The effects of mutation on the chevron plots of the individual S6 constructs are shown in Fig. 2, and the corresponding $\phi$-values (30) are mapped out structurally in Fig. 3. The probe mutations and numerical data are listed in Tables 1–6. On the whole, the results show that the S6 $\phi$-values respond readily to circular permutation and that the response of the individual S6 permutuants is different. The effects are clearly seen by comparing the $\phi$-value changes at the level of secondary structure elements. For example, the $\phi$-values of $\beta_4$ are low overall in S6wt, P68–69, and P81–82 but considerably higher in $\beta_3$ (Fig. 3). Notably the latter group also shows the most pronounced increase of $\phi$ in Fig. 1. A similar pattern is observed for the adjacent $\alpha_2$, whereas $\alpha_1$ at the opposite side of the proteins shows the reversed response by having the highest $\phi$-values in S6wt, P88–89, and P81–82. It is also apparent that the $\phi$-values of the central $\beta$-strand 1 remain quite similar around 0.5 in all permutuants, thus presenting a pivot point around which the distribution of nucleating contacts can move to either side of the S6 structure depending on detailed wiring of the polypeptide backbone. In addition to this general redistribution of the nucleating contacts upon circular permutation we note that an isolated cluster of neighboring side chains in P13–14 shows $\phi$-values that are anomalously high, i.e. V75A and V88A (Table 2). The cause of the anomaly is not

### Table 1

| Process | $\phi$-Values | $\phi_{element}$ | $\Delta L_d$ | $\Delta L_w$ |
|---------|---------------|-----------------|-------------|-------------|
| Wild type | 1.38 | 1.34 | | |
| V6A | 1.38 | 1.34 | | |
| L10A | 1.38 | 1.34 | | |
| L19A | 1.38 | 1.34 | | |
| L26A | 1.38 | 1.34 | | |
| L30A | 1.38 | 1.34 | | |
| V37A | 1.38 | 1.34 | | |
| V40A | 1.38 | 1.34 | | |
| L64A | 1.38 | 1.34 | | |
| F60A | 1.38 | 1.34 | | |
| L61A | 1.38 | 1.34 | | |
| V63A | 1.38 | 1.34 | | |
| V65A | 1.38 | 1.34 | | |
| V72A | 1.38 | 1.34 | | |
| L75A | 1.38 | 1.34 | | |
| L81A | 1.38 | 1.34 | | |
| V85A | 1.38 | 1.34 | | |
| V88A | 1.38 | 1.34 | | |

* Value set to average $m_6$ because MP < 2 AU.
* Data excluded because $\Delta L_d < 0.70$. 

### Table 2

| Process | $\phi$-Values | $\phi_{element}$ | $\Delta L_d$ | $\Delta L_w$ |
|---------|---------------|------------------|-------------|-------------|
| Wild type | 1.38 | 1.34 | | |
| V6A | 1.38 | 1.34 | | |
| L10A | 1.38 | 1.34 | | |
| L19A | 1.38 | 1.34 | | |
| L26A | 1.38 | 1.34 | | |
| L30A | 1.38 | 1.34 | | |
| V37A | 1.38 | 1.34 | | |
| V40A | 1.38 | 1.34 | | |
| L64A | 1.38 | 1.34 | | |
| F60A | 1.38 | 1.34 | | |
| L61A | 1.38 | 1.34 | | |
| V63A | 1.38 | 1.34 | | |
| V65A | 1.38 | 1.34 | | |
| V72A | 1.38 | 1.34 | | |
| L75A | 1.38 | 1.34 | | |
| L81A | 1.38 | 1.34 | | |
| V85A | 1.38 | 1.34 | | |
| V88A | 1.38 | 1.34 | | |

* Value set to average $m_6$ because MP < 2 AU.
* Data excluded because $\Delta L_d < 0.70$. 

