Complete change of the protein folding transition state upon circular permutation

Magnus Lindberg[1], Jeanette Tångrot[1,2] and Mikael Oliveberg[1]

[1]*Department of Biochemistry and* [2]*UCMP/Department of Computing Science, Umeå University, S-901 87 Umeå, Sweden.*

Published online 7 October 2002; doi:10.1038/nsb847

Reversing the loop lengths of the small protein S6 by circular permutation has a dramatic effect on the transition state structure: it changes from globally diffuse to locally condensed. The phenomenon arises from a biased dispersion of the contact energies. Stability data derived from point mutations throughout the S6 structure show that interactions between residues that are far apart in sequence are stronger than those that are close. This entropy compensation drives all parts of the protein to fold simultaneously and produces the diffuse transition-state structure typical for two-state proteins. In the circular permutant, where strong contacts and short sequence separations are engineered to concur, the transition state becomes atypically condensed and polarized. Taken together with earlier findings that S6 may also fold by a 'collapsed' trajectory with an intermediate, the results suggest that this protein may fold by a multiplicity of mechanisms. The observations indicate that the diffuse transition state of S6 is not required for folding but could be an evolutionary development to optimize cooperativity.

Imagining that protein stability needs to be kept within some limits for biological function is easy. However, is there any corresponding pressure on protein folding and, hence, the dynamic fluctuations of the native state? Naturally evolved two-state proteins show little variation in folding behavior: the archetypal transition-state structure is overall diffuse with a global distribution of fractional native contacts[1,2]. Severely polarized transition states are rare. According to ϕ-value analysis[2], the fractional contacts are spread evenly throughout the transition state structure, often radiating from a nucleation core of long-range interactions[3–5]. The formation of secondary and tertiary structure takes place more or less concomitantly[5] and seems to involve a majority of the native contacts. For proteins where the transition state shows a clear degree of polarization[6,7], the majority of contacts are still diffusely fractional. For the SH3 domain, the fully established interactions are mostly observed in turns that need to be formed before the rest of the chain can come together[8]. The results suggest that the diffuse transition state is an expanded version of the native structure[5], consistent with the observation that the folding rate constant can be correlated with the topological features of the native state[9].

In this study we demonstrate that the diffuse transition-state structure of the ribosomal protein S6 is not essential for successful folding but may be completely changed by circular permutation. The permutation links the N- and C-termini, which fold last in the wild type protein, and splits the original folding nucleus between either ends of the sequence[10]. As a consequence, the globally diffuse nucleus of the wild type protein polarizes into the linked region and becomes atypically condensed: the folding order is reversed and the nucleation takes place more locally than observed previously. At the same time, the permutant folds faster than wild type S6. The result shows that the diffuse transi-

