Trimming Down a Protein Structure to Its Bare Foldons

SPATIAL ORGANIZATION OF THE COOPERATIVE UNIT

Received for publication, October 11, 2011 Published, JBC Papers in Press, November 23, 2011, DOI 10.1074/jbc.M111.312447

Ellinor Haglund, Jens Danielsson, Saraboji Kadhirvel, Magnus O. Lindberg, Derek T. Logan, and Mikael Oliveberg

From the Department of Biochemistry and Biophysics, Arrhenius Laboratories of Natural Sciences, Stockholm University, S-106 91 Stockholm and the Department of Biochemistry and Structural Biology, Lund University, S-221 00 Lund, Sweden

Background: Structural cooperativity safeguards native proteins but is yet poorly understood.

Results: Guided by the folding nucleus, we reduced the size of a protein without compromising its structural integrity.

Conclusion: Folding nuclei are used in a modular fashion to extend or reduce the cooperative units of proteins.

Significance: Understanding cooperativity is crucial for understanding protein function and for rational design of structural properties.

Folding of the ribosomal protein S6 is a malleable process controlled by two competing, and partly overlapping, folding nuclei. Together, these nuclei extend over most of the S6 structure, except the edge strand β2, which is consistently missing in the folding transition states; despite being part of the S6 four-stranded sheet, β2 seems not to be part of the cooperative unit of the protein. The question is then whether β2 can be removed from the S6 structure without compromising folding cooperativity or native state integrity. To investigate this, we constructed a truncated variant of S6 lacking β2, reducing the size of the protein from 96 to 76 residues (S6Δβ2). The new S6 variant expresses well in Escherichia coli and has a well dispersed heteronuclear single quantum correlation spectrum and a perfectly wild-type-like crystal structure, but with a smaller threestranded β-sheet. Moreover, S6Δβ2 displays an archetypical v-shaped chevron plot with decreased slope of the unfolding limb, as expected from a protein with maintained folding cooperativity and reduced size. The results support the notion that foldons, as defined by the structural distribution of the folding nuclei, represent a property-based level of hierarchy in the build-up of larger protein structures and suggest that the role of β2 in S6 is mainly in intermolecular binding, consistent with the position of this strand in the ribosomal assembly.

The biological function of proteins is not only a matter of shape and stability, but also relies on the energy landscape controlling the structural dynamics and occupancy of alternatively structured states (1). As a painful reminder of the latter, protein misfolding diseases (2, 3) arise recurrently from mutations that would appear perfectly benign if judged by native structure and activity alone (3, 4). Of particular interest here are the role and molecular origins of folding cooperativity, which safeguard the integrity of the native structure by reducing the occupancy of partly unfolded or misfolded states (5, 6). A clue to how folding cooperativity is maintained and propagated over large structural distances was uncovered by the folding behavior of the ribosomal protein S6 (7, 8). The structure of S6 seems to be composed of two cooperative subunits or “foldons,” σ1 and σ2, that also act as competing nuclei in the folding process (8) (Fig. 1). S6 can thus fold along either of two parallel, and sequentially opposed, pathways, one starting with the nucleation of σ1 and the other starting with the nucleation of σ2 (9) (Fig. 1). Even so, the nucleation events of σ1 and σ2 are not entirely independent, but are coupled by a structural overlap in the form of the shared strand β1 (10). Such overlap between competing nuclei stands out as an efficient way of linking small cooperative subunits into larger structures without compromising global cooperativity or folding kinetics (10). Folding of large proteins would, in principle, simplify folding of smaller proteins that are structurally overlapping to parallel. However, from studies of S6 where the bias between the σ1 and σ2 pathways has systematically been shifted by circular permutation, it is apparent that the edge strand β2 never participates in the nucleation process (6, 8, 10). Despite being an integral part of the native sheet, β2 appears to be outside the cooperative unit of S6 (Fig. 1). It is then expected that S6 should also be able to fold cooperatively without β2, providing that the σ1 and σ2 foldons are sufficiently stable on their own. In this study, we have tested this possibility by simply removing β2 from the S6 structure; the strand was shuffled to the end of the sequence through circular permutation and then cut off (see Fig. 2). The truncated protein (S6Δβ2) displays a highly dispersed solution NMR spectrum characteristic of rigid tertiary structures and a crystal structure that closely matches that of the wild-type protein. Moreover, S6Δβ2 retains the cooperative folding behavior of wild-type S6 and the circular permuted from which it is derived. Taken together, this shows that β2 is not required for folding or native state integrity, but rather plays another role in the S6 structure, possibly in functional shape adjustment. This division of globular domains into cooperative core and subsidiary secondary structure is not only useful for guiding protein engineering, but
also sheds light on dynamic behavior and structural constraints in protein evolution.