**Notes:**
- $\beta_4$ = $m_j/(m_j - m_i)$.
- $\phi$ is calculated from $\log k_i^{14}$ and $\log k_i^{15}$ (Equation 1), and $\Delta L_d$ and $\Delta L_w$ are calculated according to Equations 4–6. $\Delta G_{D_N}$ and $\Delta G_{W_N}$ are in units of kcal/mol. $\phi_{element}$ is the average $\phi$-value for the individual secondary structure elements.
Influence of Sequence Separation and Backbone Neighbors

From the mutant data in Figs. 2 and 3, it is apparent that circular permutants of the S6 structure leads to significant changes of positions 75 and 88 in P13–14. To examine these φ-value changes quantitatively, we have previously modeled the effect of circular permutation by the site-specific parameter ΔL (Equations 4 and 5) (19). In some analogy with the global parameter contact order (32), ΔL provides a measure of sequence separation between side-chain contacts but not for the individual long range contacts probed by the φ-value mutations (14). The simplistic analysis shows that the φ-values increase in positions where ΔL goes down and diminish where ΔL goes up. The correlation is not unique for S6 but has also been resolved for permutants of the SH3 domain (19), suggesting that the φ-value response to increased sequence separation is general: if the loops between individual contact pairs are extended, then their contribution to the folding nucleus is reduced. In this study, the addition of the final permutants...
TABLE 5

| mp86–69 | Pseudo wt | V6A | I8A | L10A | L19A | I26A | L30A | V37A | V40A | L48A | F60A | L61A | V37A | V40A | L48A | F60A |
|---------|-----------|-----|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|
| m^3 | -1.27 | -1.54 | -1.42 | -1.50 | -1.15 | -1.21 | -1.20 | -1.10 | -1.24 | -1.22 | -1.18 | -1.22 | -1.24 | -1.24 | -1.18 |
| m^2 | -2.52 | -2.76 | -2.78 | -2.66 | -2.86 | -2.93 | -2.93 | -2.86 | -2.86 | -2.95 | -2.78 | -2.86 | -2.86 | -2.86 | -2.86 |
| m | 0.65 | 0.66 | 0.61 | 0.64 | 0.60 | 0.63 | 0.63 | 0.61 | 0.66 | 0.54 | 0.55 | 0.70 | 0.49 | 0.60 | 0.63 |
| m^2 | 2.72 | 1.59 | 1.90 | 1.88 | 2.54 | 2.56 | 1.76 | 1.41 | 1.76 | 1.21 | 1.84 | 1.22 | 3.88 | 1.30 | 0.63 |
| m^3 | 5.24 | 3.17 | 4.30 | 2.32 | 5.29 | 4.13 | 3.62 | 3.57 | 4.94 | 1.76 | 4.61 | 3.88 | 4.06 | 6.15 | 4.68 |
| log k^H2O | -0.15 | 0.30 | 0.42 | 0.57 | 0.18 | 0.43 | 0.86 | 0.17 | 1.46 | 0.25 | 0.35 | 0.59 | 1.17 | 1.50 | 0.35 |
| log k^H2 | 1.45 | 0.40 | 0.37 | 0.57 | 0.20 | 1.11 | 1.40 | 1.30 | 1.46 | 1.24 | 1.42 | 0.77 | 1.76 | 1.72 | 1.50 |
| log k^H2O | 0.08 | 0.40 | 0.38 | 0.30 | 0.21 | 0.35 | 0.31 | 0.30 | 0.29 | 0.29 | 0.40 | 0.12 | 0.12 | 0.35 | 0.27 |

**Table 6**

| mp81–82 | Pseudo wt | V6A | I8A | L10A | L19A | I26A | L30A | V37A | V40A | L48A | F60A | L61A | V37A | V40A | L48A | F60A |
|----------|------------|-----|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|
| m^3 | -1.20 | -1.24 | -1.20 | -1.20 | -1.90 | -1.24 | -1.24 | -1.24 | -1.24 | -1.24 | -1.24 | -1.24 | -1.24 | -1.24 | -1.24 |
| m^2 | -1.08 | -0.11 | -0.38 | -0.82 | -0.82 | -0.82 | -0.82 | -0.82 | -0.82 | -0.82 | -1.09 | -1.63 | -0.97 | -0.97 | -0.97 |
| m | 0.70 | 0.84 | 0.59 | 1.00 | 0.59 | 0.59 | 0.59 | 0.59 | 0.59 | 0.59 | 1.90 | 1.90 | 1.90 | 1.90 | 1.90 |
| m^2 | 1.89 | 2.04 | 1.78 | 2.02 | 1.84 | 1.76 | 1.76 | 1.76 | 1.76 | 1.76 | 1.89 | 1.89 | 1.89 | 1.89 | 1.89 |
| m^3 | 3.72 | 2.22 | 3.40 | 2.39 | 3.40 | 3.40 | 3.40 | 3.40 | 3.40 | 3.40 | 3.40 | 3.40 | 3.40 | 3.40 | 3.40 |
| log k^H2O | -0.60 | 0.40 | 0.45 | 0.43 | 0.39 | 0.43 | 0.43 | 0.43 | 0.43 | 0.43 | 0.43 | 0.43 | 0.43 | 0.43 | 0.43 |
| log k^H2 | 1.44 | 1.72 | 1.36 | 2.09 | 1.71 | 1.74 | 1.74 | 1.74 | 1.74 | 1.74 | 1.74 | 1.74 | 1.74 | 1.74 | 1.74 |
| log k^H2O | 1.70 | 2.21 | 2.70 | 4.68 | 2.70 | 2.70 | 2.70 | 2.70 | 2.70 | 2.70 | 2.70 | 2.70 | 2.70 | 2.70 | 2.70 |