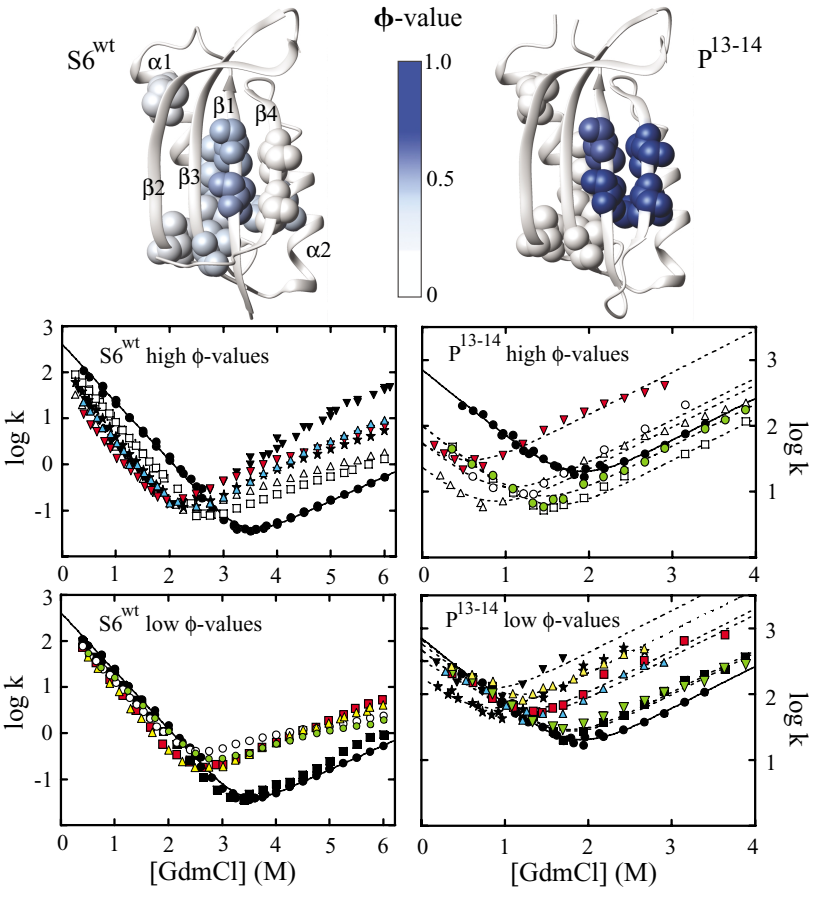

Fig. 1 The diffuse *versus* polarized distribution of ϕ-values in the transition state of $S6^{wt}$ (PDB entry 1RIS) and the circular permutant P^{13-14}. In $S6^{wt}$, the ϕ-values are overall fractional, radiating from a nucleation center between β-strand 1 and helix 1 (top). In P^{13-14}, the ϕ-values tend to be either 1 or 0, and their distribution reveals an atypically condensed and polarized transition state shifted to the helix-2 region. In the chevron plots used for the analysis (bottom), rate constants are in units of s^{-1}. Residues are numbered according to the wild type sequence: $S6^{wt}$ or P^{13-14} (black circle), V6A (open triangle), I8A (red triangle), L19A (red square), I26A in $S6^{wt}$ or I26V in P^{13-14} (blue triangle), L30A (closed triangle), V37A (yellow triangle), V40A (black square), V65A (black star), V72A (green triangle), L75A (open square), V88A (green circle) and V90A (open circle). In most cases, mutations on $S6^{wt}$ affect both chevron limbs, whereas the same mutations on P^{13-14} cause changes of either the refolding- or the unfolding-limb. This means that the ϕ-values of $S6^{wt}$ are predominantly fractional, whereas the ϕ-values of P^{13-14} tend to be either 1 or 0 (Table 1). The only clear exceptions are I8A (red triangle) and V65A (black star), which display fractional ϕ-values in both $S6^{wt}$ and P^{13-14}.

tion state is not a requirement of the folding physics but may be under biological control.

Diffuse to polarized transition state structure

The S6 structure[11] consists of two halves that are mirror images if the difference in chain connectivity is disregarded: the helix-1 side is composed of interactions that are close in sequence, whereas the helix-2 side is split between the N- and C-termini (Fig. 1). To investigate the effect of changing the loop lengths of this entropically biased fold, we have compared the transition-state structures of wild type S6 ($S6^{wt}$) and a circular permutant ($P^{13–14}$) (ref. 10) that links the N- and C-termini and cuts the wild type nucleus by an incision between residues 13 and 14 (Fig. 1). Wild type enumeration is used for both proteins. The ϕ-value analysis of $S6^{wt}$ reveals a characteristically diffuse transition state of fractional contacts that encompasses most of the native structure[12]. The center of the folding nucleus is composed of residues Ile 8 (ϕ = 0.46) and Ile 26 (ϕ = 0.40) connecting β-strand 1 and α-helix 1 through the center of the hydrophobic core[12]. The least structured region of the wild type transition state is the entropically disfavored interface between the N- and C-terminal β-strands 1 and 4 (Fig. 1) that consolidates mainly on the downhill side of the folding barrier. All ϕ-values of $S6^{wt}$ are <0.5 at the transition midpoint (Table 1).

Table 1 ϕ-values, transition midpoints, $\Delta\Delta G_{D\text{-}N}$ and L^{mean} for mutations on $S6^{wt}$ and $P^{13–14}$

Mutant[1]	ϕ-value[2]		Midpoint (M)[3]		$\Delta\Delta G_{D\text{-}N}$ (kcal mol^{-1})[4]		L^{mean}[5]	
	$S6^{wt}$	$P^{13–14}$	$S6^{wt}$	$P^{13–14}$	$S6^{wt}$	$P^{13–14}$	$S6^{wt}$	$P^{13–14}$
wt	–	–	3.34	1.81	–	–	–	–
V6A	0.52	0.92	1.97	0.45	2.73	2.26	63	32
I8A	0.46	0.63	1.89	0.29	3.14	2.85	59	37
L19A	0.24	0.17	2.44	1.12	1.39	1.51	39	52
I26A	0.40	–	2.27	–	1.69	–	36	–
I26V	–	0.15	–	1.32	–	0.98	–	56
L30A	0.34	0.13	2.08	0.78	1.35	1.70	36	43
V37A	0.24	0.07	2.37	1.07	1.66	1.74	23	23
V40A	–	–0.05	3.11	1.65	–0.08	0.28	–	21
V65A	0.38	0.20	2.20	1.04	1.76	0.67	42	31
V72A	–	0.04	–	1.45	–	1.10	–	24
L75A	0.40	1.55	2.60	1.31	1.44	0.76	42	37
V88A	0.14	1.10	2.59	1.30	2.51	0.82	63	14
V90A	0.14	0.70	2.39	0.94	2.85	1.27	56	15

