doi:10.1016/j.jmb.2006.10.079 J. Mol. Biol. (2007) 365, 1559–1577

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

The Transition State of the *ras* Binding Domain of Raf Is Structurally Polarized Based on Φ-Values but Is Energetically Diffuse

F.-X. Campbell-Valois[1,2] and S. W. Michnick[1]*

[1]*Département de Biochimie Université de Montréal C.P. 6128, Succ. centre-ville Montréal, Québec Canada H3C 3J7*

[2]*Programme de Biologie Moléculaire, Université de Montréal, C.P. 6128 Succ. centre-ville, Montréal Québec, Canada H3C 3J7*

**Corresponding author*

The *ras* binding domain (RBD) of the Ser/Thr kinase c-Raf/Raf-1 spans 78 residues and adopts a structure characteristic of the β-grasp ubiquitin-like topology. Recently, the primary sequence of Raf RBD has been nearly exhaustively mutated experimentally by insertion of stretches of degenerate codons, which revealed sequence conservation and hydrophobic core organization similar to that found in an alignment of β-grasp ubiquitin-like proteins. These results now allow us to examine the relationship between sequence conservation and the folding process, particularly viewed through the analysis of transition state (TS) structure. Specifically, we present herein a protein engineering study combining classic truncation (Ala/Gly) and atypical mutants to predict folding TS ensemble properties. Based on classical Φ-value analysis, Raf RBD TS structure is particularly polarized around the N-terminal β-hairpin. However, all residues constituting the inner layer of the hydrophobic core are involved in TS stabilization, although they are clearly found in a less native-like environment. The TS structure can also be probed by a direct measure of its destabilization upon mutation, $\Delta\Delta G_{U-\ddagger}$. Viewed through this analysis, Raf RBD TS is a more diffuse structure, in which all residues of the hydrophobic core including β-strands 1, 2, 3 and 5 and the major α-helix play similar roles in TS stabilization. In addition, Φ-values and $\Delta\Delta G_{U-\ddagger}$ reveal striking similarities in the TS of Raf RBD and ubiquitin, a structural analogue displaying insignificant sequence identity (<12%). However, ubiquitin TS appears more denatured-like and polarized around the N-terminal β-hairpin. We suggest that analysis of Φ-values should also consider the direct impact of mutations on differences in free energy between the unfolded and TS ($\Delta\Delta G_{U-\ddagger}$) to ensure that the description of TS properties is accurate. Finally, the impact of these findings on the modeling of protein folding is discussed.

Keywords: protein folding; chevron curves; protein engineering; Φ-value; Raf *ras* binding domain (RBD)

Introduction

At the heart of the protein folding problem lies the search for unifying principals that adequately describe the processes by which a polypeptide chain spontaneously folds from the denatured state into a unique three-dimensional structure. The discovery of apparently two-state folding proteins has put much emphasis in the last decade on the study of the only specie that is rate limiting in such a system,[1] i.e. the transition state (TS). The application of protein engineering to folding through the development of the Φ-value analysis method has provided a framework for straightforward interpretations of the importance of a given amino acid residue in stabilizing the TS.[2] The importance of the protein engineering methods in the development of our contemporary view of

Abbreviations used: RBD, *ras* binding domain; TS, transition state; β1, β-strand 1; β2, β-strand 2; α1, α-helix; β3, β-strand 3; β4, β-strand 4; β5, β-strand 5; N-terminal, amino-terminal; C-terminal, carboxy-terminal; Gdm-HCl, guanidinium-hydrochloride; AcP, acylphosphatase; CI2, chymotrypsin inhibitor 2.

E-mail address of the corresponding author: stephen.michnick@umontreal.ca

protein folding is clear,[3–18] despite the need to address mounting concerns regarding the interpretation of Φ-values.[19–23] Among the key questions are the precision and reproducibility of Φ-values, the comparability of the TS characteristics of distinct polypeptide chains, the indirect assessment of topological element formation and the role of non-native contacts in TS stabilization.

The role of a residue during the folding process is assessed by the measurement of variations in folding or unfolding rate and hence, TS stability, induced upon insertion of non-disruptive mutations (e.g. Ala or Gly in the case of wt Ala residues). Classically, the change in free energy of activation ($\Delta\Delta G_{U\text{-}\ddagger}$ or $\Delta\Delta G_{F\text{-}\ddagger}$) that is directly derived from folding/unfolding rates is normalized by the change in free energy ($\Delta\Delta G_{F\text{-}U}$), yielding the Φ-value, which indicates the native-likeness of contacts established by the atoms deleted in the mutant. A collection of Φ-values obtained for a significant number of residues allows for delineating a TS structural properties. The structures of protein TS are often described as polarized or diffuse depending on the topological distribution of residues having high *versus* low Φ-values. Delocalized TS has been very well described for the important model CI2 and several other proteins,[4,9,16,24] while polarized TS structures have been described for barnase, two SH3 domains, protein-L and more recently a cold shock protein and ubiquitin.[3,6,7,15,24–26] These categorizations of TS structures are reminiscent of the models for folding known as nucleation–condensation and framework, respectively. However, it is not clear that this classification of the TS is totally free from artifacts that could be introduced by the experiments and parameters utilized to describe the TS.

A useful way to approach the understanding of TS is to compare their structural properties among proteins that have the same topology. For instance, the functionally homologous SH3 domains of src, spectrin and fyn share TS displaying similar characteristics.[6–8] The combination of Φ-value analysis and molecular dynamics suggest high similarity of the TS of three members of the homeodomain superfamily. However, their pathways to the native states are divergent, folding in manners consistent with framework, nucleation–condensation or hybrid models.[18] This study suggests a continuum among folding models, in which the framework or hydrophobic collapse models are simply extreme cases of nucleation–condensation that occur when the secondary structure or long-range interactions, respectively, are overstabilized.[27] Two members of the immunoglobulin-like Greek key fold with ~10% sequence identity were shown to display striking similarities in the dispersion of the most significant Φ-values,[12,13], although simulations confirmed large differences in the level of compaction of their TS.[28] The activation domain of procarboxypeptidase A2 and muscle acylphosphatase (AcP) (~10% sequence identity), both members of the AcP topology, possess similar folding nuclei.[9] Nevertheless, this clear-cut effect of the native state topology on the definition of the folding process is being challenged by contrasting results for other members of the AcP and SH3 folds.[10,29] Also, the symmetrically organized IgG binding domains of protein-L and G display notable differences in the structure of their TS as revealed by the contrasting roles of their N and C terminus β-hairpins.[15,16] Interestingly, the combination of 11 mutations was sufficient to shift the nucleus TS from the C to the N terminus β-hairpin of protein-G.[30] These results suggest that some folds could form through distinct pathways and TS. This hypothesis is attractive, because it could provide an explanation for the structural and functional versatility of the topologies that are recurrent in the natural protein universe. Several factors might explain the diversity in the TS characteristics observed in structural analogues, including variation in the secondary structure propensities, H-bonds network, ratio of hydrophobic amino acids, hydrophobic core organization and overall stability of the native state. Discrepancies in these characteristics become more important as the sequence identities diminish between pairs of structurally analogous proteins. To date, only seven topologies have at least two of their members that have been thoroughly characterized by Φ-value analysis.[31] Thus, folding studies of more structural superfamilies could shed additional light on the relationship between sequence and folding mechanisms. The most frequently occurring topologies are particularly interesting because they enable the comparison of structural analogues with widely varying levels of sequence identity.

The *ras* binding domain of c-Raf/Raf-1 (Raf RBD) Ser/Thr kinase is a member of the highly populated and functionally diverse β-grasp ubiquitin-like topology (aka ubiquitin-roll topology), according to the Structural Classification of Proteins database (SCOP)†. This domain is composed of 78 amino acids and folds independently into a compact globular structure formed by the packing of an α-helix (α1) against a mixed β-sheet displaying the following connectivity: β2-β1-β5-β3-β4 (Figure 1(a) and (b)).[32] Recently, we reported an exhaustive sequence perturbation of Raf RBD. Most important for the current study was the observation that our results were consistent with evolutionary conservation of ubiquitin fold sequences and a bi-level organization of the hydrophobic core (Figure 1(c) and (d)).[33] These results constituted the basis for the design of mutants presented below.

We have engineered 51 mutants of Raf RBD that can be separated into two categories: (1) side-chain truncation (Ala/Gly), and (2) atypical, usually disruptive mutations. The resulting dataset, in combination with the insights gained from the sequence perturbation experiment, has allowed for testing different hypotheses concerning the relationships between sequence, function and evolutionary conservation of stability *versus* folding/unfolding

† http://scop.mrc-lmb.cam.ac.uk/scop/

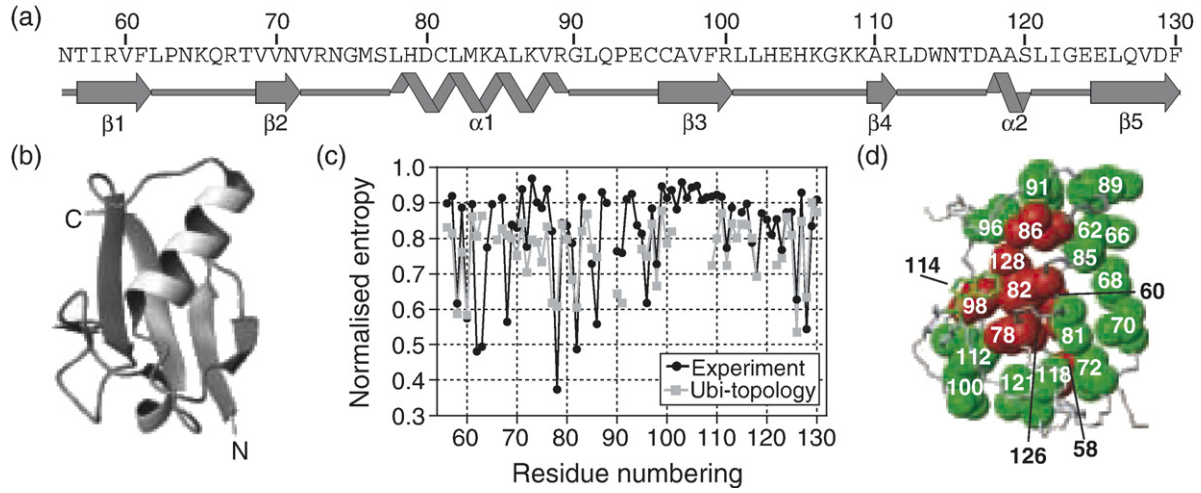

Figure 1. (a) Primary, secondary structure and (b) tertiary structure of Raf RBD. (c) Comparison of normalized entropy of Raf RBD obtained experimentally *versus* an alignment of 54 protein structures sharing the ubiquitin topology. (d) Hierarchical arrangement of the hydrophobic core into an inner (red) and outer core (green) as revealed by the sequence perturbation experiment.

kinetic.[34] Here, we report specifically the kinetic parameters of these mutants and the TS structure derived from Φ-values and a pure measure of TS destabilization ($\Delta\Delta G_{U-\ddagger}$). Strikingly, the structure of Raf RBD TS appears polarized based on interpretation of Φ-values, but diffusely distributed based on $\Delta\Delta G_{U-\ddagger}$. Further, we demonstrate that Raf RBD and ubiquitin fold through very similar TS.

Results

In total, 37 of the 78 residues of the Raf RBD were mutated to Ala (except wt Ala residues that were mutated to Gly). The main subset of residues was selected on the basis of their low tolerance to mutation in the sequence perturbation experiments.[33] These include all residues of the hydrophobic core, except for W114, which served as the fluorescent probe to follow folding/unfolding reactions (Figure 1(d)),[35] and some residues involved in the topological arrangement of the domain and in the binding interface with *ras*. Additional mutations were analyzed to cover all regions of the structure, including residues with solvent-exposed side-chains. Atypical mutations were also designed based on the literature (I58F, V72I and R89L), sequence alignments (A118L, Δ104-6 and Δ101-8+AG) or the Raf RBD sequence perturbation experiments (N56M, I58L, S77T, C81I, C96L, C96M, H2 and H2_F62L; see Materials and Methods and Campbell-Valois *et al.*[34] for description of the mutants).