**Notes:**
- a Value set to average m because MP < 2 m.
- b Data excluded because ΔG(DN) < 0.70.

Circular Permutants of S6

33–34 and p81–82 makes it possible to investigate this effect also at the level of individual sequence positions. Overall the P33–34 and p81–82 φ-values fall nicely within the scattered data of the other permutants but do not contribute to improve the correlation with ΔL. Rather the global value of r decreases from 0.78 (19) to 0.62 (Fig. 4). The reason for this apparent reversal is in part that some of the most influential data points at positive ΔL values break the correlation by being too high. Importantly the latter, deviating data points are unlikely due to experimental errors but indicate that the minimal treatment of the sequence separation in Equation 4 is in part unrealistic: it assumes that the residues act independently of their sequence neighbors. The problem is best illustrated by the φ-value of V88A in p81–82 that changes in good accord with those of the neighboring probes in β4, i.e. V85A and V90A. Yet V88A yields a ΔL value that is of opposite sign of its sequence neighbors, making the mutation an outlier in the plot (Fig. 4). The cause of this deviation is simply that some of the long range interactions of V88A extend to α2, hence bridging the gap of the backbone incision. When it comes to control of the folding reaction, however, the contribution of these interactions is small in relation to the more extensive and conflicting contact bias of the other residues in β4. Thus, to introduce some persistence length of the backbone we included a weight from the nearest neighbors in the L calculation according to Equation 6, Lw. The correction leads to a small improvement of the correlation with Δφ: r
Circular Permutants of S6

![Chevron plots for the conservative mutations of S6\textsuperscript{wt}, P\textsuperscript{13–14}, P\textsuperscript{33–34}, P\textsuperscript{54–55}, P\textsuperscript{68–69}, and P\textsuperscript{81–82}, forming the basis for \( \phi \)-value analysis. The mutants are grouped according to their positions in the secondary structure elements following the enumeration of S6\textsuperscript{wt}. The kinetic parameters accompanying the data are given in Tables 1–6.](image)

increases to 0.69 (Fig. 4). The result indicates, not surprisingly, that the action of a residue in the folding transition state is coupled to the action of its sequence neighbors. Even so, the most significant contributor to obstruct the correlation between \( \Delta \phi \) and \( \Delta L_w \) seems to persist. A group of \( \phi \)-values at large negative values of \( \Delta L_w \) appears to be insensitive to circular permutation, giving rise to a trumpet-like shape of the plot. Again this pattern is not due to experimental errors but stems from the mechanistic details of the S6 folding reaction.

**Folding at the Level of Individual Side Chains and Secondary Structure Elements**—Upon closer examination of how the transition-state ensemble responds to circular permutation, it is apparent that the \( \Delta \phi \) values at fixed sequence positions display considerable variation. It is also apparent that the \( \Delta \phi \) values can be grouped according to their positions in the S6 structure. For example, the side chains in \( \beta 4 \) are biased to negative \( \Delta L_w \) and display correspondingly the highest values of \( \Delta \phi \) (Fig. 5). Vice versa, the data points for \( \alpha 1 \) are found at the bottom left-hand side of the correlation (Fig. 5). The favorable span of \( \Delta \phi \) and \( \Delta L_w \) values associated with \( \beta 4 \) and \( \alpha 1 \) result also in relatively high individual correlation coefficients of \( r = 0.64 \) and \( r = 0.80 \), respectively. The effect of circular permutation on \( \Delta L_w \) for mutations in \( \beta 3 \) and \( \alpha 2 \), however, is smaller overall, manifested in data points that are grouped rather symmetrically around the origin. Accordingly the contribution from \( \beta 3 \) and \( \alpha 2 \) to the global correlation between \( \Delta \phi \) and \( \Delta L_w \) is relatively modest. In apparent contrast, the \( \Delta \phi \) values of \( \beta 1 \) seem not to respond to circular permutation at all. Despite the large negative \( \Delta L_w \) values of this secondary structure element, the \( \Delta \phi \) values fall along the x axis, save the odd data point of V6A in the anomalously condensed transition state of P\textsuperscript{13–14} (Fig. 5). Intriguingly it turns out that this deviating behavior of \( \beta 1 \) is a direct prediction from a folding mechanism with two competing folding nuclei that are partly overlapping.