**EXPERIMENTAL PROCEDURES**

**Protein Expression and Purification**—Mutations were performed with the QuikChange Multi site-directed mutagenesis kit (Stratagene, Agilent Technologies, Santa Clara, CA). Oligonucleotides were purchased from Eurofins MWG Operon (Ebersberg, Germany), and all mutations were confirmed by sequencing (Eurofins MWG Operon). The protein was overexpressed in *Escherichia coli* strain BL21 and purified as described in Ref. 11.

**NMR Spectroscopy**—All NMR data were obtained at 25 °C, pH 6.3, and with a protein concentration of 0.5 mM, on a Bruker 700-MHz spectrometer (Bruker Avance, Karlsruhe, Germany) equipped with a cryogenically cooled triple resonance probe. Assignment was by standard 15N-{1H} HSQC, 2 HNCA, HN(CO)CA, HN(CA)CO, HNCO, 15N-edited NOESY and total correlation spectroscopy experiments. Dynamics were measured by standard T1, T2, and steady-state heteronuclear NOE experiments. Spectra were transformed using NMRPipe (12) and analyzed with the program Sparky (45). In the T1 and T2 experiments, the signal attenuation, from 10 different relaxation delays, was fitted to a single exponential decay and the relaxation rates were determined. The fitting routine was performed using MATLAB (MathWorks, Natick, MA).

**X-ray Crystallography**—Crystals of S6ΔH2 were grown using the hanging drop vapor diffusion method at 20 °C from drops containing equal volumes of protein (18 mg/ml) and a reservoir solution composed of 1.9 M ammonium sulfate, 0.1 M MES, pH 6.5, and 8% (v/v) 1,4-dioxane. The addition of dioxane was necessary to avoid the formation of crystal clusters due to excessive nucleation. Crystals appeared after 4 days and reached a maximum size of 0.2 × 0.2 × 0.1 mm in 10 days. The crystals were cryoprotected using 20% glycerol in the mother liquor. Crystallographic data were collected at 100 K at station 1911-2 of MAX-lab, Lund, Sweden, using a 165mm charge coupled device (CCD) detector from MarResearch (Norderstedt, Germany). Images were indexed, integrated, and scaled using XDS (13). Crystals diffracted to 0.96 Å resolution, and data collection statistics are shown in Table 1.

<table>
<thead>
<tr>
<th>Data statistics</th>
<th>X-ray wavelength (Å)</th>
<th>1.038</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>20.0-0.96 (0.99-0.96)</td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P21 21 21</td>
<td></td>
</tr>
<tr>
<td>Unit cell parameters (Å)</td>
<td>α = 22.2, β = 40.8, γ = 68.66</td>
<td></td>
</tr>
<tr>
<td>Unit cell volume (Å³)</td>
<td>62,189</td>
<td></td>
</tr>
<tr>
<td>Solvent content (%)</td>
<td>29.8</td>
<td></td>
</tr>
<tr>
<td>Mosaicity (%)</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>No. of measured reflections</td>
<td>230,289</td>
<td></td>
</tr>
<tr>
<td>No. of independent reflections</td>
<td>38,300</td>
<td></td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.3 (98.2)</td>
<td></td>
</tr>
<tr>
<td>Redundancy (%)</td>
<td>6.0 (4.9)</td>
<td></td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>13.7</td>
<td></td>
</tr>
<tr>
<td>Rfree (%)</td>
<td>11.3</td>
<td></td>
</tr>
<tr>
<td>No. of water molecules</td>
<td>129</td>
<td></td>
</tr>
<tr>
<td>r.m.s. deviation from ideal values</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td>Average distances (Å)</td>
<td>0.033</td>
<td></td>
</tr>
<tr>
<td>Average B-factor, protein (water) (Å²)</td>
<td>11.6 (28.2)</td>
<td></td>
</tr>
<tr>
<td>Average anisotropy for all protein atoms</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>Ramachandran plot statistics*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most favored regions (%)</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