The folding traces of wt Raf RBD in guanidinium-hydrochloride (Gdm-HCl) or urea could be fit to four exponential terms.[35,36] The three slower phases are resolved at below 3 M urea and they appear to be insensitive to denaturant concentration (Figure 2(a)).[36] The two-state transition is embodied by the fastest phase in the refolding traces of Raf RBD and hence, variation in its rate upon mutation is used to delineate the TS structure. The dependence of folding/unfolding rates on urea concentration was monitored by stopped-flow fluorescence spectroscopy (see Materials and Methods). The chevron curves of the various mutants are displayed in Figure 2(b)–(j)), according to their position in the defined secondary structure segments.

Accurate derivation of kinetic and thermodynamic parameters and Φ-values

The fast folding rate of Raf RBD and its relatively high resistance to urea denaturation can increase errors in k_f^{H2O} and k_u^{H2O} and thus, the rate of folding at 1.6 M ($k_f^{1.6\ M}$) and of unfolding at 5.8 M and 8 M ($k_u^{5.8\ M}$ and $k_u^{8\ M}$, respectively) are also reported (Table 1 and Supplementary Data, Table S1). The $k_f^{1.6\ M}$ value spans from 18 to 4092 s^{-1} (wt ~322 s^{-1}) for I58A and R89L, respectively. The $k_u^{5.8\ M}$ value spans from 0.11 to 57.0 s^{-1} (wt ~0.53 s^{-1}) for Δ104-6 and L78A, respectively. Importantly, these non-extrapolated parameters were used to derive more precise Φ-value estimates, as discussed below.

The average values of m_f and m_u are 1.25±0.15 and 0.39±0.06, respectively (Table 1 and Supplementary Data, Figure S1). The m_f and m_u values can be used to calculate the β-Tanford value [$\beta_t = m_f/(m_f+m_u)$], which indicates the TS position relative to the ground states. Variations in β_t upon perturbations (e.g. mutation, chemical, etc.) are utilized to describe TS shift by applying the Hammond or the rarer anti-Hammond postulates.[37–39] Twelve mutants of Raf RBD displayed a significant increase in β_t (e.g. $\beta_t > 0.8$; Supplementary Data, Table S1), indicating Hammond-like behavior. However, variations in the properties of the ground states must be ruled out before concluding that a true shift in the TS position is occurring. This can be tested indirectly by verifying that variations in m_f are compensated by an opposite change in m_u and *vice versa*.[24,40] In

contrast, the data on Raf RBD reveals a very good correlation of m_u+m_f *versus* m_f (slope = 0.86; $R = 0.89$), but not m_u (slope = 0.36; $R = 0.17$) (Figure 3(a)). Identical correlations of m_u+m_f *versus* m_f are observed for strictly Ala/Gly mutants (data not shown). Thus, the access of solvent to the denatured

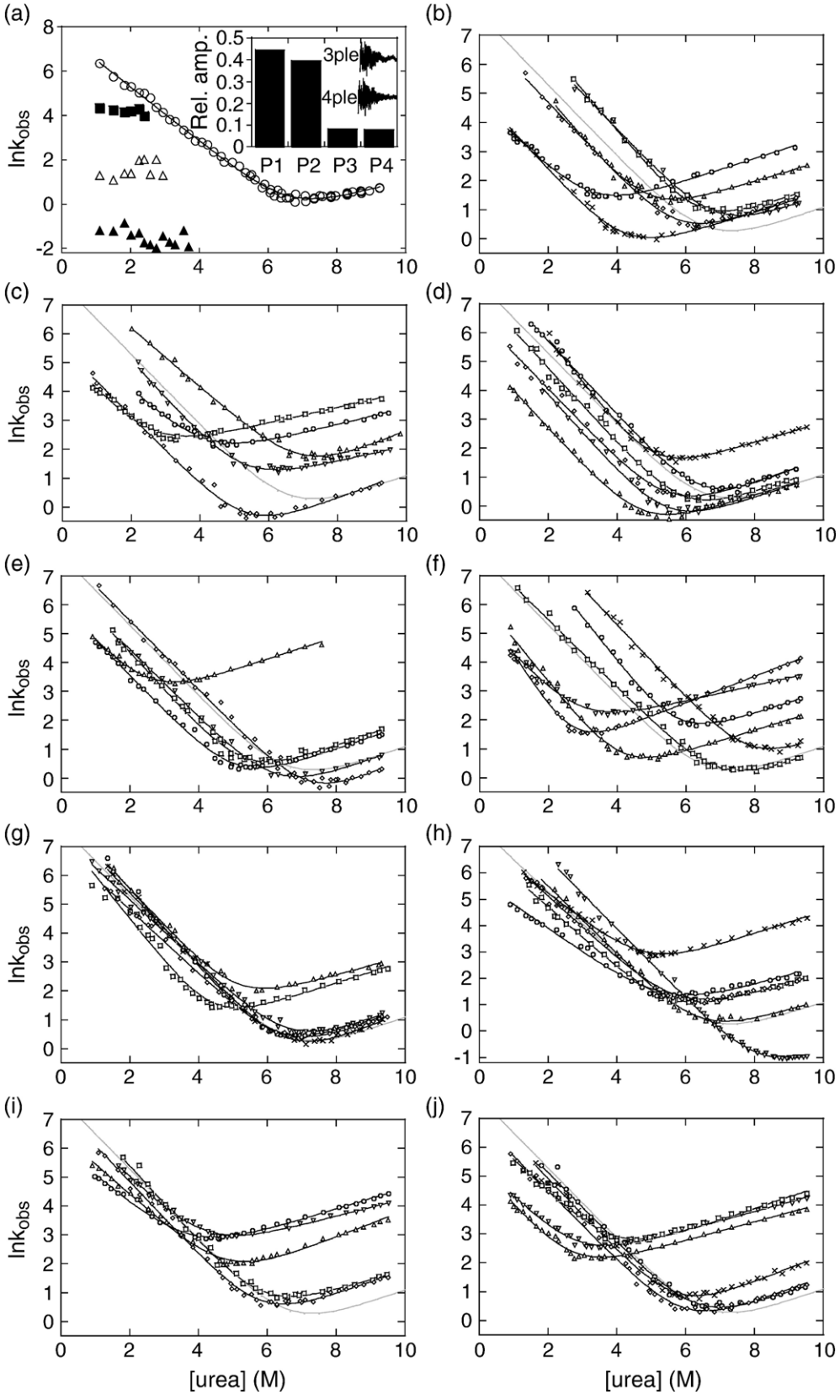

Figure 2 (*legend on opposite page*)

state, but not to the native state, is significantly affected upon mutation of this domain, making β_t an unreliable measure of TS shift.

Urea unfolding curves were obtained from the endpoint fluorescence signal of unfolding traces. The *m*-values obtained from equilibrium and kinetic data were similar overall, yielding mean values of 3.90±0.33 and 4.39±0.40, respectively (Table 2).[34] Three $\Delta\Delta G_{F\text{-}U}$ estimates were calculated, again taking care to reduce errors due to large extrapolations.[15,41] The $\Delta\Delta G^{kin}_{F\text{-}U}$ value was derived from $k_f^{1.6\ M}$ and $k_u^{8\ M}$ (see Material and Methods, equation (1)). The $\Delta\Delta G^{Cm}_{F\text{-}U}$ and $\Delta\Delta G^{5.8\ M}_{F\text{-}U}$ values were calculated using the mean *m*-values mentioned above and real *m*-values, respectively, and *Cm* (equations (2) and (3)) obtained from equilibrium experiments. The $\Delta\Delta G^{Cm}_{F\text{-}U}$ is more precise because it can be calculated from a smaller number of independent parameters. The kinetic and equilibrium estimates of the change in free energy correlate well (slope = 0.98, $R = 0.96$; Figure 3(b)), confirming the suitability of the two-state assumption to model Raf RBD folding. In addition, this observation suggests that the variations in the denatured state properties, as described above (Figure 3(a)), are energetically insignificant.

Three Φ-value estimates, similar to those described by Kim *et al.*[15], were obtained to provide rigorous internal controls. Briefly, Φ_F^{kin} was calculated from $k_f^{1.6\ M}$ and $\Delta\Delta G^{kin}_{F\text{-}U}$, Φ_F^{Cm}, from k_f^{H2O} and $\Delta\Delta G^{Cm}_{F\text{-}U}$ and 1-Φ_U, from $k_u^{5.8\ M}$ and $\Delta\Delta G^{5.8\ M}_{F\text{-}U}$, using equations (4), (5) and (6), respectively (Materials and Methods). The comparison of these estimates provides a supplementary assessment of experimental errors and of the two-state assumption. No atypical Φ-value ($\Phi < 0$ or $\Phi > 1$) was consistently observed in the three estimates for a given non-disruptive mutation once those with $\Delta\Delta G < |2|$ kJ mol^{-1} were excluded (Table 3). Overall, we found a good correlation between the Φ_F^{kin} and Φ_F^{Cm} (slope = 1.07; $R = 0.91$) or 1-Φ_U (slope = 1.03; $R = 0.91$) (Supplementary Data, Figure S2). The most significant discrepancies occur in Φ_F^{Cm}, which is predictable considering that fixed *m*-value and k_f^{H2O} were used in its calculation. The Φ_F^{exp}, which was calculated from fully extrapolated equilibrium and kinetic data, is also provided for reference (Supplementary Data, Table S3). We found that the average of the standard deviation to the mean of the three Φ-value estimates for Raf RBD mutants was comparable to protein-L as reported by Kim *et al.*[15] (0.41±0.11 and 0.26±0.10, respectively). Because of the smaller extrapolation used in their calculations, the Φ_F^{kin}-values (thereof Φ or Φ-value) are more precise (Table 3), particularly for less destabilizing mutations.[20] Hence, they are used in the structural description of Raf RBD folding TS and in all Figures related that follow.

Raf RBD TS structure is polarized according to Φ-value analyses

The overview of the distribution of Φ-values for Raf RBD suggests, according to their level of structure in the TS, a breakdown of the data in three regions: (1) the N-terminal β-hairpin; (2) the α1; and (3) the C-terminal region including β3, α2 and β5 (Figure 4; Φ of a given mutation appear in parenthesis in the Results, unless mentioned otherwise).

The N-terminal β-hairpin is the most structured region

The highest Φ-values occur mainly in the N-terminal β-hairpin, which spans residues T57–N71. The mutant V60A (0.82), in the middle of β1, has a higher Φ-value than I58A (0.54) and L62A (0.44), located at the edges of β1. The residues constituting the β-turn1, P63A and N64A, display low and very high Φ-values (0.32 and 0.98, respectively). The latter residue might be involved in forming side-chain backbone H-bonds that could stabilize the β-turn at the rate limiting step,[32] as suggested for N14 of protein-L.[15] Q66A and T68A found in a loop, which constitutes part of the *ras* binding interface, produced insignificant $\Delta\Delta G_{F\text{-}U}$, indicating that they are conserved solely for binding.[33,34] The mutants V70A (0.96) and V72A (0.98) that are located in β2 showed the highest Φ-values at hydrophobic core positions. V69A, which is located on the solvent exposed face of β2, showed Φ-value well above average (0.69). The β2 being extremely short, the high propensity of Val for β-strand might be important for its stabilization in the initial folding steps even at this exposed position. Taken together these results suggest that this β-hairpin is well structured in the TS, particularly in β2 and in the β-turn. However, the inner hydrophobic core residues of β1, which make direct contact with those of α1, are in a less native-like environment.