**Folding at the Level of Overlapping Nuclei**—In our previous studies of the S6 permutants P\textsuperscript{13–14} (14), P\textsuperscript{54–55}, and P\textsuperscript{68–69} (19) and the sequence-divergent S6 from *Aquifex aeolicus* (22), it was indicated that the S6 structure is composed of two competing folding nuclei: \( \beta 1 + \alpha 1 + \beta 3 \) and \( \beta 1 + \alpha 2 + \beta 4 \). These two-strand helix motifs match the size of an independently folding unit, a so-called foldon (33, 34), and also represent alternative starting points for the folding reaction (20). Depending on permutation and sequence details, however, the different S6 variants will choose different routes across this folding landscape (see Ref. 35). The diffuse transition-state structure of P\textsuperscript{81–82} suggests that this protein folds by both channels, i.e. P\textsuperscript{81–82} has similar probabilities to fold by either the \( \alpha 1 \) or the \( \alpha 2 \) nuclei, whereas the polarized transition state of P\textsuperscript{13–14} indicates...
Circular Permutants of S6

a strong preference for the α2 channel. Thus, although the overall features of the S6 energy landscape is determined by topology, the folding pathway of a specific S6 variant is ultimately determined by the sequence details and could vary. This modular organization of the protein structure into competing foldons seems not to be unique for S6 but is also discerned in proteins with other types of topologies (35–38). An interesting feature of the S6 nuclei, however, is that they are partly overlapping: by sharing β1 the folding of one side of the protein is coupled to the folding of the other. To facilitate the comparison of the different S6 permutants, we have schematically color-coded the secondary structure elements according to their average ϕ-values (Tables 1–6 and Fig. 6). First it can be seen that the ϕ-value distributions of all S6 variants encompass β1. In addition, all proteins indicate the involvement of either or both of the helices. S6wt and P81–82 show average ϕ-values above 0.15 in both α1 and α2, whereas P13–14, P33–34, P54–55, and P68–69 tend to be polarized to one or the other (Tables 1–6 and Fig. 6). Finally the ϕ-values indicate the contribution of a flanking strand on either side of β1, making it possible to outline the contours of the two-strand helix motif in each of the permutant transition-state structures. Taking into account also the lower ϕ-values it is possible to discern a second, partly overlapping, two-strand helix motif in the diffuse transition-state ensemble of P81–82 (Fig. 6).

Additional Support for Parallel Pathways: the Nuclei Overlap Region Displays Constant ϕ-Values—The critical question is then: what says that the diffuse ϕ-value distributions observed for S6wt and P81–82 represent parallel folding over two competing pathways rather than folding over a “single” pathway with a globally expanded, transition-state structure? In support of the former scenario, all S6 variants have very similar folding rate constants despite radical changes of the appearance of the folding nucleus. The energetic costs of forming the structurally analogous α1 and α2 nuclei are essentially the same. On the other hand, if the folding reaction would be uniform the large changes in the character of the transition state are expected to have larger kinetic impact (17). The folding nuclei for the different S6 variants also have a very similar solvent-accessible surface area as measured from the kinetic m-values (Fig. 1 and Tables 1–6). Such invariable m-values are in perfect accord with a shift between two equivalent foldons but are difficult to reconcile with a single, structurally variable nucleus. Yet another point is that the existence of two competing pathways provides a simple explanation for why the folding kinetics of the S6 variants are relatively insensitive to changes in protein stability and contact order of the folded states (31). In a system with two competing pathways even major entropic or energetic perturbations in one part of the structure could have a relatively modest impact on the global refolding rate constant: the flux of molecules simply is redistributed from the penalized nucleus to the other. Now coming back to the relation between Δϕ and ΔLm, a direct test of the competing pathway model is provided by how the ϕ-values in the foldon-overlap region change with circular permutation. As the interactions anchoring the pivotal β1 will change from one side of the protein to the other, the associated ϕ-values are expected to change less than those in the other secondary structure elements. In the balanced case, the ϕ-values of β1 could even be completely unaffected by pathway changes: half of the contacts are with the α1 nucleus, and half are with the α2 nucleus, yielding constant, fractional ϕ-values below 0.5 regardless of how the protein progresses down the folding funnel. Consistent with this scenario, the experimentally observed ϕ-values of β1 display no appreciable change upon circular permutation despite substantial changes in ΔLm (Fig. 5). The sole exception is the high
Circular Permutants of S6