*Calculated using Molprobity.

**TABLE 1**

Crystallographic data-collection and refinement statistics for S6ΔH2 Values in parentheses correspond to the highest resolution shell.

The structure of S6ΔH2 was determined by the molecular replacement method using the program Phaser (14), with the structure of ribosomal protein S6 from *Thermus thermophilus* (Protein Data Bank (PDB) code: 1RIS) (15) as a starting model. The model was built using the automatic building procedure in ARP/wARP (16), and the model was completed by manual building using the molecular graphics program Coot (17). In the initial stages, all refinement was carried out to a resolution of 1.2 Å using phenix.refine (18), with 5% of the total reflections randomly set aside for cross-validation. Subsequently, the resolution was gradually extended to the full resolution range, and refinement was carried out using SHELXL-97 (19) in conjugate-gradient least-squares mode. Disordered side chains were modeled, and their occupancies were refined using the restraint that the occupancies should sum to unity. Hydrogen atoms were not refined, but were included in *riding* positions. Anisotropic refinement was carried out in a series of rounds.
with SHELXL. After each of these refinement rounds, the model was inspected manually with Fourier maps generated from SHELX files using Coot. The refinement statistics are summarized in Table 1. The stereochemistry of the model was evaluated using MolProbity (20). Structural alignments and illustrations were made using the SSM algorithm (21) and PyMOL (22), respectively.

**RESULTS**

**Design of S6\(\Delta\beta_2\)—**The new construct S6\(\Delta\beta_2\) was derived from the circular permutant P54–55, which has a linker between the wild-type N and C termini and an incision between Lys-54 and Asp-55 to localize \(\beta_2\) at the C terminus (permutant P54–55). Second, \(\beta_2\) was truncated from the C terminus of P54–55 to obtain S6\(\Delta\beta_2\). The latter step also included the insertion of gatekeepers in \(\beta_3\) (F7T and L8T) and in \(\alpha_1\) (A62Q and L63N) to better adjust the newly exposed protein surface to solvent and avoid unwanted aggregation. The positions of these gatekeepers in the tentative S6\(\Delta\beta_2\) structure are shown in red.

**Stopped-flow Kinetics—**Stopped-flow measurements and curve fitting were performed as in Ref. 10. All kinetic parameters were derived from the linear regime near the bottom of the chevron plots, i.e. at \(\pm 2\) m from the transition midpoint, to avoid contributions from the curvatures in the unfolding and refolding limbs at high and low [GdmCl]. The data were fitted according to the standard equation

\[
\log k_{\text{obs}} = \log(k_f + k_u)
\]

\[
= \log(10^\log k_f + m_f[GdmCl]) + 10^\log k_u + m_u[GdmCl])
\]  
(Eq. 1)

where \(k_f\) and \(k_u\) are the extrapolated values of the refolding \((k_f)\) and unfolding \((k_u)\) rate constants at 0 M GdmCl, and \(m_f\) and \(m_u\) are the slopes of the refolding and unfolding limbs, respectively.