Figure 2. (a) Dependence on urea concentration of the rate of the three slowest phases modeled in the fitting of refolding traces of wt Raf RBD (■, △ and ▲ are P2, P3 and P4, respectively) and chevron curve obtained with the fastest refolding phase (○ is P1). The inset shows the relative amplitude of each phase at 2.6 M urea and the residuals for fitting this trace either with a triple or a quadruple exponential function (3ple and 4ple, respectively). Chevron curves of each mutant designed in this study follow in (b)–(j), according to their distribution in the native structure, starting from the N terminus. (b) β1: N56M (□), I58A (○), I58L (◊), I58F (△), R59A (▽) and V60A (×). (c) β-Turn1: L62A (□), P63A (○), N64A (◊), H2 (△) and H2_F62L (▽). (d) β2: Q66A (□), T68A (○), V69A (◊), V70A (△), V72A (▽) V72I (×). (e) N-capping residues and α1 N terminus: M76A (□), S77A (○), S77T (◊), L78A (△) and D80A (▽). (f) C terminus of α1: C81A (□), C81I (○), L82A (◊), A85G (△), L86A (▽) and R89L (×). (g) Loop following the α-helix: L91A (□), P93A (○), C95A (◊), C96A (△), C96L (▽) and C96M (×). (h) β3 and the following loop: A97G (□), V98A (○), R100A (◊), E104A (△), Δ104-6 (▽) and Δ101-8 (×). (i) β4, loop and α2: K109A (□), L112A (○), D117A (◊), A118G (△) and A118L (▽). (j) β5: L121A (□), E124A (○), E125A (◊), L126A (△), V128A (▽) and D129A (×). The modeled wt chevron curve (grey line) is shown for comparison in (b) through (j).

Table 1. Kinetic parameters

	$-m_f$ (M^{-1})	k_f^{H2O} (s^{-1})	$k_f^{1.6\ M}$ (s^{-1})	m_u (M^{-1})	$k_u^{5.8\ M}$ (s^{-1})	$k_u^{8\ M}$ (s^{-1})
Wt	1.22(±0.02)	2260(±140)	322(±12)	0.41(±0.04)	0.53(±0.05)	1.30(±0.04)
N56M	1.40(±0.04)	10,900(±1600)	1170(±110)	0.38(±0.05)	1.26(±0.15)	2.92(±0.10)
I58A	1.16(±0.06)	112(±10)	17.6(±0.6)	0.38(±0.02)	6.78(±0.25)	15.77(±0.59)
I58L	1.24(±0.03)	1300(±110)	178(±9)	0.43(±0.03)	0.96(±0.07)	2.49(±0.08)
I58F	1.40(±0.05)	2146(±370)	228(±21)	0.35(±0.03)	3.30(±0.20)	7.20(±0.22)
R59A	1.23(±0.03)	5960(±780)	831(±68)	0.39(±0.05)	0.86(±0.13)	2.03(±0.10)
V60A	1.29(±0.03)	142(±10)	18.0(±0.5)	0.37(±0.02)	1.10(±0.05)	2.48(±0.09)
L62A	1.29(±0.06)	205(±19)	26.2(±1.0)	0.27(±0.01)	17.45(±0.41)	31.50(±0.76)
P63A	1.14(±0.06)	568(±96)	92.2(±6.8)	0.29(±0.02)	9.22(±0.38)	17.44(±0.40)
N64A	1.21(±0.03)	260(±20)	37.7(±1.5)	0.44(±0.03)	0.51(±0.04)	1.35(±0.05)
H2	1.03(±0.02)	3870(±380)	749(±46)	0.49(±0.04)	1.79(±0.23)	5.26(±0.24)
H2_F62L	1.33(±0.04)	2320(±310)	279(±19)	0.27(±0.02)	2.72(±0.16)	4.89(±0.13)
Q66A	1.30(±0.03)	1600(±130)	201(±9)	0.40(±0.05)	0.67(±0.08)	1.62(±0.07)
T68A	1.27(±0.02)	3850(±310)	504(±19)	0.47(±0.06)	0.74(±0.10)	2.10(±0.09)
V69A	1.23(±0.03)	680(±56)	94.9(±4.2)	0.41(±0.04)	0.92(±0.08)	2.27(±0.09)
V70A	1.29(±0.03)	197(±13)	24.9(±0.8)	0.36(±0.02)	0.65(±0.03)	1.45(±0.04)
V72A	1.33(±0.04)	677(±96)	80.6(±6.5)	0.40(±0.04)	0.56(±0.05)	1.34(±0.06)
V72I	1.51(±0.03)	6800(±750)	606(±35)	0.35(±0.02)	4.21(±0.19)	9.04(±0.21)
M76A	1.30(±0.03)	1060(±110)	131(±8)	0.42(±0.03)	1.23(±0.09)	3.08(±0.11)
S77A	1.22(±0.02)	401(±27)	56.7(±1.9)	0.41(±0.03)	1.11(±0.07)	2.72(±0.09)
S77T	1.20(±0.03)	2560(±240)	378(±23)	0.52(±0.08)	0.22(±0.05)	0.68(±0.05)
L78A	1.26(±0.05)	399(±31)	53.5(±1.5)	0.39(±0.01)	57.0(±1.0)	133.5(±4.9)
D80A	1.18(±0.02)	908(±75)	138(±7)	0.43(±0.04)	0.49(±0.06)	1.25(±0.05)
C81A	1.17(±0.02)	2470(±170)	381(±16)	0.42(±0.06)	0.45(±0.07)	1.14(±0.06)
C81I	1.52(±0.05)	24,900(±4400)	2180(±230)	0.38(±0.03)	4.10(±0.30)	9.36(±0.27)
L82A	1.86(±0.06)	453(±47)	23.2(±0.8)	0.46(±0.02)	12.25(±0.29)	33.7(±1.0)
A85G	1.48(±0.04)	530(±46)	49.8(±1.9)	0.37(±0.01)	2.31(±0.10)	5.20(±0.20)
L86A	1.28(±0.04)	251(±18)	32.4(±0.8)	0.27(±0.01)	13.54(±0.25)	24.51(±0.46)
R89L	1.26(±0.04)	30,600(±5300)	4090(±490)	0.46(±0.13)	0.59(±0.26)	1.63(±0.24)
L91A	1.50(±0.07)	1790(±300)	163(±13)	0.38(±0.04)	4.26(±0.36)	9.74(±0.60)
P93A	1.34(±0.02)	3700(±310)	433(±22)	0.37(±0.04)	0.82(±0.08)	1.84(±0.06)
C95A	1.11(±0.02)	1149(±59)	194(±6)	0.43(±0.04)	0.63(±0.07)	1.62(±0.06)
C96A	1.30(±0.04)	3450(±380)	434(±25)	0.36(±0.03)	6.17(±0.45)	12.82(±0.43)
C96L	1.10(±0.01)	1579(±70)	274(±7)	0.49(±0.04)	0.60(±0.06)	1.74(±0.05)
C96M	1.25(±0.02)	2693(±190)	367(±15)	0.47(±0.04)	0.49(±0.05)	1.38(±0.05)
A97G	1.17(±0.02)	1151(±77)	178(±6)	0.37(±0.02)	1.88(±0.12)	4.28(±0.11)
V98A	0.92(±0.03)	308(±24)	70.8(±2.7)	0.37(±0.04)	2.55(±0.23)	5.80(±0.20)
R100A	1.19(±0.02)	1810(±110)	271(±9)	0.39(±0.03)	1.84(±0.12)	4.30(±0.10)
E104A	1.29(±0.03)	3240(±350)	412(±27)	0.37(±0.04)	0.69(±0.08)	1.56(±0.07)
K109A	1.31(±0.04)	3020(±360)	372(±25)	0.34(±0.04)	1.40(±0.14)	2.99(±0.12)
L112A	1.00(±0.04)	432(±31)	87.3(±2.4)	0.36(±0.01)	22.25(±0.69)	49.3(±1.0)
D117A	1.26(±0.02)	1600(±120)	211(±9)	0.41(±0.03)	1.05(±0.08)	2.59(±0.08)
A118G	1.05(±0.03)	671(±44)	125(±4)	0.46(±0.02)	6.79(±0.32)	18.63(±0.48)
A118L	1.15(±0.03)	1450(±120)	233(±8)	0.31(±0.01)	19.37(±0.59)	38.24(±0.65)
L121A	1.14(±0.04)	833(±74)	135(±5)	0.40(±0.02)	19.97(±0.77)	48.2(±1.4)
E124A	1.21(±0.03)	1910(±230)	276(±19)	0.40(±0.06)	0.73(±0.11)	1.79(±0.10)
E125A	1.21(±0.02)	1002(±67)	146(±6)	0.43(±0.03)	0.74(±0.06)	1.91(±0.06)
L126A	1.34(±0.06)	208(±18)	24.2(±0.8)	0.32(±0.01)	14.96(±0.36)	30.43(±0.74)
V128A	1.21(±0.05)	244(±19)	35.5(±1.0)	0.33(±0.01)	21.21(±0.46)	43.46(±0.93)
D129A	1.25(±0.03)	1680(±150)	227(±12)	0.46(±0.03)	1.37(±0.10)	3.82(±0.11)
Δ104-6	1.32(±0.02)	9660(±810)	1170(±70)	0.32(±0.10)	0.11(±0.04)	0.22(±0.03)
Δ101-8+AG	1.04(±0.03)	1470(±100)	280(±9)	0.42(±0.01)	15.94(±0.71)	40.55(±0.93)

See Materials and Methods for a description of the parameters.

α1 is most structured in its central section, but incompletely and non-optimally packed

The α-helix N-capping usually consists of a hydroxyl bearing and negatively charged side-chains in order to satisfy the backbone amide H-bond propensity and neutralize the induced dipole, respectively.[42] Mutation of the matching residues of Raf RBD, S77 and D80, to Ala yielded high Φ-values (0.70 and 1.05, respectively), suggesting that α1 secondary structure is significantly formed in this region at the TS. Strikingly, mutation to Ala of the first inner hydrophobic core residues of α1, L78, yielded low Φ-values (0.28). Previous work has indicated that α1 could be N-capped in trans by the backbone carbonyls of residues W114 and T116, as indicated by the H-bond networks they formed with L78 and H79.[32] T116 is connected through H-bonds to D117, an exposed residue that displayed a more significant amino acid selection in the sequence perturbation and relatively low Φ-value (0.38). W114M has a very minor effect on TS formation (Φ_F^{kin}~0.07; the Met was the least destabilized mutation tested; A. Vallée-Belisle & S.W. M., unpublished data). The low Φ-value of S77T (0.20) indicates that this residue is also involved in later events of folding, which could involve tighter packing with N115. Taken together, these results

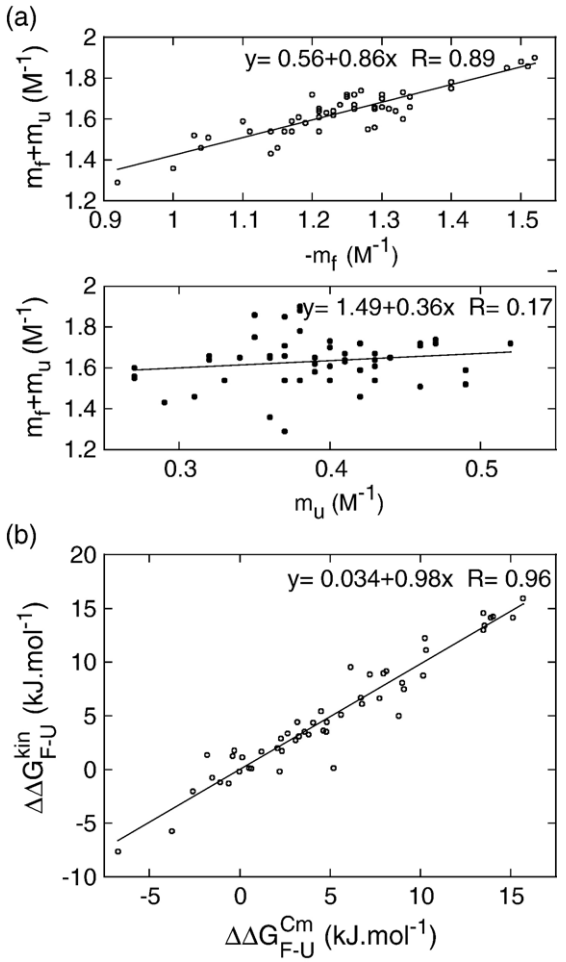

Figure 3. (a) Plot of m_f (○) or m_u (●) *versus* m_f+m_u. A very high correlation is seen with m_f, while a poor correlation is obtained with m_u (slope = 0.86; $R = 0.89$ *versus* slope = 0.36; $R = 0.17$, respectively). Note that the most deviant mutant, L82A was removed from both graphs. (b) Plot of the difference in the free energy of folding between mutants and the wt determined from equilibrium *versus* kinetic experiments (e.g. $\Delta\Delta G^{Cm}_{F-U}$ *versus* $\Delta\Delta G^{kin}_{F-U}$) and best linear fit of the data (slope = 0.98; $R = 0.96$).

correlate better than S77A and D80A with the low Φ-value of L78A, suggesting that N-capping *in trans* occurs after the TS and this could have an impact on the formation of the hydrophobic core. In agreement with this hypothesis, L78A displays a lower Φ-value than L82A and L86A (0.45 and 0.44, respectively), which are located in the center of α1. It is also noteworthy that the L78, in contrast with L82 and L86, establishes no direct side-chain contacts with the β-hairpin (Supplementary Data, Table S2). The native-likeness of the contact formed by the outer core residue A85G (0.57) at the TS is compatible with the nearby residues. In a mutagenesis study that delineated the RBD directly involved in *ras* binding, the R89L mutation was found to completely disrupt formation of the complex.[43] The non-obvious choice of Leu to mutate R89 motivated us to determine its folding/unfolding kinetics. Interestingly, the R89L strongly stabilizes the RBD, showing a one order of magnitude increase in k_f but little change in k_u, yielding a $\Phi \sim 1$. In contrast, the L91A, which probes the Gly/Asn-aliphatic C-capping motif, displayed a low Φ-value (0.25), indicating that the C terminus of α1 is denatured-like in the TS. Hence, the C-terminal end of α1 appears suboptimal for stability and folding rate. Overall, the relatively low Φ-values of L78, L82 and L86 correlate well with those of their principal inner core contacting residues in β1 (e.g. I58 and L62), β3 and β5, as discussed below, and suggest that α1 is not completely packed against the β-sheet in the TS.