[1]For the complete data set of $S6^{wt}$ see ref. 12. '–' indicates that the value is not determined.
[2]ϕ-values are calculated according to Eq. 1.
[3]Transition midpoints are derived from the intersect between log k_f and log k_u in the chevron plots (Fig. 1).
[4]$\Delta\Delta G_{D\text{-}N}$ for nonpermuted S6 is calculated according to Eq. 4. Notably, these values of $\Delta\Delta G_{D\text{-}N}$ calculated from individual m-values in some cases differ slightly from those determined by equilibrium unfolding and the average m-value. This variation does not affect the result and conclusions of this study. $\Delta\Delta G_{D\text{-}N}$ for $P^{13–14}$ is estimated by linear extrapolation of log k_f and log k_u to 0 M [guanidinium-Cl].
[5]L^{mean} shows the data plotted in Fig. 2 and is calculated for mutations with a solvent exposure of <20% and c = 10 (Eq. 2).

The permutant $P^{13–14}$ shows a different nucleation pattern: the ϕ-values tend to cluster around 0 or 1. The behavior is clearly seen by direct comparison of the chevron data for the wild type and permutant proteins (Fig. 1). With only one exception (I8A), mutations on $P^{13–14}$ affect either the refolding or unfolding limb (log k_f or log k_u, respectively), rather than both, as observed with $S6^{wt}$. In addition, the nucleus of $P^{13–14}$ is abruptly shifted towards the loop connecting the former N- and C-termini. Residues Leu 75 (helix 2), Val 88 and Val 90 (strand 4), which are largely unstructured in the transition state of $S6^{wt}$, all show ϕ-values near 1 in $P^{13–14}$. Conversely, Ile 26 and Leu 30 in helix 1, which play a central role in the $S6^{wt}$ nucleus, show ϕ-values of close to 0 in $P^{13–14}$. The residues in β-strand 1 (Val 6 and Ile 8) are part of both the wild type and the P^{13-14} nucleus, although they form interactions towards different sides of the S6 structure. I8A and, to a lesser extent, V65A are the only mutations with intermediate ϕ-values (Fig. 1; Table 1). Thus, the folding nucleus of $P^{13–14}$ has not only shifted to a part of the protein that was previously unstructured in the transition state, but it has also collapsed into an unusually compact and polarized species that is atypical for naturally evolved two-state proteins. The reason for such a radical change is not clear from the alterations of loop entropy alone: other two-state proteins maintain diffuse transition state structures despite considerable variation in loop entropy[1].

Large loops balanced by strong contacts

To see if the residue contact energies are uniformly distributed in S6 or balanced against topological features, we have mapped out the change in stability upon point mutation ($\Delta\Delta G_{D\text{-}N}$) in various parts of the wild type structure (Fig. 2*a*). There is a bias of the $\Delta\Delta G_{D\text{-}N}$ values: mutation of residues connecting the N- and C-termini causes larger stability loss than in the rest of the protein. In an attempt to capture the effect graphically, we have plotted $\Delta\Delta G_{D\text{-}N}$ against the average sequence separation between the deleted contacts, L^{mean} (Eq. 2). Calculation of L^{mean} depends on two parameters: the contact radius (set to 5.0 Å), and the cluster size — that is, the number of local residues at either side of the mutation that are excluded. If the cluster size (c) is set too small, local interactions will dominate L^{mean} and diminish the difference in lost long-range contacts for the mutations. If c is set too large, there will be too few interactions left. The best correlation between $\Delta\Delta G_{D\text{-}N}$ and L^{mean} (R = 0.76) is obtained for mutations of buried side chains (with a solvent exposure of <20%) and c = 10 (Fig. 2*b*). Upon inclusion of solvent-exposed mutations, the correlation is still apparent but goes down to R = 0.68. The most destabilizing mutations (blue data points) are clustered at the highest values of L^{mean} (Fig. 2*b*). The trend is completely erased by the permutation (R = 0.12), where the residues with high $\Delta\Delta G_{D\text{-}N}$ are moved close in sequence (Fig. 2*c*). When $\Delta\Delta G_{D\text{-}N}$ is plotted against the loss of loop entropy in the Flory form[13], ΔS^{Flory} (Eq. 3), R goes down to 0.43 for $S6^{wt}$ (Fig. 2*d*) and to 0.25 for $P^{13–14}$ (data not shown). The reason for this decrease seems to be that the value of ΔS^{Flory} is coupled to the protein density[2] (d) (R = 0.88, Fig. 2*e*), which, in turn, shows a relatively weak correlation with $\Delta\Delta G_{D\text{-}N}$ (R = 0.58, Fig. 2*f*). For comparison, the combined data on the two-state proteins ADA 2h[14], ACBP[15], CI2 (ref. 3), src SH3 domain[6], suc1 (L. Itzaki, F. Rousseau and J. Schymkowitz, pers. comm.), U1A[4] and villin[16] yield a slightly better correlation between $\Delta\Delta G_{D\text{-}N}$ and d (R = 0.64, Fig. 3*a*), which is in good agreement with previous investigations[17]. Consistently, the reference proteins also display a better correlation between $\Delta\Delta G_{D\text{-}N}$ and ΔS^{Flory} (R = 0.59, Fig. 3*b*), coupled by a correlation between