![Circular Permutants of S6](https://via.placeholder.com/150)

FIGURE 4. Upon circular permutation the individual \( \varphi \) values of S6 undergo systematic changes (\( \Delta \varphi \)) that can be related to differences in sequence separation between pairwise contacts \( \Delta L \) (Equations 4–6). A, plot of \( \Delta \varphi \) versus \( \Delta L \) according to Equations 4 and 5. B, plot of \( \Delta \varphi \) versus \( \Delta L_w \) according to Equation 6. The improvement of the correlation for \( \Delta L_w \) suggests that the action of residues that are close in sequence is to some extent coupled. Even so, the plot of \( \Delta \varphi \) versus \( \Delta L_w \) shows a trumpet-like shape because the \( \varphi \) values in \( \beta_1 \) (the overlap of the two competing S6 nuclei) seem insensitive to permutation. C, upon removal of the \( \varphi \) values in \( \beta_1 \) the correlation between \( \Delta \varphi \) versus \( \Delta L_w \) increases to \( r = 0.82 \). D, plot of \( \Delta \varphi \) versus \( \Delta L_w \) where the individual data points are averaged over the individual secondary structure elements and \( \beta_1 \) is excluded (\( r = 0.89 \)). The improved correlation of this plot indicates that the persistence length of the S6 sequence with respect to folding nucleation approaches the length of a secondary structure element.

The \( \varphi \) value of V6A in the anomalously condensed transition state of \( \beta_1 \) (Fig. 4, B) shows little variation across the different S6 constructs (Tables 1–6). Looking at the absolute levels of \( \varphi \), however, these are substantially skewed by heterogeneities in the contact free energies: the interactions keeping together the entropically favorable contacts as done by Wang and co-workers (25, 39). Note also the strategies used in Refs. 40 and 41.

The Wild-type \( \varphi \) Values Are Partly a Function of Heterogeneous Contact Energies—It is thus apparent that changes of \( \varphi \) upon circular permutations can be simply predicted from sequence separation, but can this rationale also be used to predict the original \( \varphi \) values of S6\(^{wt} \)? The answer is no. The reason is that the relation between \( \Delta \varphi \) and \( \Delta L \) only captures the chain entropy component of the folding process and does not explicitly account for heterogeneities in the contact potential. As S6\(^{wt} \) is the base line, any heterogeneity in the contact free energies is cancelled by subtraction leaving just the effect of the chain entropy differences. In support of this assumption, the stability loss upon mutation in a specific sequence position (\( \Delta \Delta G_{D,N} \))
Circular Permutants of S6

An explanation for the biased contact potential of S6 is that there is a selective pressure to optimize the folding cooperativity to avoid misfolding and aggregation (14). The potential is tuned to make each folding and unfolding event equally (unlikely) (17). In principle, such equalization of the folding probability reduces the preference for any specific trajectory that could otherwise be associated with native-state fraying or a population of structurally promiscuous intermediates. From a macroscopic view, such a broad folding progression is manifested in a globally diffuse transition-state ensemble where each contact is present to a fractional extent (17). Within the context of two competing pathways, optimization of folding cooperativity would mean that both pathways are about equally populated as indicated for S6wt and P\textsuperscript{81–82}. The data presented in this study suggest that the balance between the two pathways is buffered by structural overlap.