C-terminal region is not well structured, except at inner hydrophobic core residues

The V98A (0.50), affecting the sole inner core residue of β3, yields the highest Φ-value in the C-terminal half of the RBD. Indeed, the mutations probing proximal residues of the outer core, such as C96A (−0.14) and R100A (0.13) have among the lowest Φ-values in our dataset. Other outer core residues that are closer to the C terminus, such as L112A, A118G and L121A, also yielded low Φ-values (0.26, 0.26 and 0.19, respectively). The A118G suggests that the α2 is not properly packed in the hydrophobic core at the TS. Several proteins adopting the ubiquitin-roll topology display a contact triad between the side-chains of residues matching I58 (β1), V72 (β2) and A118 (α2) of the RBD. Since the contacts between these three residues are spatially constrained and α2 is packed strictly through A118, we reasoned that introducing bulkier amino acids at any of the sites of the triad could be used to confirm its disruption at the TS. I58F and A118L displayed low Φ-values, while V72I mutation produced a negative Φ-value, suggesting that it induces formation of non-native interactions before the TS. Overall, these results confirm that α2 is not properly packed at the TS. The residues E104 to K106 that constitute part of the loop bridging β3 and β4 in c-Raf/Raf-1 are absent in a-Raf and b-Raf. The Δ104-6 mutant in which these residues are deleted in the c-Raf background increases k_f and decreases k_u ($3.5 \cdot k_f^{1.6\ M}$ wt and $0.2 \cdot k_u^{5.8\ M}$ wt, respectively), yielding a Φ-value of 0.42. In fact, Δ104-6 is so stable that it cannot be completely unfolded by urea. However, the impact of Δ104-6 mutation on TS stability is equivalent in Gdm-HCl ($\Phi_F^{kin} \approx 0.39$).[34] The strong impact of this mutation was unexpected based on other mutants (e.g. R100A, E104A and K109A) and the sequence perturbation data.[33] Finally, it suggests that this region, although apparently unimportant for TS formation in the wt, might induce significant variation in the folding mechanism of other Raf RBD family members. The E125 is the first residue of β5 and its mutation to Ala, in contrast with its putative salt bridge partner R59, produced a high Φ-value (0.67), suggesting that it stems from more local contacts. The inner core mutants of β5, L126A

Table 2. Thermodynamic parameters used for the various Φ-value calculations

	m (kJ mol^{-1} M^{-1})	Cm (M)	$\Delta\Delta G^{kin}_{F\text{-}U}$ (kJ mol^{-1})	$\Delta\Delta G^{Cm}_{F\text{-}U}$ (kJ mol^{-1})	$\Delta\Delta G^{5.8\ M}_{F\text{-}U}$ (kJ mol^{-1})
Wt	3.8(±0.2)	6.30(±0.04)			
N56M	3.8(±0.2)	6.59(±0.05)	−1.20(±0.11)	−1.1(±0.3)	−1.1(±2.0)
I58A	4.1(±0.1)	2.83(±0.04)	13.40(±0.07)	13.5(±1.2)	14.2(±1.4)
I58L	4.3(±0.2)	5.47(±0.03)	3.08(±0.08)	3.3(±0.3)	3.3(±1.7)
I58F	4.0(±0.2)	4.87(±0.03)	5.10(±0.11)	5.6(±0.5)	5.6(±1.5)
R59A	3.6(±0.2)	6.46(±0.06)	−1.24(±0.11)	−0.6(±0.3)	−0.4(±1.7)
V60A	3.6(±0.1)	3.70(±0.02)	8.76(±0.07)	10.1(±0.9)	9.4(±1.3)
L62A	3.6(±0.1)	2.43(±0.02)	14.13(±0.07)	15.1(±1.3)	14.2(±1.3)
P63A	3.6(±0.5)	4.73(±0.06)	9.54(±0.09)	6.1(±0.6)	5.8(±2.8)
N64A	3.9(±0.2)	5.15(±0.04)	5.41(±0.07)	4.5(±0.4)	4.4(±1.8)
H2	4.0(±0.4)	6.77(±0.06)	1.37(±0.09)	−1.8(±0.3)	−2.0(±2.7)
H2_F62L	3.6(±0.1)	5.12(±0.02)	3.64(±0.09)	4.6(±0.4)	4.4(±1.4)
Q66A	3.8(±0.2)	5.71(±0.04)	1.72(±0.08)	2.3(±0.3)	2.3(±1.7)
T68A	3.8(±0.2)	6.14(±0.05)	0.07(±0.08)	0.7(±0.2)	0.6(±1.8)
V69A	3.9(±0.2)	5.07(±0.03)	4.41(±0.08)	4.8(±0.4)	4.8(±1.6)
V70A	4.0(±0.2)	4.32(±0.04)	6.62(±0.07)	7.7(±0.7)	7.9(±1.6)
V72A	4.2(±0.2)	5.08(±0.03)	3.51(±0.10)	4.8(±0.4)	4.9(±1.7)
V72I	3.9(±0.1)	5.32(±0.03)	3.24(±0.08)	3.8(±0.4)	3.8(±1.5)
M76A	3.1(±0.1)	5.26(±0.02)	4.36(±0.08)	4.1(±0.4)	3.6(±1.4)
S77A	4.0(±0.1)	4.57(±0.01)	6.14(±0.07)	6.8(±0.6)	6.8(±1.3)
S77T	4.0(±0.2)	6.97(±0.03)	−1.99(±0.10)	−2.6(±0.3)	−2.7(±1.8)
L78A	4.0(±0.1)	2.28(±0.01)	15.93(±0.07)	15.7(±1.3)	16.1(±1.3)
D80A	3.5(±0.1)	5.76(±0.02)	1.99(±0.08)	2.1(±0.2)	2.1(±1.4)
C81A	3.6(±0.2)	6.70(±0.04)	−0.75(±0.08)	−1.5(±0.2)	−1.3(±1.7)
C81I	5.0(±0.2)	4.98(±0.05)	0.15(±0.12)	5.2(±0.5)	6.1(±1.6)
L82A	4.8(±0.3)	2.84(±0.03)	14.60(±0.07)	13.5(±1.2)	16.2(±1.5)
A85G	4.4(±0.2)	4.00(±0.02)	8.07(±0.07)	9.0(±0.8)	9.9(±1.5)
L86A	4.1(±0.1)	2.85(±0.02)	12.99(±0.06)	13.5(±1.1)	14.0(±1.4)
R89L	4.1(±0.2)	7.27(±0.03)	−5.73(±0.19)	−3.8(±0.4)	−4.0(±1.9)
L91A	3.8(±0.4)	4.59(±0.14)	6.67(±0.11)	6.6(±0.9)	6.5(±2.3)
P93A	4.0(±0.1)	6.17(±0.05)	0.13(±0.08)	0.5(±0.2)	0.4(±1.5)
C95A	4.1(±0.2)	6.38(±0.03)	1.80(±0.07)	−0.3(±0.2)	−0.5(±2.0)
C96A	3.6(±0.1)	4.05(±0.01)	4.98(±0.08)	8.8(±0.8)	8.3(±1.3)
C96L	3.6(±0.1)	6.40(±0.02)	1.24(±0.06)	−0.4(±0.2)	−0.3(±1.5)
C96M	4.4(±0.2)	5.73(±0.03)	−0.17(±0.07)	2.2(±0.3)	2.2(±1.9)
A97G	4.0(±0.2)	5.49(±0.03)	4.43(±0.06)	3.2(±0.3)	3.2(±1.7)
V98A	3.5(±0.2)	3.97(±0.03)	7.46(±0.07)	9.1(±0.8)	8.3(±1.5)
R100A	3.8(±0.2)	5.62(±0.03)	3.40(±0.06)	2.7(±0.3)	2.6(±1.6)
E104A	3.9(±0.3)	6.31(±0.05)	−0.16(±0.09)	0.0(±0.1)	−0.1(±2.1)
K109A	3.9(±0.2)	6.00(±0.04)	1.71(±0.09)	1.2(±0.2)	1.2(±1.9)
L112A	3.8(±0.1)	3.67(±0.02)	12.25(±0.06)	10.3(±0.9)	10.0(±1.3)
D117A	3.6(±0.1)	5.51(±0.02)	2.76(±0.07)	3.1(±0.3)	3.0(±1.4)
A118G	3.8(±0.1)	4.27(±0.02)	8.94(±0.06)	7.9(±0.3)	7.8(±1.4)
A118L	3.3(±0.1)	4.23(±0.02)	9.19(±0.06)	8.1(±0.7)	7.2(±1.3)
L121A	4.0(±0.1)	3.66(±0.02)	11.11(±0.07)	10.3(±0.9)	10.4(±1.4)
E124A	4.1(±0.4)	6.27(±0.06)	1.17(±0.10)	0.1(±0.3)	0.0(±3.0)
E125A	3.8(±0.1)	5.73(±0.02)	2.92(±0.07)	2.3(±0.2)	2.2(±1.4)
L126A	4.1(±0.2)	2.71(±0.03)	14.24(±0.06)	14.0(±1.2)	14.5(±1.4)
V128A	4.0(±0.1)	2.74(±0.02)	14.18(±0.06)	13.9(±1.2)	14.2(±1.3)
D129A	3.9(±0.1)	5.38(±0.03)	3.53(±0.08)	3.6(±0.3)	3.6(±1.5)
Δ104-6	3.5(±0.2)	8.03(±0.04)	−7.63(±0.14)	−6.7(±0.6)	−5.8(±1.9)
Δ101-8+AG	3.6(±0.1)	4.46(±0.02)	8.88(±0.06)	7.2(±0.6)	6.8(±1.5)

See Materials and Methods for a description of the parameters.

and V128A, displayed close to average Φ-values (0.45 and 0.39, respectively). The low Φ-value of D129A (0.24) suggests that the C terminus of β5 is almost completely denatured-like.

In brief, our data supports a model in which the Raf RBD folds through formation of a folding-nucleus consolidated around the N-terminal β-hairpin. Despite the adoption of less native-like conformations, significant contacts for stabilization of the TS are also formed between α1, β3 and β5, which establish numerous contacts in the inner hydrophobic core of the native structure.