**RESULTS**

**Design of S6\(\Delta\beta_2\)—**The new construct S6\(\Delta\beta_2\) was derived from the circular permutant P54–55, which has a linker between the wild-type N and C termini and an incision between Lys-54 and Asp-55 to localize \(\beta_2\) at the C terminus (permutant P54–55). Second, \(\beta_2\) was truncated from the C terminus of P54–55 to obtain S6\(\Delta\beta_2\). The latter step also included the insertion of gatekeepers in \(\beta_3\) (F7T and L8T) and in \(\alpha_1\) (A62Q and L63N) to better adjust the newly exposed protein surface to solvent and avoid unwanted aggregation. The positions of these gatekeepers in the tentative S6\(\Delta\beta_2\) structure are shown in red.

**Stopped-flow Kinetics—**Stopped-flow measurements and curve fitting were performed as in Ref. 10. All kinetic parameters were derived from the linear regime near the bottom of the chevron plots, i.e. at \(\pm 2\) m from the transition midpoint, to avoid contributions from the curvatures in the unfolding and refolding limbs at high and low [GdmCl]. The data were fitted according to the standard equation

\[
\log k_{\text{obs}} = \log(k_f + k_u)
\]

\[
= \log(10^\log k_f + m_f[GdmCl]) + 10^\log k_u + m_u[GdmCl])
\]  
(Eq. 1)

where \(k_f\) and \(k_u\) are the extrapolated values of the refolding \((k_f)\) and unfolding \((k_u)\) rate constants at 0 M GdmCl, and \(m_f\) and \(m_u\) are the slopes of the refolding and unfolding limbs, respectively.

**RESULTS**

**Design of S6\(\Delta\beta_2\)—**The new construct S6\(\Delta\beta_2\) was derived from the circular permutant P54–55, which has a linker between the wild-type N and C termini and an incision between Lys-54 and Asp-55 to localize \(\beta_2\) at the C terminus (permutant P54–55). Second, \(\beta_2\) was truncated from the C terminus of P54–55 to obtain S6\(\Delta\beta_2\). The latter step also included the insertion of gatekeepers in \(\beta_3\) (F7T and L8T) and in \(\alpha_1\) (A62Q and L63N) to better adjust the newly exposed protein surface to solvent and avoid unwanted aggregation. The positions of these gatekeepers in the tentative S6\(\Delta\beta_2\) structure are shown in red.

**Stopped-flow Kinetics—**Stopped-flow measurements and curve fitting were performed as in Ref. 10. All kinetic parameters were derived from the linear regime near the bottom of the chevron plots, i.e. at \(\pm 2\) m from the transition midpoint, to avoid contributions from the curvatures in the unfolding and refolding limbs at high and low [GdmCl]. The data were fitted according to the standard equation

\[
\log k_{\text{obs}} = \log(k_f + k_u)
\]

\[
= \log(10^\log k_f + m_f[GdmCl]) + 10^\log k_u + m_u[GdmCl])
\]  
(Eq. 1)

where \(k_f\) and \(k_u\) are the extrapolated values of the refolding \((k_f)\) and unfolding \((k_u)\) rate constants at 0 M GdmCl, and \(m_f\) and \(m_u\) are the slopes of the refolding and unfolding limbs, respectively.
backbone chemical shifts. The results indicate three β-strands (residues 6–12, 33–39, and 45–51) and two α-helices (residues 15–29 and 56–76), the positions of which overlap precisely with those of the parent protein S6wt (supplemental Fig. S1). No peak splitting was found in sequential vicinity to the prolines, which indicates that S6Δβ2 resides in a single isomeric state. In contrast, P54–55 exhibits proline cis-trans isomerization near the C terminus, which is truncated in S6Δβ2, and around Pro-3 near the N terminus, which is homogeneous in S6Δβ2 (25). The dynamic motions of the S6Δβ2 structure were mapped out by NMR relaxation experiments. At all positions, except for the outermost N-terminal residues, the relaxation data indicate a well ordered protein with a low degree of dynamic motion. The heteronuclear 15N-1H NOE values are high and uniform along the S6Δβ2 backbone, with an average value of 0.81 ± 0.17. Although no significant dips are seen along the sequence, residues 20–25 at the center of helix 1 stand out by displaying higher NOE values (Fig. 3). This indicates that the central part of helix 1 is particularly rigid. The rotational correlation time (τc) of the S6Δβ2 molecule was determined according to the approximation (26)