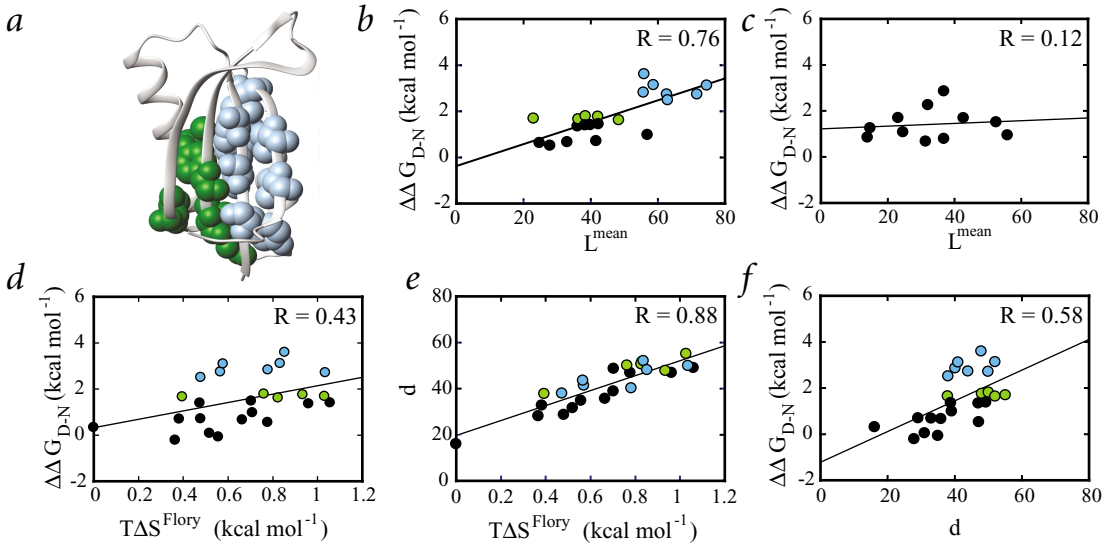

Fig. 2 The distribution of strong side chain interactions in the S6 structure. ***a***, The effect of mutation on protein stability ($\Delta\Delta G_{D\text{-}N}$) reveals that the strongest contacts are clustered at the entropically penalized interface to helix 2 and β-strand 4. Blue is $\Delta\Delta G_{D\text{-}N} > 2.5$ kcal mol^{-1}; and green, $2.5 > \Delta\Delta G_{D\text{-}N} > 1.5$ kcal mol^{-1}. ***b***, A graphical representation of the dispersed contact energies is obtained in a plot of $\Delta\Delta G_{D\text{-}N}$ *versus* L^{mean} — that is, the average sequence separation between the contacts lost upon mutation — for buried mutations with <20% side chain exposure (Eq. 2). Blue and green symbols mark residues highlighted in the protein structure, and the fitted line is y = 0.047x – 0.36 and R = 0.76. ***c***, In the permutant P^{13-14}, the correlation between $\Delta\Delta G_{D\text{-}N}$ and L^{mean} vanishes; y = 0.006x + 1.22 and R = 0.12. ***d***, Plot of $\Delta\Delta G_{D\text{-}N}$ against the Flory loop entropy ($T\Delta S^{Flory}$) for $S6^{wt}$; y = 1.83x + 0.32 and R = 0.43. ***e***, Plot of d *versus* $T\Delta S^{Flory}$ for $S6^{wt}$; y = 32.5x + 19.7 and R = 0.88. ***f***, Plot of $\Delta\Delta G_{D\text{-}N}$ *versus* protein density (d) for $S6^{wt}$; y = 0.07x – 1.22 and R = 0.58.

ΔS^{Flory} and d that is identical to that of S6 (R = 0.88, Fig. 3*c*). Even so, we observe no significant relation between $\Delta\Delta G_{D\text{-}N}$ and L^{mean} for these proteins (R = 0.41, Fig 3*d*), with the possible exception of the entropically biased structure of ACBP (R = 0.57) (data not shown). ACBP shows also a biased distribution of $\Delta\Delta G_{D\text{-}N}$ similar to that of the S6 structure (Fig. 2). Finally, the reference proteins reveal a weak correlation between d and L^{mean} (R = 0.63, Fig 3*e*), supporting the idea that there is a general preference to anchor long loops in the core[18].