Raf RBD TS structure is diffuse according to $\Delta\Delta G_{U\text{-}‡}$rel values

Although the Φ-value analysis described above supports the hypothesis of a polarized TS, there are very few residues throughout the protein that have near-zero Φ-values, suggesting that some degree of native structure is formed in all regions of the RBD. In view of these results, we sought to compare the TS characteristics delineated from Φ-value analysis with those deduced from a direct measure of TS destabilization. To do so, $\Delta\Delta G^{1.6\ M}_{U\text{-}‡}$ was normalized against

Table 3. Φ-Values, $\Delta\Delta G_{U-\ddagger}$, β_t and structural information

	$\Phi_F^{kin\ a}$	$\Phi_F^{Cm\ a}$	$1-\Phi_U^{a,b}$	$\Delta\Delta G_{U-\ddagger}^{1.6\ M}rel^a$ (kJ mol^{-1})	Structure[c]	Solvent access[d]
Wt						
N56M*					und	69
I58A	0.54(±0.01)	0.55(±0.05)	0.55(±0.06)	1.00(±0.03)	β1/IC	5
I58L	0.48(±0.05)	0.42(±0.09)	0.55(±0.30)	0.20(±0.02)	β1	5
I58F	0.17(±0.05)	0.02(±0.07)	0.19(±0.05)	0.12(±0.03)	β1	5
R59A*					β1	87
V60A	0.82(±0.02)	0.68(±0.06)	0.80(±0.16)	0.99(±0.03)	β1/IC	3
L62A	0.44(±0.01)	0.39(±0.04)	0.39(±0.04)	0.86(±0.02)	t1/OC	4
P63A	0.32(±0.02)	0.56(±0.09)	−0.22(±0.11)	0.43(±0.03)	t1	45
N64A	0.98(±0.03)	1.19(±0.13)	1.02	0.74(±0.02)	t1	93
H2*					t1	
H2_F62L	0.10(±0.05)	−0.01(±0.05)	0.07(±0.02)	0.05(±0.03)	t1	
Q66A*					und/OC	65
T68A*					und/OC	46
V69A	0.69(±0.04)	0.62(±0.08)	0.71(±0.3)	0.42(±0.02)	β2	83
V70A	0.96(±0.02)	0.78(±0.07)	0.93(±0.5)	0.88(±0.02)	β2/OC	22
V72A	0.98(±0.07)	0.63(±0.10)	0.97	0.48(±0.03)	und/OC	3
V72I	−0.48(±0.05)	−0.71(±0.11)	−0.36(±0.14)	−0.22(±0.02)	und/OC	3
M76A	0.51(±0.04)	0.46(±0.09)	0.42(±0.17)	0.31(±0.02)	t2	45
S77A	0.70(±0.02)	0.63(±0.06)	0.73(±0.18)	0.60(±0.02)	Ch	30
S77T	0.20(±0.09)	0.12(±0.11)	0.20(±0.15)	−0.06(±0.02)	Ch	30
L78A	0.28(±0.01)	0.27(±0.03)	0.28(±0.02)	0.62(±0.02)	α1/IC	0
D80A	1.05(±0.09)	1.06(±0.17)	1.10	0.29(±0.02)	α1	89
C81A*					α1/OC	3
C81I*					α1/OC	3
L82A	0.45(±0.01)	0.30(±0.03)	0.52(±0.05)	0.90(±0.02)	α1/IC	5
A85G	0.57(±0.02)	0.40(±0.05)	0.63(±0.10)	0.64(±0.02)	α1/OC	3
L86A	0.44(±0.01)	0.40(±0.04)	0.42(±0.04)	0.79(±0.02)	α1/IC	0
R89L	1.10(±0.07)	1.72(±0.21)	1.07	−0.87(±0.05)	α1/OC	83
L91A	0.25(±0.03)	0.09(±0.07)	0.21(±0.07)	0.23(±0.03)	Ch/OC	65
P93A*					t3	2
C95A*					t3	70
C96A	−0.14(±0.04)	−0.11(±0.04)	0.26(±0.04)	−0.10(±0.02)	T3/OC	19
C96L*					T3/OC	19
C96M*					T3/OC	19
A97G	0.33(±0.03)	0.53(±0.09)	0.01(±0.01)	0.20(±0.02)	β3	1
V98A	0.50(±0.02)	0.54(±0.05)	0.53(±0.11)	0.52(±0.02)	β3/IC	2
R100A	0.13(±0.04)	0.21(±0.08)	−0.19(±0.12)	0.06(±0.02)	β3/OC	25
E104A*					t4	114
K109A*					und	46
L112A	0.26(±0.01)	0.40(±0.04)	0.08(±0.01)	0.45(±0.02)	und/OC	24
D117A	0.38(±0.05)	0.28(±0.08)	0.43(±0.22)	0.14(±0.02)	und	89
A118G	0.26(±0.01)	0.38(±0.04)	0.18(±0.03)	0.32(±0.02)	α2/OC	1
A118L	0.09(±0.01)	0.14(±0.03)	−0.24(±0.05)	0.11(±0.02)	α2/OC	1
L121A	0.19(±0.01)	0.24(±0.03)	0.13(±0.02)	0.30(±0.02)	t7/OC	15
E124A*					t7	44
E125A	0.67(±0.05)	0.90(±0.14)	0.61(±0.46)	0.27(±0.02)	β5	49
L126A	0.45(±0.01)	0.42(±0.04)	0.43(±0.04)	0.89(±0.02)	β5/IC	3
V128A	0.39(±0.01)	0.40(±0.08)	0.35(±0.03)	0.76(±0.02)	β5/IC	0
D129A	0.24(±0.05)	0.20(±0.09)	0.33(±0.15)	0.12(±0.02)	β5	34
Δ104-6	0.42(±0.02)	0.53(±0.06)	0.33(±0.13)	−0.44(±0.03)	t4	
Δ101-8+AG	0.04(±0.01)	0.15(±0.04)	−0.25(±0.05)	0.05(±0.02)	t4	

[a] See Materials and Methods for details of the calculation.
[b] The experimental errors for some mutants were extremely high because of the fractional value of $\Delta\Delta G_{F-\ddagger}$. In these cases, the experimental errors are omitted.
[c] S, t and h represent β-strand, α-helix and β-turns. β-strands are: β1 (57–61), β2 (69–71), β3 (97–101), β4 (110–111) and β5 (125–130); α_helices are: α1 (78–89), α2 (118–120); β-turns are: t1 (62–65), t2 (73–76), t3 (93–96), t4 (102–105), t5 (105–108), t6 (113–116) and t7 (121–124). Ch stands for residues involved in the capping of the major helix; und stands for undefined structure (e.g. loops). IC and OC highlight inner and outer core residues.
[d] Information extracted from the DSSP file of 1RFA.
* Mutations with kinetic estimates of $\Delta\Delta G_{F-U} < |2|$ kJ mol^{-1}.

$\Delta\Delta G_{U-\ddagger}^{1.6\ M}$ max, which is the $\Delta\Delta G_{U-\ddagger}^{1.6\ M}$ of the mutant with the most destabilized TS (e.g. I58A), thus yielding the $\Delta\Delta G_{U-\ddagger}^{1.6\ M}$rel, which ranges between 0–1 for destabilizing non-disruptive mutations (Materials and Methods, equations (7) and (8)). The $\Delta\Delta G_{U-\ddagger}^{1.6\ M}$rel of I58A and V60A, followed closely in decreasing order by L62A, V70A, L82A, L86A, L126A and V128A indicate that these residues contribute the most strongly to TS stabilization (Table 3 and Figure 4). Consequently, all regions of the inner core seem to participate roughly equally to TS stabilization including residues located in α1 and β5, which are not in such a highly native-like environment. In contrast, the role of core residues in the most native-like region, the

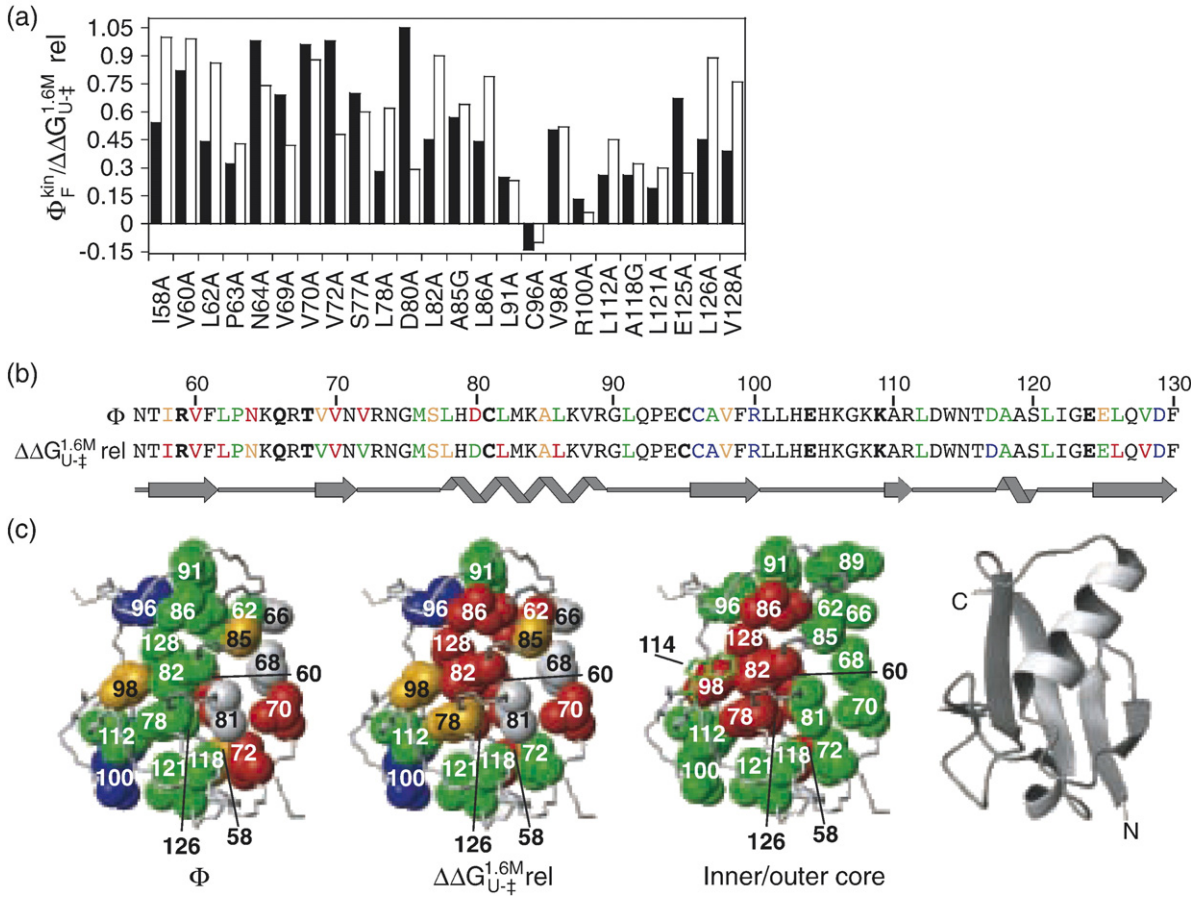

Figure 4. (a) Comparison between Φ_F^{kin} (■) and $\Delta\Delta G_{U-\ddagger}^{1.6\,M}$rel (□) for the residues most important for TS stabilization. (b) The distribution of Φ and $\Delta\Delta G_{U-\ddagger}$rel obtained from Ala/Gly mutations on the primary and secondary structure of Raf RBD are indicated using the following scale: (1) 0–0.25 (blue); (2) 0.25–0.5 (green); (3) 0.5–0.75 (orange); and (4) 0.75–1 (red). Mutations producing $\Delta\Delta G_{F-U} < |2|$ kJ mol^{-1} are highlighted by black and bold lettering (Table 3). (c) Distribution on the tertiary structure of Raf RBD of residues forming the hydrophobic core according to Φ, $\Delta\Delta G_{U-\ddagger}^{1.6\,M}$rel and the bi-level organization of the hydrophobic core (e.g. the inner/outer core; reproduced from Figure 1). The color scale used in the two leftmost panels are as described for (b), except that mutations with $\Delta\Delta G_{F-U} < |2|$ kJ mol^{-1} are colored in grey. A ribbon representation of Raf RBD is shown for reference in the right panel.

N-terminal β-hairpin, does not change as drastically, although the observation of significant and opposite variations at L62 and V72 is noteworthy. Therefore, in contrast to the interpretation implied by Φ-values, $\Delta\Delta G_{U-\ddagger}^{1.6\,M}$rel reveals a more diffuse TS structure (Figure 4(c)). To ensure that the non-thermodynamically normalized parameter $\Delta\Delta G_{U-\ddagger}^{1.6\,M}$rel does not introduce a bias for the most drastic mutation (e.g. I→A *versus* V→A), the effect of normalizing according to side-chain volume variation was tested. We found that this manipulation of the data did not significantly alter our interpretation of the Raf RBD TS or those of other proteins discussed below (Discussion, and data not shown).