$$\tau_c = \frac{1}{2\omega_n} \sqrt{\frac{6R_2}{R_1} - 7}$$ (Eq. 2)

where ωn is the Larmor frequency of nitrogen, and R1 and R2 are the longitudinal and transversal relaxation rates, respectively. This approximation yields good precision when the R1 and R2 values are homogeneous along the sequence, i.e., when conformational exchange and local dynamics are small overall. The average values of R1 and R2 were 1.8 ± 0.15 and 8.3 ± 1.4 s⁻¹, respectively (Fig. 3). By omitting some extreme ratios of R2/R1, τc was estimated to 5.1 ± 0.4 ns (Equation 2), which corresponds to a hydrodynamic radius of 17.8 ± 0.4 Å using Stokes’ law. The hydrodynamic radius expected for a globular protein of the size of S6Δβ2 is 17.4 Å. This close agreement shows that S6Δβ2 does not associate into dimers or higher-order species but remains monomeric at the high protein concentration used for the NMR experiments. The aggregation gatekeepers (23) designed into the edge-exposed β3 in S6Δβ2 seem thus to work according to prediction.

X-ray Structure—S6Δβ2 crystallizes in space group P212121 with one molecule in the asymmetric unit. The exceptionally high resolution of the diffraction data (0.96 Å) and low mosaicity of the crystals are consistent with a compact molecule with little flexibility (Fig. 4). The structure consists of a three-stranded β-sheet with two α-helices on the same side. The radius of gyration including hydration shell, calculated using CRY SOL (27), is 16.6 Å, in good agreement with the NMR-derived value of 17.4 Å. The S6Δβ2 structure is essentially identical to the in silico theoretical model for P54–55 except for the deletion of strand β2 and the following loop. It is also very similar to that of S6wt. Because this is the first reported crystal structure of a circular permutant of S6, this is an important confirmation that the permutation has caused little perturbation of the overall structure. No extended β-sheets are formed between molecules in the crystal. This is in contrast to another artificial variant, S6ΔA12, in which hydrophobic residues were engineered into β2, leading to a β-zipped tetramer in the crystal (23). The lack of oligomerization of S6Δβ2 follows from the introduction of edge-strand gatekeepers in β3 and is consistent with NMR data.

Structural perturbations caused by the removal of β2 are minimal and are mostly limited to rearrangements of side chains in β3 (Fig. 4). The latter strand relaxes to new positions due to the removal of packing constraints from side chains in β2. For example, Trp-9(62) flips back toward the central strand β1, facilitated by a rearrangement of Asn-49(7) (Fig. 4). Removal of β2 also causes a small rigid body shift of helix α1 (residues 58–76) relative to the rest of the structure. Shifts in Cα positions in this helix between S6wt and S6Δβ2 range from 1.2 to 2.6 Å when the rest of the structure is superimposed. This indicates that β2 has a role in the exact positioning of α1 in the wild-type.
protein. An important interaction in this positioning could be the salt bridge between Lys-65(23) and wild-type position 42 in β2 (supplemental Fig. S2). In contrast, α2 (residues 15–28) is not shifted upon β2 removal, except for some rearrangement at its N terminus. This rearrangement seems due to interactions with the N- and C-terminal linker (residues 41–44) introduced in the circular permutation.