Symmetry aids shift of the folding nucleus

The effect of circular permutation is different for different proteins. The transition state of the 64-residue protein CI2 does not respond to permutation[19], whereas the relatively polarized transition state of SH3 from α-spectrin displays a slight redistribution when the incision cuts the wild type nucleus[20]. When SH3 is cut in the late forming RT-loop, permutation has only minor effects. A similar behavior is observed in computer simulation: a lattice permutant that splits the 'wild type' nucleus produces distinct changes of the transition state structure, whereas a permutation that cuts in a region that is unstructured in the 'wild type' transition state shows negligible effect[21].

The radical shift of the transition state of P^{13-14} is probably coupled to the symmetry of the S6 structure. In $S6^{wt}$, the diffuse folding nucleus is biased towards the helix-1 side, possibly because the entropic cost of forming this side is smaller than for the helix-2 side. A comparable nucleus is found for the structurally analogous protein U1A[4], whereas another analog, ADA 2h, shows the highest ϕ-values in the helix-2 side[14]. This disparity suggests that the structural family may nucleate in either region, despite the entropic bias towards the helix-1 side[4]. Symmetry-related shifts of the folding nucleus are also seen in protein L and G[7], can be inferred from theory[21–23] and may be induced by computer-based redesign[24]. It is then not surprising that the permutation P^{13-14}, which more or less swaps the loop entropy between these two prospective nucleation sites, shifts the nucleus to the helix-2 side. Perhaps, the effect of permuting CI2 or the SH3 domain is smaller because there are no such alternative nuclei within energetic reach.

Contact energies control nucleation pattern

A persuasive explanation for the diffuse nucleus of $S6^{wt}$, as well as for the abrupt polarization of P^{13-14}, is provided by the biased contact energies (Fig. 2). Plotkin and Onuchic[25] recently demonstrated that such compensation of the loop entropy contributes to equalizing the free energy of the interaction along the native interactions so that any part of the protein may fold with similar probability. Proteins where the contact energies are set to balance the loop entropy display broad transition state ensembles where numerous native contacts participate at a fractional level[25,26]. When strong contacts are set to coincide with low entropy, the transition state ensemble became much narrower and the contact pattern discretely polarized with ϕ-values of either 0 or 1 (ref. 25). The results are in agreement with the experimental observations on S6 (Figs 1, 2).

Diffuse nuclei optimize cooperativity

Thus, S6 may nucleate and fold in at least two different ways without significantly compromising either the refolding rate constant or the native structure. S6 may also fold by a third, overall compact trajectory that includes a misfolded intermediate upon addition of sulfate or by mutation of certain gatekeeper residues[27]. The results show that the diffuse transition state of S6 is not required for successful folding and presents the possibility that it is produced in response to other, selective pressures. What could then be the biological advantage of a diffuse transition state? An explanation is hinted by the way the transition state controls the dynamic properties of