Contacts formed at the TS in structural regions decrease as their distance from the folding nucleus increases

For a set of mutants introduced into a given protein, the lnk_f or lnk_u is correlated with $\Delta\Delta G_{F-U}$, yielding what is called a Leffler plot (aka Brønstead). In this case, the slope (named β_F or β_U for lnk_f or lnk_u, respectively) of the best linear fit corresponds to an average Φ_F. The Leffler plot obtained by plotting lnkf against $\Delta\Delta G_{F-U}/RT$ for all Raf RBD mutants displays a good correlation (Figure 5(a); $\beta_F = 0.47$ and $R = 0.82$). This result is in agreement with the low number of residues displaying extreme Φ-values (e.g. 0 or 1; see Table 3). The scattering of this Leffler plot appears to be intermediate between those of protein-L and CI2, prototypes of highly polarized and diffuse TS, respectively.[4,15] Hence, correlations within distinct secondary structure elements and subsets of the tertiary structure, such as core or surface positions, were tested to evaluate the level of native structure that they consolidated at the TS. Indeed, we observed stronger correlations by breaking up the mutants of Raf RBD into three groups: the N-terminal β-hairpin, the segment extending from the C terminus of α1 to the N terminus of β5 (e.g. L91-L121) and a discontinuous set composed of α1 and β5, plus V98A (Figure 5(b); $\beta_F = 1.03$ and $R = 0.89$, $\beta_F = 0.25$ and $R = 0.84$, and $\beta_F = 0.50$ and $R = 0.88$, respectively, for these three regions). However, the best linear fit for the β-

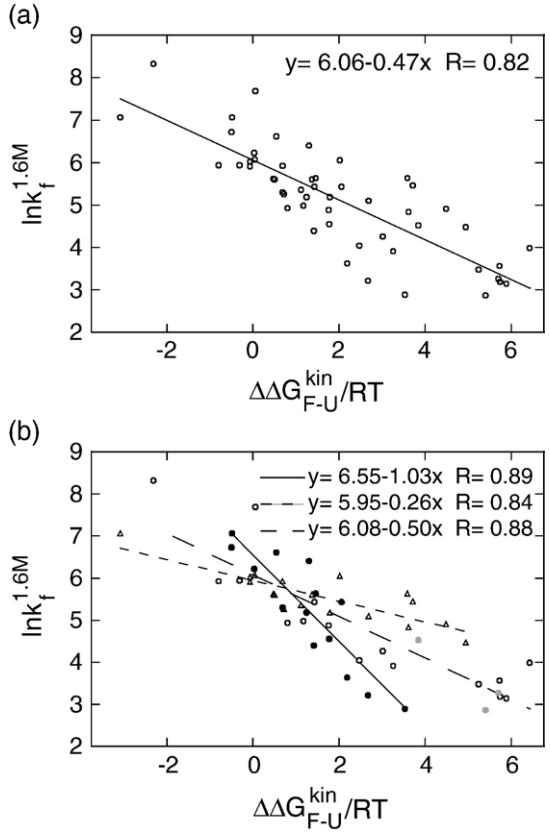

Figure 5. (a) Leffler graph obtained by plotting $\ln k_f^{1.6\,M}$ against $\Delta\Delta G^{kin}_{F\text{-}U}/RT$ for all mutants tested (○). (b) Based on the same data, separate Leffler correlations were established for mutations located in the first β-hairpin (N56-V72) (●), a segment located between α1 and β5 (e.g. L91-E124, but V98) (△), and the segments corresponding roughly to α1 (e.g. M76-R89), β5 plus V98 (○). Outlier mutants in the β-hairpin (I58A, L62A and P63) are indicated (●), but neglected for the best linear fit displayed. If the outlying mutants are nevertheless included in the β-hairpin subgroup, the correlation obtained is less strong and the slope is significantly altered ($y = 6.20 + 0.65x$; $R = 0.85$).

hairpin excluded three major outliers, I58A, L62A and P63A, which are also the most destabilizing mutants ($\Delta\Delta G^{kin}_{F\text{-}U}/RT$ for these mutants are greater than 4; the best linear fit obtained by inclusion of these mutants is: $y = 6.20 + 0.65x$; $R = 0.85$). The observations noted above with regard to the Leffler plot hold true if strictly Ala/Gly mutants are considered or for correlation of folding and unfolding rate with equilibrium data (data not shown; Supplementary Data, Figure S3). Overall, these results support the hypothesis that the residues located in the β-hairpin are in a more native-like environment, while all regions of the protein seem to participate to stabilization. However, the Φ-values diminished progressively as the residues mutated are further away from β2. This phenomenon is particularly evident for the most distal residues located between the C terminus of α1 and the N terminus of β5. These observations suggest a progressive extension of consolidated structure in the TS ensemble from the hydrophobic core nucleus formed in the amino-terminal β-hairpin. Nevertheless, it is not clear that the linearity of the dependence of $\ln k_f^{1.6\,M}$ *versus* $\Delta\Delta G^{kin}_{F\text{-}U}/RT$ is true for mutants of the N-terminal β-hairpin with $\Delta\Delta G^{kin}_{F\text{-}U}/RT > 4$. Specific curvature in Leffler plots of barnase were suggested to indicate the presence of parallel pathways in which the helix located at the N terminus was either completely structured or partially disrupted.[44] It was reported that similar patterns in Leffler plots are rare, as it was not observed in a large set of two-state-folding proteins.[24] Therefore, more mutants of Raf RBD should be obtained to verify this discrepancy in the Leffler correlation.

Discussion

Evidence for the presence of residual structure in the denatured state of Raf RBD

Several small proteins, including ribosomal protein S6, CI2, Sso7D, SH3, protein-L and G, were previously shown to display apparent variation in the denatured state structure upon mutation based on correlation of m_f *versus* $m_f + m_u$.[24,40] The study by Sanchez & Kiefhaber suggested that this could be a common phenomenon as 7 of 21 proteins that showed Hammond behavior upon mutation or chemical/physical perturbations according to β_t, in fact, underwent substantial structural changes in their denatured state. We have demonstrated that Raf RBD mutants display such changes in their denatured state, similarly to what is observed in ubiquitin (Figure 3(a) and by analyzing data by Went & Jackson, respectively). The most drastic changes occur in both cases upon mutation of hydrophobic core residues, arguing in favor of a denatured state structure stabilized by the interactions of hydrophobic side-chains. A corollary of the correlation of m_f *versus* $m_f + m_u$ is that β_t is not a reliable measure of TS shift. On the other hand, the data suggests that the native state is insensitive to mutation (Figure 3(a)). Therefore, Raf RBD mutants that showed a decrease in m_u could indicate true Hammond behavior and reveal properties of the folding pathway, as discussed later.

From a theoretical viewpoint, structure in the denatured state, particularly its sensitivity to mutation, causes a problem for the interpretation of Φ-values. A central assumption of protein engineering is that the denatured state does not vary in energy upon mutation.[2] The correlation between m_f and $m_f + m_u$ solely indicates variation in exposure to solvent of the denatured state, not necessarily a significant variation in energy. In this sense, contacts formed in the denatured state could be so weak and transiently formed that it is reasonable to state that the majority of $\Delta\Delta G_{U\text{-}\ddagger}$ occurs because of TS destabilization.[24] This assumption is in agreement with the correlation observed among the Φ-value

estimates determined from folding and unfolding data on Raf RBD mutants (Table 3 and Supplementary Data, Figure S2). Direct experimental evidence suggests that native and non-native contacts are involved in stabilizing fluctuating substructures in the denatured state of several proteins, including barnase, CI2 and engrailed homeodomain for which Φ-values analysis were reported.[45–47]

Raf RBD TS characteristics: Φ *versus* $\Delta\Delta G_{U\text{-}\ddagger}$rel values

The Φ-value analysis suggests that Raf RBD forms a structurally polarized TS that is characterized by a nucleus located in the amino-terminal β-hairpin. In fact, this segment contains four of the five hydrophobic core residues with the highest Φ-values (I58, V60, V70 and V72) (Table 3). The nucleus extends, albeit with less native-like characteristics, to every residues in the inner core including the α1 (specifically L82A and L86A), β3 (V98A) and β5 (L126A and V128A), all of these showing close to average Φ-values (Figures 4 and 5(a)). The high Φ-values obtained upon mutation of N-capping residues (S77A and D80A) and by introducing a Gly mutation in α1 (A85G) confirmed that this element of secondary structure is significantly structured between D80 and L86. The ratio of native contacts formed at the TS by any residue is roughly correlated with its proximity and direct contacts to the N-terminal β-hairpin (Figure 4 and Supplementary Data, Table S2). This extension of the nucleus from a polarized structure is coherent with the Leffler plot obtained (Figure 5 (b)), which demonstrates that regions most distal to the β-hairpin have lower β_F. Next, we reasoned that it would be interesting to compare Φ to $\Delta\Delta G^{1.6\ M}_{U\text{-}\ddagger}$rel values, and see how it could challenge our interpretation of TS structure. We found using this direct measure of the stabilizing contributions of residues to the TS (e.g. k_f variation) that it is more diffuse than expected from the Φ-values analysis (Table 3 and Figure 4). This "energy perspective" on the TS ensemble properties suggests that the inner hydrophobic core plays a central role, although N-terminally located outer core residues, L62 and V70, are also significantly implicated. Finally, Φ-values and $\Delta\Delta G^{1.6\ M}_{U\text{-}\ddagger}$rel suggest that residues probed between L91–L121, excluding V98, play a minor role in TS stabilization. According to our model, the side-chains in the β-hairpin should be more rigid than those in other parts of the structure, particularly in the most distal segment, L91–L121. The context in the TS of the side-chains of hydrophobic core residues located outside of the β-hairpin might approximate those found in molten globule states. Therefore, most of the topology and some secondary structures are already determined at the TS of Raf RBD in agreement with the near native placement of the TS ($\beta_t^{avge} = 0.76 \pm 0.04$), while rearrangements in the hydrophobic core are still required to reach the native state.

Φ-Values analysis allows for predicting the TS ensemble structural properties by measuring the level of native-like contacts formed in each region of a protein. This capacity stems from the introduction of $\Delta\Delta G_{F\text{-}U}$ as a denominator in Φ calculation and the assumption that only native contacts are significantly formed at the TS. On that basis, it is assumed that the TS is a less compact version of the native structure in which a fraction of the native contacts are missing and non-native contacts are not significantly formed. This reasoning highlights the fact that Φ is not a pure measure of TS structure, but a relative measure of its native-likeness. Consequently, the residues impacting equally on $\Delta\Delta G_{U\text{-}\ddagger}$ (e.g. k_f), but differently on $\Delta\Delta G_{F\text{-}U}$ display different Φ-values, converging to one for mutants in which $\Delta\Delta G_{U\text{-}\ddagger}$ approximates $\Delta\Delta G_{F\text{-}U}$. Therefore, the impact of mutations that induce the greatest destabilizing effect on TS, but that have not completely formed their native contacts at this stage, are underestimated by Φ-values. This could lead to a physical paradox, in which residues with higher Φ-values are considered more important for TS stabilization than residues affecting more strongly the folding rate and hence, TS stability. The comparison of Φ and $\Delta\Delta G^{1.6\ M}_{U\text{-}\ddagger}$rel for several mutations in Raf RBD illustrates this phenomenon (e.g. N64A, V72A, S77A and inner core residues of β1 and β2 *versus* α1; see Table 3). Proteins displaying moderately polarized TS in combination with a rather homogenous contribution of hydrophobic core residues to native state stability should behave similarly to the RBD. Nevertheless, three characteristics of TS description using the normalized parameter $\Delta\Delta G_{U\text{-}\ddagger}$rel should be reminded to interpret results correctly. Foremost, this parameter does not provide direct structural information, but rather the relative importance of each mutated side-chains in stabilizing the TS. On the other hand, the estimates of $\Delta\Delta G_{U\text{-}\ddagger}$rel have lower incertitude than Φ-values. Second, it is dependant on normalization by the most destabilizing single point mutations, which could vary for the same protein depending on the set of mutations tested. Third, $\Delta\Delta G_{U\text{-}\ddagger}$rel does not implicitly correct for the severity of mutations (e.g. the level of side-chain volume variation). The normalization of the $\Delta\Delta G^{1.6\ M}_{U\text{-}\ddagger}$rel parameter by the variation in side-chain volume upon mutation did not dramatically change the TS energetic properties in this case, as discussed in detail below.