**Folding Kinetics**—S6wt displays a v-shaped chevron plot with approximately linear GdmCl dependence of the logarithmized refolding and unfolding rate constants (logk_u and logk_f) (Fig. 5), characteristic for cooperative, two-state folding behavior according to Scheme 1

\[
D \leftrightarrow \ddagger \leftrightarrow N \text{ \ (SCHEME 1)}
\]

where D, ‡, and N are the denatured ensemble, the transition state and the native state, respectively. The protein stability for such a system is defined by

\[
\Delta G_{D-N} = -2.3RT \log K_{D-N} = -2.3RT(\log k_u - \log k_f) \quad (\text{Eq. 3})
\]

where \(K_{D-N} = [D]/[N] = k_u/k_f\) is the first-order equilibrium constant for folding, and

\[
\log K_{D-N} = \log k_u^{H2O} + m_{D-N}[\text{GdmCl}]
= \log k_u^{H2O} + m_u[\text{GdmCl}] - \log k_f^{H2O} - m_f[\text{GdmCl}] \quad (\text{Eq. 4})
\]

where \(m_{D-N} = m_u - m_f\) are the m-values. Because \(m_f\) and \(m_u\) are measures of changes in the solvent-exposed surface area for the transitions D → ‡ and N → ‡ (Scheme 1), they are commonly used as experimental reaction coordinates for folding (28–30). For S6wt, the transition midpoint is at \([\text{GdmCl}] = 3.4\) M, as measured by the intersect between \(\log k_u\) and \(\log k_f\). At \([\text{GdmCl}] = 0\) M, the stability of S6wt is estimated to \(\Delta G_{D-N}^{0} = 8.5\) kcal/mol (Equation 3, Table 2). For the circular permutant P54–55, the transition midpoint and protein stability increase to \([\text{GdmCl}] = 4.1\) M and \(\Delta G_{D-N}^{0} = 9.1\) kcal/mol, respectively (Table 2). The major part of this stabilization is likely due to the linkage of the N and C termini as mutations/incisions in the dynamic loop between β2 and β3 have generally small effects on protein stability (10). Coupled to the stability gain, P54–55 displays an increased refolding rate constant, which has previously been suggested to arise from a stabilization of the α2 pathway (10). Finally, the removal of β2 from P54–55 to obtain S6Aβ2 results in a pronounced decrease of the transition midpoint; the stability decreases by 5.1 kcal/mol to \(\Delta G_{D-N}^{0} = 4.0\) kcal/mol (Table 2). Most of this stability loss is due to an increased unfolding rate constant (Fig. 5), yielding an overall \(\phi\)-value for β2 removal of approximately zero (Equation 5)

\[
\phi = \Delta \log k_f / (\Delta \log k_u + \Delta \log k_f) \quad (\text{Eq. 5})
\]

where \(\Delta \log k_u\) and \(\Delta \log k_f\) are the change in rate constants upon β2 removal (31). The result shows that β2 has only marginal impact on the folding transition state ensemble and selectively stabilizes the folded state (10). This supports the conclusion that β2 is outside the cooperative core of the S6 structure. Moreover, S6Aβ2 displays a pronounced decrease of \(m_u\) (Eq. A), indicating that its folded state buries less surface area than the parent protein P54–55 (Fig. 5 and Table 2). Such a decrease in buried surface area is expected as the truncated S6Aβ2 is effectively a smaller protein than P54–55 and wild-type S6.

**DISCUSSION**

Foldons as Property-based Level in Structural Hierarchy—Comparison of structurally divergent protein domains indicates that structural evolution to some extent occurs progressively by insertions, deletions, and substitutions of smaller structural units (32). Mutation or insertion of additional residues in a flexible part of the protein can cause it to condense into an ordered secondary structure element that adds to the mother domain. When the mother domain is a well optimized two-state protein (1), the structural integrity of this newly acquired secondary structure element will not be critical for global stability or folding; it can unfold locally or be replaced with a suitable disordered sequence without rupturing the scaffold to which it docks. In sharp contrast, the corresponding alteration of a secondary structure element within the cooperative unit of the mother protein is expected to cause global

![FIGURE 5. Chevron plots of S6wt (black), P54–55 (blue), and β2-truncated variant S6Aβ2 (red), showing logkobs = log(k_u + k_f) versus [GdmCl]. Rate constants are in units of s⁻¹. The overall wild-type-like chevron plot of S6 displays a pronounced decrease of the transition midpoint; the stability decreases by 5.1 kcal/mol to \(\Delta G_{D-N}^{0} = 4.0\) kcal/mol (Table 2). Most of this stability loss is due to an increased unfolding rate constant (Fig. 5), yielding an overall \(\phi\)-value for β2 removal of approximately zero (Equation 5)](image)