**Competing Nuclei: a Simple Reason for Complex Behavior**

In summary, the folding landscape of S6 can be schematically visualized as shown in Fig. 8 where the main trajectories are linked to folding of either of the two competing nuclei: a funnel of funnels (see Refs. 36 and 42). Notably these global features of the S6 landscape are linked to topology but do not imply that all variants or divergent members of the S6 family need to sample both trajectories (22, 35) (see Fig. 3). Rather the landscape sets the boundaries for how the folding reaction is free to vary. Depending on the specific sequence details one protein could favor trajectory 1, and another protein could favor trajectory 2, but upon directed perturbations of the systems the missing trajectory is expected to show up. In addition to the nuclei in Fig. 8 there are likely to be other combinatorial ways of nucleating the S6 structure that have been omitted because they seem relatively rare. For example, the simultaneous formation of both nuclei and an additional path linked to a third nuclei composed of strands \(\beta_1, \beta_3,\) and \(\beta_4\) (see the structure of WW domains (43)). Despite these limitations, the two-component landscape captures at a structural level the most essential characteristics of the S6 folding reaction. It identifies the mutations inducing a Hammond postulate shift between consecutive transition-state ensembles (28) and mutations that shift the folding reaction to parallel transition states (20). Notably shifts between the \(\alpha_1\) and \(\alpha_2\) trajectories upon point mutation are in several cases associated with anti-Hammond behavior (44, 45) indicating that the early \(\alpha_2\) nucleus is slightly more expanded than the \(\alpha_1\) nucleus (see Fig. 8). In parallel studies, Werbeck and Itzhaki (37) and Lowe and Itzhaki (38) have arrived at an analogous folding scheme for modular ankyrin repeat proteins based on pathway perturbations following point mutations. This indicates that deconstruction of protein structures into component nuclei presents a general strategy for dissecting complex folding reactions. Even so, the simplistic landscape in Fig. 8 fails to shed any light on the exaggerated \(\Delta\phi\) response in part of P\textsuperscript{13–14} (see Fig. 5). The cluster of anomalously high \(\phi\)-values in this permutant (i.e. positions 75, 88, and possibly also 6) seems due to transient misfolding and needs an extended description based on non-native contacts.\(^4\) The distinctive detail of the S6 reaction in Fig.

---

4. E. Haglund and M. Oliveberg, unpublished data.
8, however, is the resolution of the nuclei overlap. Identification of such a structural coupling between multiple, elementary nucleation motifs not only has implications for structural evolution but also presents a simple explanation for how folding cooperativity is mechanistically controlled in larger proteins with more complex topology.