A sketch of Raf RBD folding pathway

In near physiological conditions, the denatured state ensemble displays transitory hydrophobic contacts and secondary structures (Figures 3(a) and 6).[27] β-Hairpins with equivalent importance in the stabilization of the TS of protein-G and ubiquitin were shown to fold independently at rates two to three orders of magnitude above the full-length proteins,[48–50] suggesting that these elements constitute a first significant step on the folding pathway. Accordingly, the β-hairpin was found to adopt the most native conformation in the TS ensemble of Raf

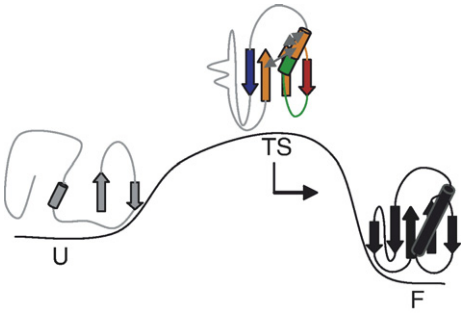

Figure 6. Schematic model of the folding pathway of Raf RBD derived from the protein engineering experiments. The role of various structural regions in stabilization of the TS is colored using the following scale in increasing order of importance: grey<blue<green <orange<red. Hammond behavior on a broad energy barrier induced upon mutation of certain residues was suggested by significant variation of m_u (the arrow indicate direction of TS shift induced by mutations). The co-localization in the native structure of the mutated residues displaying this effect and other results suggest that α1 is in the process of consolidating its packing over the β-hairpin and β5 at the TS (indicated by the grey double-head arrows).

RBD. The α1 adopts a partially native-like structure particularly strong between D80 and L86, but the relatively low Φ-value of α1 and β5 inner core residues as well as L62 and V98 in β1 and β3, respectively, indicates that the α-helix is not properly packed against the β-sheet. Interestingly, variations in m_u indicating a maximum reduction in solvent exposure of the TS of 33% in comparison with the native state (e.g. considering that $m_u^{L62,\ L86}=0.27$, $m_u^{wt}=0.41$), were noted for Ala mutations of several of these residues, which are structurally juxtaposed: L62, P63 (β-turn1), L86 (α1), L126 and V128 (β5) (Table 1 and Supplementary Data, Figure S1). The behavior of these mutations follows the Hammond postulate, which is frequently observed on the broad energy barrier typical of protein folding and arises only at residues that form contacts close to the top of the barrier.[38,51] Therefore, we propose that the sensitivity of the TS to mutation indicates that concomitant to nucleus formation centered on the β-hairpin, native hydrophobic contacts consolidate preferentially between the C termini of α1 with β1 and β5 (Figure 6), which would thereafter be extended cooperatively to the rest of the structure on the downhill side of the reaction. Accordingly, the mutants R89L, which is adjacent in the tertiary structure to L86 and the β-hairpin through direct contacts with L62, yield high Φ-value (≈1), indicating that better hydrophobic packing at this end of α1 improved folding efficiency. The mechanistic implication of our model is that among the extended folding nucleus formed by inner core residues, the C terminus of α1 and the residues of the β-sheet with which it interacts in the native state play a distinctive "switch-like" role at the TS that determine the completion of the folding reaction. This is in agreement with the detailed formalism by Hedberg *et al.* that was used to explain similar movement of the TS induced by discrete mutations introduced in the protein L23.[51] In the latter study, the residues displaying such behavior were used to define the "critical contact layer", which adds decisively to the interactions formed by a polarized nucleus en route to the native state. In contrast, CI2, which adopts a diffuse TS, display comparable Hammond behavior upon the introduction of mutations in all regions of the structure.[38] A TS shift towards a destabilized native state could also indicate the presence of successive energy barriers,[52] but there is no evidence arguing for the presence of intermediate states on the Raf RBD folding pathway.

Recently, synergistic results obtained in experiments and simulations have shed light on the generality of partially structured denatured states and the presence of intermediate states,[27] emphasizing the importance of *in silico* studies in improving the description of the folding process. In this perspective, the sketch of the folding pathway presented above represents a testable hypothesis for theorists. The Raf RBD is a particularly interesting study case, because it adopts the highly populated β-grasp ubiquitin-like topology, allowing for comparison with several other structural analogues, among which ubiquitin is the most relevant.

Similarities in the folding mechanisms of Raf RBD and ubiquitin

Raf RBD and ubiquitin belong to the ubiquitin-like superfamily of the β-grasp ubiquitin-like topology according to the SCOP database, but they have insignificant sequence identity (<12%). Their refolding traces were previously shown to display similar heterogeneities and sensitivities to temperature and stabilizing salts, despite significant differences in native state stability and TS placement on the reaction coordinates.[35] According to Φ-values, their TS are similarly consolidated around their N-terminal β-hairpin, although the TS of ubiquitin is less native-like and more polarized[26] (Figures 7(a) and 8(a)). A notable difference is that β5 and the residue corresponding to V98 (e.g. L43) are not involved at all in stabilizing the ubiquitin TS. The discrepancy in the level of compaction of the TS of these structural analogues are reminiscent of that observed between two members of the immunoglobulin-like Greek key fold, although in this case the reduction in compaction appears to be more evenly distributed across the protein structure in the most diffuse TS.[12,13,28] Nevertheless, we found a low but significant correlation between Raf RBD and ubiquitin Φ-values (slope=0.51; $R=0.63$; Supplementary Data, Figure S4; see also Materials and Methods), which can be improved by ignoring the most deviant pair and correcting for differences in packing (R values of 0.75 and 0.84, respectively). This is comparable to the correlations published on seven pairs of structurally similar proteins studied previously[31]. From the perspective of $\Delta\Delta G_{U\text{-}\ddagger}$rel,

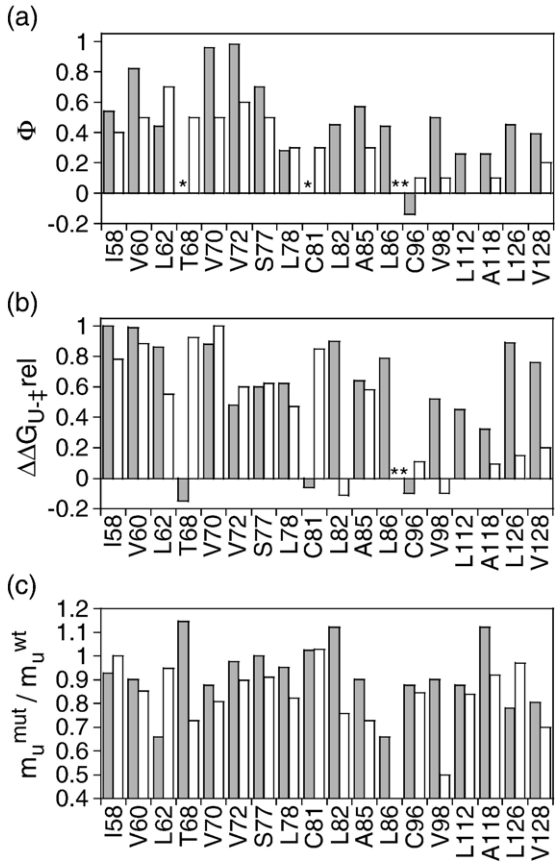

Figure 7. Comparison of the residues involved in the TS stabilization and change of m_u in Raf RBD (■) *versus* ubiquitin (□). (a) Φ-Value. (b) $\Delta\Delta G_{U\text{-}‡}$rel. (c) The m_u variations expressed using the ratio m_u^{mut}/m_u^{wt}. See Materials and Methods for details on the Φ-values and $\Delta\Delta G_{U\text{-}‡}$rel utilized for comparing Raf RBD and ubiquitin. The residues numbering of the Raf RBD is used throughout all panels. The residues of both proteins are matched using the secondary structure alignment[34]. Two of the inner core residues of the α-helix could not be straightforwardly aligned. Therefore, residues V26 and I30 of ubiquitin should be compared with their equivalent in the tertiary structure of Raf RBD, L82 and L86, respectively. Mutations marked with one star (*) indicate $\Delta\Delta G_{F\text{-}U} < |2|$ kJ mol^{-1} (Table 3). The two stars (**) highlight an ubiquitin mutation that has not been reported. The remaining columns without bars indicate untested positions. Only Ala/Gly mutants were considered.

ubiquitin TS extends to α1 as described for Raf RBD, but still excludes β5 and the contribution of residues in β2 is more substantial (Figures 7(b) and 8(a)). These differences between Raf RBD and ubiquitin TS ensemble could stem from slight native structure dissimilarities due to differences in the packing of α1 over the β-sheet, which introduces local variations in packing of the protein.[33,34] In addition, it is noteworthy that the $\Delta\Delta G_{U\text{-}‡}$ patterns observed for Raf RBD and ubiquitin are virtually identical to $\Delta\Delta G_{U\text{-}‡}$rel (Supplementary Data, Figure S4), indicating that matching residues in the two proteins, excluding those in β5, induce comparable destabilization of the TS. We also provide the correction for volume of $\Delta\Delta G_{U\text{-}‡}$rel for Raf RBD and ubiquitin to demonstrate that it does not affect drastically the

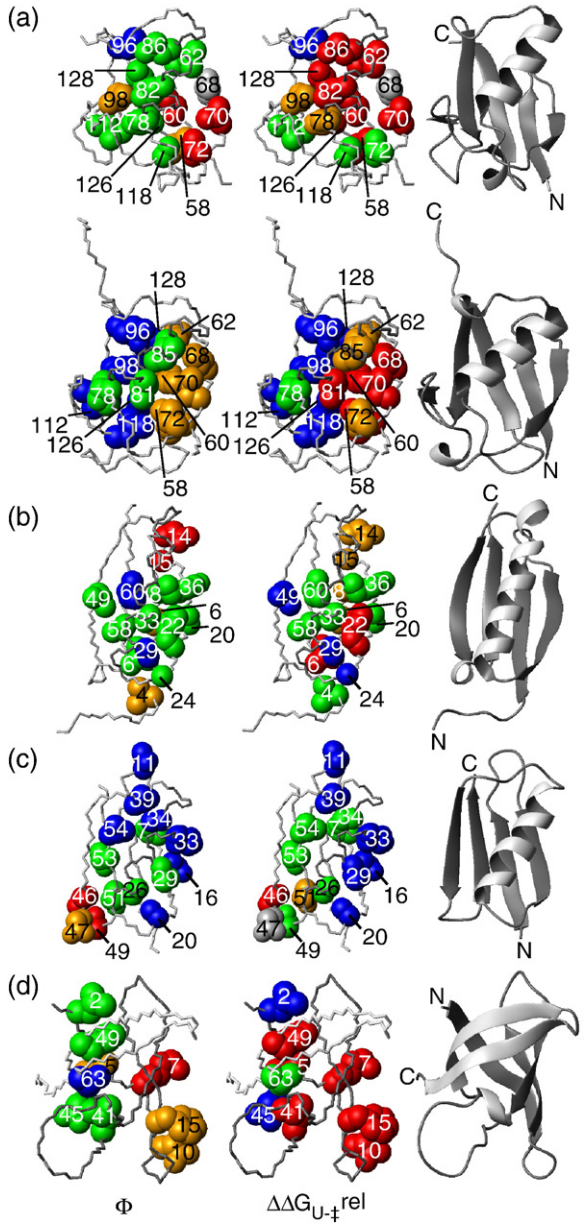

Figure 8. Role of residues in the stabilization of TS structure using Φ-value and $\Delta\Delta G_{U\text{-}‡}$ rel (left and middle panel, respectively): comparison of Raf RBD, ubiquitin and other structural analogues and cold shock protein case. (a) Raf RBD *versus* ubiquitin (top and bottom panel, respectively). The matching hydrophobic core residues for which data were available for Raf RBD and ubiquitin are shown here. The residue numbering of the Raf RBD is used for both proteins. In comparing the two proteins, consider the misalignment of inner core residues in α1 described (Figure 7 and Discussion). (b) Protein-L. (c) Protein-G. (d) Cold-shock protein. Details on the various Φ and $\Delta\Delta G_{U\text{-}‡}$rel estimates used in drawings are described in Materials and Methods. The color scale is the same as that used previously (Figure 4). Ribbon representations of the protein structures are shown in the right panels for reference.

properties of the TS predicted for these proteins (Supplementary Data, Figure S4).