**TABLE 2**

<table>
<thead>
<tr>
<th></th>
<th>logk_f[H2O]</th>
<th>m_u</th>
<th>logk_u[H2O]</th>
<th>m_f</th>
<th>MP^a</th>
<th>m_D-N</th>
<th>logK_{D-N}</th>
<th>ΔG_{D-N}^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>S6wt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P54–55</td>
<td>2.64 ± 0.02</td>
<td>1.26 ± 0.03</td>
<td>-3.63 ± 0.14</td>
<td></td>
<td>0.57 ± 0.03</td>
<td>3.42 ± 0.04</td>
<td>1.83 ± 0.06</td>
<td>6.57 ± 0.02</td>
</tr>
<tr>
<td>S6Aβ2</td>
<td>2.30 ± 0.06</td>
<td>1.09 ± 0.01</td>
<td>-3.40 ± 0.10</td>
<td></td>
<td>0.69 ± 0.02</td>
<td>4.09 ± 0.03</td>
<td>1.64 ± 0.03</td>
<td>6.70 ± 0.16</td>
</tr>
</tbody>
</table>

^a Derived from chevron data according to Eq. 1.

^b Transition midpoint derived from the intersect between logk_u and logk_f.

^c m_{D-N} = m_u - m_f according to Eq. 4.

^d logK_{D-N} = logk_{D-N}^{H2O} + m_{D-N}[\text{GdmCl}] = logk_u + m_u[\text{GdmCl}] - logk_f - m_f[\text{GdmCl}] according to Eq. 4.

^e ΔG_{D-N} = -2.3RTlogK_{D-N} = -2.3RT(logk_u - logk_f) according to Eq. 3.
unfolding because the stabilities of two-state domains are generally very sensitive to core substitutions (1,31). To distinguish between such property-based differences within protein structures, we have employed an additional level of structural hierarchy based on foldons (7, 8, 10, 33) (Fig. 1): a minimal cooperative unit that, in principle, would be able to fold cooperatively on its own (cf. Ref. 34). This does not always mean that the foldon would be thermodynamically stable on its own, or even contiguous in sequence, but its folding transition, albeit to a high-energy state, would proceed over a free-energy barrier (1). The structure of S6 seems to be based on two partly overlapping foldons, α1 and α2, identified from the spatial boundaries of two competing folding nuclei (10) (Fig. 1). A notable feature of this arrangement is that β2, regardless of how the protein is rewired through circular permutation, falls outside the boundaries of the two folding nuclei (10); despite β2 being part of the S6 sheet, it seems not to be part of the cooperative core. The ability to remove β2 without compromising folding or stability lends further support to the notion that the properties of the various parts of a protein can be predicted from the spatial distribution of foldons. In the simplest case, the foldons are then expected to outline the inflexible parts of the protein machinery. However, there are also instances where the actual unfolding/folding transitions of foldons are used to control functional rearrangements. For example, reversible unfolding/folding of foldons and domains can lend elasticity to arrayed repeat proteins (35), and the “cracking” of strategically placed foldons can relieve local stress and thereby lower the barriers for specific global motions (36,37).