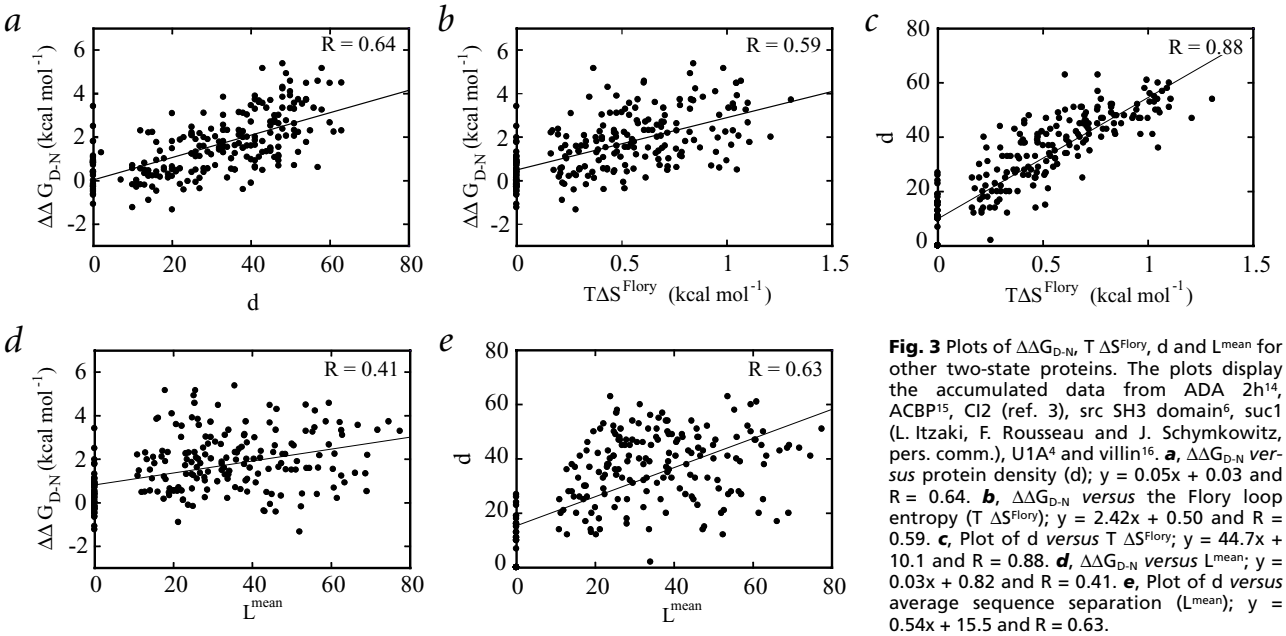

Fig. 3 Plots of $\Delta\Delta G_{D\text{-}N}$, $T\,\Delta S^{Flory}$, d and L^{mean} for other two-state proteins. The plots display the accumulated data from ADA 2h[14], ACBP[15], CI2 (ref. 3), src SH3 domain[6], suc1 (L. Itzaki, F. Rousseau and J. Schymkowitz, pers. comm.), U1A[4] and villin[16]. ***a***, $\Delta\Delta G_{D\text{-}N}$ *versus* protein density (d); y = 0.05x + 0.03 and R = 0.64. ***b***, $\Delta\Delta G_{D\text{-}N}$ *versus* the Flory loop entropy ($T\,\Delta S^{Flory}$); y = 2.42x + 0.50 and R = 0.59. ***c***, Plot of d *versus* $T\,\Delta S^{Flory}$; y = 44.7x + 10.1 and R = 0.88. ***d***, $\Delta\Delta G_{D\text{-}N}$ *versus* L^{mean}; y = 0.03x + 0.82 and R = 0.41. ***e***, Plot of d *versus* average sequence separation (L^{mean}); y = 0.54x + 15.5 and R = 0.63.

the native state. The globally diffuse transition state with its extensive network of native interactions implies that few regions of the native state may unfold locally unless the protein is exited over the top of the barrier (Fig. 4). Cooperativity is optimized and native fluctuations reduced. Evidence that native fluctuations can be biologically disadvantageous was recently provided by a disease-associated mutation of human lysozyme that seemed to trigger erroneous aggregation and amyloid formation *in vitro* by exposing certain contiguous stretches of sheet residues[28]. Similar to here, the authors point at the possibility that cooperativity is an evolved feature and an essential requirement for biological function. In further support of this idea, the amyloidogenic properties of β_2-microglobulin are observed to stem from local unfolding of the N-and C-terminal region[29]. Notably, S6 is susceptible to similar local unfolding phenomena upon mutation in the entropically penalized interface to the α2–β4 region[12]. Because the α2–β4 region is vital for specific binding to the ribosomal RNA[30], excessive floppiness in this part of the structure likely impairs function. The role of the energetic bias could be to prevent this from happening by integrating the affected interactions in the folding nucleus. In other instances, it is easy to imagine that certain parts of the native protein need to be flexible for functional reasons. Such dynamic regions would then be required to be outside the folding nucleus and would be indicated by heterogeneous transition state structures.

Special adaptation to entropy-biased fold

Although there are several reasons to suspect that the diffuse transition state structure is under biological influence in other two-state proteins, there are few indications that the tuning is achieved in the same way as in S6. For the reference proteins (Fig. 3), the correlation between $\Delta\Delta G_{D\text{-}N}$ and average sequence separation (L^{mean}) is hardly significant at R = 0.41, even though