REFERENCES
1. Oliveberg, M., and Wolynes, P. G. (2005) Q. Rev. Biophys. 38, 245–288
2. Chiti, F., Calamai, M., Taddei, N., Stefani, M., Ramponi, G., and Dobson, C. M. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, Suppl. 4, 16419–16426
3. Sandelin, E., Nordlund, A., Andersen, P. M., Marklund, S. S., and Oliveberg, M. (2007) J. Biol. Chem. 282, 21230–21236
4. Otzen, D. E., Kristensen, O., and Oliveberg, M. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 9907–9912
5. Richardson, J. S., and Richardson, D. C. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 2754–2759
6. Thirumalai, D., Klimov, D. K., and Dima, R. I. (2003) Curr. Opin. Struct. Biol. 13, 146–159
7. Matysiak, S., and Clementi, C. (2006) J. Mol. Biol. 363, 297–308
8. Stoycheva, A. D., Brooks, C. L., III, and Onuchic, J. N. (2004) J. Mol. Biol. 340, 571–585
9. Silow, M., Tan, Y. J., Fersht, A. R., and Oliveberg, M. (1999) Biochemistry 38, 13006–13012
10. Yang, S., Cho, S. S., Levy, Y., Cheung, M. S., Levine, H., Wolynes, P. G., and Onuchic, J. N. (2004) Proc. Natl. Acad. Sci. U. S. A. 101, 13786–13791
11. Guo, Z., and Eisenberg, D. (2006) Proc. Natl. Acad. Sci. U. S. A. 103, 8042–8047
12. Bennett, M. I., Choe, S., and Eisenberg, D. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 3127–3131
13. Fandrich, M., Fletcher, M. A., and Dobson, C. M. (2001) Nature 410, 165–166
14. Lindberg, M., Tangrot, J., and Oliveberg, M. (2002) Nat. Struct. Biol. 9, 818–822
15. Durnouille, M., Canet, D., Last, A. M., Pardon, E., Archer, D. B., Muyldermons, S., Wüns, L., Matagne, A., Robinson, C. V., Redfield, C., and Dobson, C. M. (2005) J. Mol. Biol. 346, 773–788
16. Nordlund, A., and Oliveberg, M. (2006) Proc. Natl. Acad. Sci. U. S. A. 103, 10218–10223
17. Plotkin, S. S., and Onuchic, J. N. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 6509–6514
18. Jackson, S. E. (1998) Fold. Des. 3, R81–R91
19. Lindberg, M. O., Haglund, E., Hubner, I. A., Shakhnovich, E. I., and Oliveberg, M. (2006) Proc. Natl. Acad. Sci. U. S. A. 103, 4083–4088
20. Lindberg, M. O., and Oliveberg, M. (2007) Curr. Opin. Struct. Biol. 17, 21–29
21. Zarrine-Afsar, A., Larson, S. M., and Davidson, A. R. (2005) Curr. Opin. Struct. Biol. 15, 42–49
22. Olofsson, M., Hansson, S., Hedberg, L., Logan, D. T., and Oliveberg, M. (2007) J. Mol. Biol. 365, 237–248
23. Lappalainen, I., Hurley, M. G., and Clarke, J. (2008) J. Mol. Biol. 375, 547–559
24. Matysiak, S., and Clementi, C. (2004) J. Mol. Biol. 343, 235–248
25. Wu, L., Zhang, J., Wang, J., Li, W. F., and Wang, W. (2007) Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 75, 031914
26. Lindahl, M., Svensson, L. A., Liljas, A., Sedelnikova, S. E., Eliseikina, I. A., Fomenkova, N. P., Nevskaya, N., Nikonorov, S. V., Garber, M. B., Munanova, T. A., Rykonova, A. I., and Amons, R. (1994) EMBO J. 13, 1249–1254
27. Lindberg, M. O., Tangrot, J., Otzen, D. E., Dolkah, D. A., Finkelnest, A. V., and Oliveberg, M. (2001) J. Mol. Biol. 314, 891–900
28. Otzen, D. E., Kristensen, O., Proctor, M., and Oliveberg, M. (1999) Biochemistry 38, 6499–6511
29. Oliveberg, M. (1998) Acc. Chem. Res. 31, 765–772
30. Fersht, A. R. (1999) Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding. W. H. Freeman and Co., New York
31. Miller, E. J., Fischer, K. F., and Marqusee, S. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 10359–10363
32. Plaxco, K. W., Simons, K. T., and Baker, D. (1998) J. Mol. Biol. 277, 985–994
33. Panchenko, A. R., Luthey-Schulten, Z., and Wolynes, P. G. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 2008–2013
34. Panchenko, A. R., Luthey-Schulten, Z., Cole, R., and Wolynes, P. G. (1997) J. Mol. Biol. 272, 95–105
35. Chavez, L. L., Gosavi, S., Jennings, P. A., and Onuchic, J. N. (2006) Proc. Natl. Acad. Sci. U. S. A. 103, 10254–10258
36. Ferreiro, D. U., Cho, S. S., Komives, E. A., and Wolynes, P. G. (2005) J. Mol. Biol. 354, 679–692
37. Werbeck, N. D., and Itzhaki, L. S. (2007) Proc. Natl. Acad. Sci. U. S. A. 104, 7863–7868
38. Lowe, A. R., and Itzhaki, L. S. (2007) Proc. Natl. Acad. Sci. U. S. A. 104, 2679–2684
39. Chen, J., Wang, I., and Wang, W. (2004) Proteins 57, 153–171
40. Weikl, T. R., and Dill, K. A. (2003) J. Mol. Biol. 332, 953–963
41. Weikl, T. R. (2008) Arch. Biochem. Biophys. 469, 67–75
42. Ferreiro, D. U., and Komives, E. A. (2007) Proc. Natl. Acad. Sci. U. S. A. 104, 7735–7736
43. Ferguson, N., Johnson, C. M., Macias, M., Oschkinat, H., and Fersht, A. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 13002–13007
44. Matthews, J. M., and Fersht, A. R. (1995) Biochemistry 34, 6805–6814
45. Otzen, D. E., and Oliveberg, M. (2002) J. Mol. Biol. 317, 613–627