Foldon Expansion and Reduction—The solution structures and backbone dynamics of S6\textsuperscript{wt} and P\textsuperscript{54−55} have earlier shown that β2 displays dynamic motions in the region connecting to the seemingly flexible β2-β3 loop (25). Together with the autonomous stability of S6\textsuperscript{Alz} (Fig. 5 and Table 2), this indicates that β2 is free to unfold locally in the folded ground state of the wild-type protein. Even so, it is reasonable to believe that S6 could evolve to integrate β2 with the cooperative core, e.g. to better resist aggregation (2) or proteolytic digestion (38,39) or to allow further extension of the mother domain. Evidence that such changes of the foldon architecture indeed occur is provided by the S6-like protein muscle acylphosphatase (PDB code 1APS (40)), which comprises a C-terminal extension that folds up against β2 to form a five-stranded sheet (Fig. 6). Kinetic analysis suggests that the fifth strand is also integrated in the folding nucleus of acylphosphatase together with β2 and α1 (41) (Fig. 6). It is not yet established, however, whether this expansion of the cooperative unit leads to the introduction of a third foldon, and thus, three competing pathways in the folding energy landscape, cf. Ref. 8. Conversely, it should be possible to reduce further the S6 structure to a single foldon comprising just two strands and a helix. This minimal folding unit would then correspond to the smallest observed proteins that are able to fold into globular structures without co-factors (8). Indications that such a reduction could indeed occur are provided by ground state rupture of β4 in the pre-equilibrium of the unfolding reaction at high [GdmCl] (42). As the remaining structure still undergoes cooperative unfolding, it is expected to persist as an independent cooperative unit even in the absence of GdmCl. Consistently, the structures of α2 and β4 show relatively low

---

\textsuperscript{3} As an extension of the original definition by Panchenko et al. (34), the foldon is here not restricted to be contiguous in sequence, but solely determined by three-dimensional packing. Studies of circular permutations show that the sequence connectivity of the foldon structure has little effect on its folding behavior (10).
protection factors as measured by hydrogen-exchange NMR (9); in contrast to the α1 foldon (Fig. 1), which exchanges by global unfolding (EX1), α2 and β4 exchange by local motions in the native free-energy basin (EX2), suggesting that these secondary structure elements can be removed with maintained folding barrier. Although S6 can be expressed and purified with extensive side-chain truncations in the α2 and β4 region, we have not yet been able to isolate the α1 foldon as an independent domain. The isolated α1 requires further optimization of core packing and solvent interface to populate under equilibrium conditions. An example of such an optimization is provided by the sequence-divergent domain of flavoprotein deodecin from Halobacterium salinarum (PDB 2V18) (43), which displays a topology very similar to S6 as it would appear without α2 and β4 (Fig. 6).

Possible Roles of β2—The question then arises: if not required for folding (Fig. 5) or structural integrity (Figs. 3 and 4), what is the role of β2? One possibility is that β2 originates from a loop between α1 and β3, which serves to co-localize the N and C termini in the superfamily of ferredoxin-like folds. An advantage of such co-localization, which is common among globular proteins, could be that it facilitates structural evolution by gene shuffling; the fold can be freely evolved, inserted, or deleted from loops of other domains. A more straightforward explanation is, however, that β2 is maintained for functional reasons. In the S6 structure, β2 provides an anchor for the loop that wraps over S18 in the ribosomal assembly (Fig. 6). Furthermore, comparison with other ferredoxin-like structures shows that β2 is the preferred interface for dimerization (see e.g. ferredoxin-like folds at the SCOP: Structural Classification of Proteins database). A representative example is monooxygenase from Streptomyces coelicolor (PDB 1L0Q and several others) (44) where β2 joins two monomers by forming an anti-parallel sheet (Fig. 6). Interestingly, S6 can be made to assemble in an analogous fashion by truncation of charged gatekeeper side chains in β2 (23) (Fig. 6). In the wild-type protein, these side chains are employed as aggregation gatekeepers that prevent promiscuous edge-to-edge interactions, and upon their removal, S6 assembles into ordered homotetramers (23) (Fig. 6). It is conceivable that the position of β2 outside the cooperative unit of the protein allows such quaternary assemblies to evolve relatively freely, with only minor penalties for the structural properties of the constituent monomers. A basic, two-component, recipe for designing proteins would then be to assemble in an overlapping fashion suitable foldons to shape the cooperative scaffold and then decorate this unit with mutationally tolerant loop and border segments for function and binding.

REFERENCES


Structure of the Cooperative Unit of S6