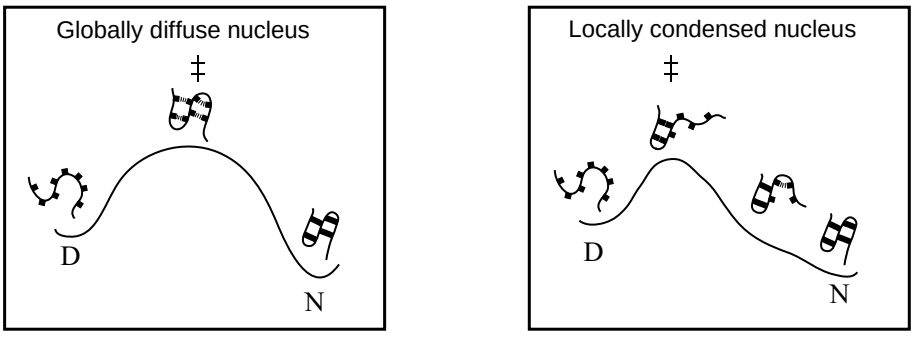

Fig. 4 Schematic illustration of the folding free-energy profiles for a globally diffuse nucleus and a locally condensed nucleus. The diffuse nucleus captures all native contacts in the transition state, whereas the condensed nucleus involves only a small region of the native structure. The parts excluded from the condensed nucleus will form on the downhill side of the folding barrier and be at risk to unfold locally by fluctuations in the ground state. In energetic terms, small nuclei (arising in systems dominated by local contacts) can be inferred from theory to produce early transition states and relatively low energetic gradients towards the native basin[25]. As a side effect, however, the integrity of the native structure will be poor because even minor fluctuations could lead to population of partly structured states — that is, local unfolding. At the extreme, the barrier vanishes completely and all structural cooperativity is lost. *Vice versa*, diffuse nuclei (arising in systems dominated by long-range contacts) produce the highest barriers and, hence, the most distinct dynamic enclosures of the native state[25,31]. Notably, such a dynamic enclosure will contribute to kinetically safeguard the native structure and to compensate for the overall low thermodynamic stability of the native state: rate is traded for structural specificity.

most proteins acquire a similar positive slope. Nevertheless, there is a weak correlation between L^{mean} and d (R = 0.63), indicating that the longest sequence loops are generally tied together in the hydrophobic core (Fig. 3). Consistently, database surveys have shown that strong contacts — that is, residue pairs with >15 side chain contacts — are enriched at both short and long sequence separations[18]. The phenomenon was suggested to reflect the evolutionary optimization of the folding kinetics and thermodynamic stability[18,31]. The current data provide an alternative explanation for the concurrence of good packing and long sequence separations: optimization of native state rigidity. Bringing together the longest loops with hydrophobic residues in the core could constitute a topological strategy to achieve diffuse nuclei that is more common than the polarized $\Delta\Delta G_{D\text{-}N}$ distribution (Fig. 2). Thus, S6 may represent a special case, in which functional limitations, as well as the symmetric but entropically biased topology, have favored the modulation of the contact energies to achieve uniform stability.

Conclusions

The complete change of the S6 transition state upon circular permutation shows that the diffuse transition state of the wild type protein is not required for folding *per se* but is probably under biological control. $S6^{wt}$ maintains the diffuse transition state structure by biasing strong contacts to the entropically penalized interface between the C- and N-termini. When this bias is broken by circular permutation, the folding nucleus polarizes towards the artificially linked C- and N-termini. The observations are in agreement with predictions from theory[25]. We speculate that the biological advantage of the generally observed diffuse transition state structure is to optimize folding cooperativity and safeguard the structural integrity of the native state. The mechanism by which S6 promotes the diffuse transition state, however, is not general but represents a specific adaptation to the symmetric structure and functional requirements of this protein.

Methods

Mutagenesis and kinetic analysis. Mutagenesis, expression and chemicals were as described[10]. Buffer was 50 mM MES, pH 6.3, at 25 °C, and stopped-flow analysis was done as described[10]. To reduce the effects of kinetic curvatures and varying m-values, ϕ-values[2] were calculated from chevron data at the transition midpoint according to[12]:

$$\phi = (\log k_f^{wt:0.5M} - \log k_f^{mut:0.5M}) / ((\log k_f^{wt:0.5M} - \log k_f^{mut:0.5M}) + (\log k_u^{mut:3M} - \log k_u^{wt:3M})) \quad (1)$$

where the superscript wt:0.5M and mut:0.5M refers to the guanidinium-Cl concentrations at which the wild type and mutant data were taken.

Calculation of topological parameters. The average sequence separation between deleted contacts (L^{mean}) was calculated according to:

$$L^{mean} = (1 / n)\Sigma L^i \quad (2)$$

where L^i is the separation in sequence (loop length) between the carbon–carbon contacts (<5 Å radius) lost upon mutation, and n is the number of contacts lost. Changes in loop entropy were estimated by standard polymer theory[13]:

$$\Delta S^{Flory} = (-3 / 2R)\Sigma L^i_F \quad (3)$$

where L^i_F is the lengths of the loops connecting all the residues interacting with the deleted moiety within a 5 Å radius.

Protein density (d) was defined as the number of methylene groups within 5 Å of the deleted carbon atoms. Solvent exposure was calculated in MolMol[32] (solvent radius = 1.4 Å), which was also used for the protein graphics.

Stability analysis. Protein stability was derived from chevron data[12] according to

$$\Delta G_{D\text{-}N} = (m_u^{3\text{–}5M} - m_f)\,[\text{guanidinium-Cl}]^{50\%} = m_{D\text{-}N}[\text{guanidinium-Cl}]^{50\%} \quad (4)$$

where $[\text{guanidinium-Cl}]^{50\%}$ is the transition midpoint measured by equilibrium unfolding, m_f is the slope of log k_f and $m_u^{3\text{–}5M}$ is the slope of log k_u estimated from data between 3 and 5 M guanidinium-Cl to reduce effects of curvatures at high [guanidinium-Cl] (ref. 12).

Acknowledgments

We thank S. Plotkin, P. Wolynes and D. Thirumalai for stimulating discussions and Katarina Wallgren for technical assistance. The work was supported by the Swedish Research Council.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence should be addressed to M.O. *email: mikael.oliveberg@chem.umu.se*

Received 7 June, 2002; accepted 12, September, 2002.

1. Jackson, S.E. *Folding Des.* **3**, R81–91 (1998).
2. Fersht, A.R. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (W.H. Freeman, New York; 1999).
3. Itzhaki, L.S., Otzen, D.E. & Fersht, A.R. *J. Mol. Biol.* **254**, 260–288 (1995).
4. Ternstrom, T., Mayor, U., Akke, M. & Oliveberg, M. *Proc. Natl. Acad. Sci. USA* **96**, 14854–14859 (1999).
5. Fersht, A.R. *Proc. Natl. Acad. Sci. USA* **92**, 10869–10873 (1995).
6. Grantcharova, V.P., Riddle, D.S., Santiago, J.V. & Baker, D. *Nature Struct. Biol.* **5**, 714–720 (1998).
7. McCallister, E.L., Alm, E. & Baker, D. *Nature Struct. Biol.* **7**, 669–673 (2000).
8. Martinez, J.C., Pisabarro, M.T. & Serrano, L. *Nature Struct. Biol.* **5**, 721–729 (1998).
9. Plaxco, K.W., Simons, K.T. & Baker, D. *J. Mol. Biol.* **277**, 985–994 (1998).
10. Lindberg, M.O. *et al. J. Mol. Biol.* **314**, 891–900 (2001).
11. Lindahl, M. *et al. EMBO J.* **13**, 1249–1254 (1994).
12. Otzen, D.E. & Oliveberg, M. *J. Mol. Biol.* **317**, 613–627 (2002).
13. Flory, P.J. *J. Am. Chem. Soc.* **78**, 5222–5235 (1956).
14. Villegas, V., Martinez, J.C., Aviles, F.X. & Serrano, L. *J. Mol. Biol.* **283**, 1027–1036 (1998).
15. Kragelund, B.B. *et al. Nature Struct. Biol.* **6**, 594–601 (1999).
16. Choe, S.E., Li, L., Matsudaira, P.T., Wagner, G. & Shakhnovich, E.I. *J. Mol. Biol.* **304**, 99–115 (2000).
17. Chakravarty, S. & Varadarajan, R. *Structure Fold. Des.* **7**, 723–732 (1999).
18. Doyle, R., Simons, K., Qian, H. & Baker, D. *Proteins* **29**, 282–291 (1997).
19. Otzen, D.E. & Fersht, A.R. *Biochemistry* **37**, 8139–8146 (1998).
20. Viguera, A.R., Serrano, L. & Wilmanns, M. *Nature Struct. Biol.* **3**, 874–880 (1996).
21. Li, L. & Shakhnovich, E.I. *J. Mol. Biol.* **306**, 121–132 (2001).
22. Wolynes, P.G. *Proc. Natl. Acad. Sci. USA* **93**, 14249–14255 (1996).
23. Thirumalai, D. & Klimov, D.K. *Folding Des.* **3**, R112–118 (1998).
24. Nauli, S., Kuhlman, B. & Baker, D. *Nature Struct. Biol.* **8**, 602–605 (2001).
25. Plotkin, S.S. & Onuchic, J.N. *Proc. Natl. Acad. Sci. USA* **97**, 6509–6514 (2000).
26. Shea, J.E., Onuchic, J.N. & Brooks, C.L. III *J. Chem. Phys.* **113**, 7663–7671 (2000).
27. Otzen, D.E. & Oliveberg, M. *Proc. Natl. Acad. Sci. USA* **96**, 11746–11751 (1999).
28. Canet, D. *et al. Nature Struct. Biol.* **9**, 308–315 (2002).
29. McParland, V.J., Kalverda, A.P., Homans, S.W. & Radford, S.E. *Nature Struct. Biol.* **9**, 326–331 (2002).
30. Agalarov, S.C., Sridhar Prasad, G., Funke, P.M., Stout, C.D. & Williamson, J.R. *Science* **288**, 107–113 (2000).
31. Abkevich, V.I., Gutin, A.M. & Shakhnovich, E.I. *J. Mol. Biol.* **252**, 460–471 (1995).
32. Koradi, R., Billeter, M. & Wüthrich, K. *J. Mol. Graph.* **14**, 51–55 (1996).