A variant of the amino terminal β-hairpin of ubiquitin has been found to fold independently at a rate of 17 μs^{-1},[50] hence agreeing with the β-hairpin folding-first hypothesis. The independent folding of various N-terminal peptides of wt ubiquitin also supports the hypothesis that structure could be acquired initially in an N-terminal segment including the β-hairpin and α1.[49,53] In addition, the most significant decrease in m_u upon mutations of ubiquitin occurs in the same structural region as that of the RBD, although additional decreases occur for mutation in β3 and β2 (Figure 7(c)). These resemblances of the m_u profiles suggest that α1 of both proteins pass through similar processes for consolidating their TS. A simulation using Cα Go-type models of ubiquitin suggested that the C terminus of α1 docks to the β-sheet, first at the level of the β-turn1, β2 and β3 and that the β5 adopts native state contacts only late in folding.[54] The results of the Ψ-value analysis, which is based on the relative effects on the transition and native state stability of the coordination of metal ion to various residue pairs, are also in agreement with the C-terminal part of α1 being preferentially consolidated at the TS of ubiquitin.[55] In summary, the comparison of TS structure of Raf RBD *versus* ubiquitin suggests that while these two proteins have very different sequences, they fold *via* similar TS and pathways. However, the higher polarization and less compact TS of ubiquitin *versus* Raf RBD (Figure 8 and β_t of 0.66 *versus* 0.76), differentiate these structural analogues. The similarities in the sequence entropy profile of Raf RBD variants obtained experimentally and of natural proteins classified in five ubiquitin-related superfamilies suggest that the determinants of formation and stabilization of the native state could be broadly conserved (Figure 1).

Comparison with distant structural analogues and discussion of other fold

Protein-L and G, which have been subjected to Φ-value analysis,[15,16] are classified in the IgG binding domain superfamily of the β-grasp ubiquitin-like topology and as such, they are distant structural analogues of Raf RBD. The protein-G and L nucleus are located in one of two, symmetrically disposed C and N-terminal β-hairpins. In contrast to Raf RBD and ubiquitin, the role of the hydrophobic core seems reduced in the IgG binding domains, whereas topology forming elements (e.g. the β-turns) gain importance. Nonetheless, the dispersion in protein-L structure of mutations generating high Φ-values is slightly more polarized than that of protein-G and corresponds more closely to ubiquitin and Raf RBD TS. In addition, the TS structures obtained by the classical method do not change dramatically from the perspective of $\Delta\Delta G_{U-\ddagger}$rel for these IgG binding domains (Figure 8(b) and (c)). In particular, the role of the major α-helix, which is analogous to α1, appears drastically diminished. Further evidence obtained from sequence perturbation and the introduction of several Gly into this α-helix confirmed its disruption at the TS of protein-L.[56] The IgG binding domain and ubiquitin-like superfamily share the same succession of secondary elements, but the β-sheet and α-helix are extended longitudinally and their axes more parallel in the former superfamily. Taken together, these observations indicate the inherent complexities in relating structural analogues and superfamilies and highlight how fold classification may be challenged by folding kinetics. From this perspective, it would be interesting to examine whether the TS structures of Raf RBD and ubiquitin are conserved across the diverse superfamilies of the β-grasp ubiquitin-like topology.

For most of the other proteins that we have scrutinized, the change in TS structure by using $\Delta\Delta G_{U-\ddagger}$rel was minor. However, we found one example in which the TS of a cold-shock protein was described to be polarized based on Φ-value analysis,[25] but appeared more diffuse according to $\Delta\Delta G_{U-\ddagger}$rel (Figure 8(d)). Such a delocalized TS would correspond better to the extremely high β_t (0.9) measured for this cold-shock protein.

Similarities in the folding TS of structural analogues: beyond sequence identity

Moderately high sequence identity (>30%) in a pair of structural analogues is a good indicator of similarities in the folding mechanism. However, as the sequence identity diminishes to lower level, several cases of discrepancies in the properties of the TS of structural analogues have been reported, as described in the Introduction for protein-L *versus* protein-G, AcP and SH3 folds. Hence, there are clearly other general factors impacting on the TS properties that escape detection by mere sequence comparison. For example, the reversal of the nucleus of protein-G to approximate protein-L TS suggests that secondary structure propensities influence TS properties.[30] In the same vein, secondary structure propensities are supposed to have an impact on the folding mechanism adopted by a polypeptide chain. Indeed, sequences that have high secondary structure propensities for their native conformation would favor the framework model.[18,27] The propensity of stretches of sequence to form β-turns has been shown to be critical for nucleus formation in several cases.[6,15,16] In distant structural analogues, the occurrences of insertions or deletions in turns and connecting loops are high. This phenomenon could in itself change the properties of TS. Similarly, the difference between the respective TS of Sso7d and of SH3 domains could originate from replacement of the C-terminal β-strand by a α-helix.[29] In Raf RBD and ubiquitin, the β-hairpin, which constitutes the most structured region of the folding nucleus is the segment with the most native-like calculated secondary structure propensities.[34,57] In contrast, the C-termi-

nal β-strand, β5, has been shown to adopt non-native structure in ubiquitin,[58,59] while it is unfolded at the TS.[26] Since ubiquitin and the RBD belong to the same superfamily they share tighter structural similarities and evolutionary relationships than predictable from their sequence identity alone. Thus, it is not surprising that the arrangement and identities of amino acids observed in the hydrophobic cores of these proteins is more conserved than the rest of the sequence,[33,34] even though the small distinctions in the organization of the hydrophobic core have an effect on the TS structures (Figure 8(a)). In contrast, the remotely related IgG binding domains adopt hydrophobic core organization and TS that are drastically different than those of the members of the ubiquitin-like superfamily analyzed herein (Figure 8(a)–(c)). The demonstration that the stability is conserved throughout artificial evolution highlight its potential role in defining the folding process and illustrate further the limitation imposed on the mutation of hydrophobic core residues.[34] Given the generally important role of the hydrophobic core in the formation of the TS, the pronounced conservation of its properties throughout evolution could be the major constraints leading structural analogues to fold *via* similar TS. The ratio of hydrophobic residues within a polypeptide chains can also influence the mechanism and type of TS formed via the hydrophobic collapse effect.[27,60] Nevertheless, as highlighted in elegant circular permutation experiments,[61] the entropic factor determined by the polypeptide chain connectivity dictate broadly the assembly of native contacts. Consequently, we expect generally speaking that evolutionary related structural analogues should fold *via* similar TS, notwithstanding the exceptions discussed above.

Conclusions and implications for modeling protein folding

The comparison of the TS for folding of proteins displaying structural analogies and remote sequence identities is instrumental to improve our comprehension of protein folding. At present, less than ten topologies have been thoroughly analyzed. Above, we have presented a model of the TS ensemble for folding of a member of the β-grasp ubiquitin-like topology, the Raf RBD, which has allowed comparison with data on structural analogues already documented in the literature. Specifically, the TS model was built using classic Φ-value interpretation and comparison with the newly introduced parameter $\Delta\Delta G_{U\text{-}\ddagger}$rel. The TS ensemble of Raf RBD appears polarized around the β-hairpin according to the Φ-value, but delocalized to all inner hydrophobic core residues according to the $\Delta\Delta G_{U\text{-}\ddagger}$rel. Clearly, the interpretation of Φ-values analysis is improved by considering $\Delta\Delta G^{1.6\ M}_{U\text{-}\ddagger}$rel and we believe this contributes to building a more comprehensive view of Raf RBD TS. In addition, we have shown that the TS and folding pathways for Raf RBD and ubiquitin are highly similar, despite insignificant sequence identity. This observation is in agreement with the concept that the folding mechanism is defined by coarse amino acid composition and topological characteristics.

Φ-Value analyses have been used successfully as constraints in folding simulations to build detailed TS models.[62,63] However, our results suggest that for proteins, such as Raf RBD, in which the TS is polarized according to native contacts formed while energetically diffuse, the addition of the $\Delta\Delta G_{U\text{-}\ddagger}$ to the computational method should improve the accuracy of the model.

Materials and Methods

Mutants: description, cloning and purification

Details on the methods for cloning and purification of mutants can be found in a complementary article.[34] The atypical mutants with no self-explanatory name are described: H2, in which residues 62–65 (Leu-Pro-Asn-Lys) are replaced with the amino acid Phe-Thr-Asp-Gly, was recovered from the sequence perturbation experiment;[33] H2_F62L is derived from H2 by reverting residue 62 to the wt amino acid (Leu-Thr-Asp-Gly); Δ104-6 and Δ101-8 + AG are deletion mutants in which residues 104 to 106 are deleted (Glu-His-Lys) and 101 to 108 are replaced by Ala-Gly, similarly to ubiquitin, respectively.

Kinetics and chevron curves

The design of kinetic experiments were designed in agreement with consensus experimental standards[36]. The kinetic reactions were followed using Applied Photophysics SX.18MV stopped-flow. The experiments were performed at 25(±0.1) °C in 50 mM sodium phosphate buffer (pH 7.0) containing 1 mM DTT, using urea as denaturant. Kinetic traces were performed with 3–4 μM RBD, using an excitation wavelength of 281 nm and a 320 nm cut-off filter to detect the fluorescence emitted by W114, the unique tryptophan of this polypeptide. The refolding reactions were initiated from proteins diluted in 9–9.75 M urea. The unfolding reactions were initiated from proteins diluted in 1–2 M urea to avoid aggregation. Data between 3.5 ms and 10 s were used for fitting traces and four to seven traces were averaged for each data point. Refolding traces at low denaturant concentration were fit to four exponentials as described for wt Raf RBD.[36] No deviation from linearity of denaturant dependence of $\ln k_f$ was observed at either end of the chevron plot of all mutants tested, except below 2 M urea in the refolding arm of V72A. In this case, these data points were removed from the chevron curves. The unfolding traces were fit to simple exponentials. Chevron curves were fit to two-state equations.

$\Delta\Delta G_{F\text{-}U}$ and Φ-value estimates

Φ is particularly sensitive to errors in $\Delta\Delta G_{F\text{-}U}$ (as $\Phi = \Delta\Delta G_{U\text{-}\ddagger}/\Delta\Delta G_{F\text{-}U}$). To address this issue, studies on CI2 and protein-L presented Φ-value estimates obtained from independent calculations of free energy changes.[15,41]

Three equations were used to calculate free energy changes:

$$\Delta\Delta G^{\mathrm{kin}}_{\mathrm{F-U}} = RT\left(\ln\left(k_f^{1.6M(wt)}/k_u^{8M(wt)}\right) - \ln\left(k_f^{1.6M(mut)}/k_u^{8M(mut)}\right)\right) \quad (1)$$

$$\Delta\Delta G^{Cm}_{\mathrm{F-U}} = \langle m\rangle\left(Cm^{(wt)} - Cm^{(mut)}\right) \quad (2)$$

$$\Delta\Delta G^{5.8M}_{\mathrm{F-U}} = m^{(wt)}(Cm^{(wt)} - 5.8) - m^{(mut)}(Cm^{(mut)} - 5.8) \quad (3)$$

where $\langle m\rangle$ is the mean m value (3.90(±0.33) kj mol^{-1} M^{-1}) derived from all mutants tested, $Cm^{(wt)}$ and $Cm^{(mut)}$ are the concentrations of urea at which 50% of wt and mutant proteins are folded, $k_f^{1.6\ M(wt)}$ and $k_f^{1.6\ M(mut)}$ are the folding rates in 1.6 M urea of the wt and mutants, respectively. Similarly, $k_u^{8\ M(wt)}$ and $k_u^{8\ M(mut)}$ are the unfolding rates in 8 M urea for the wt and mutants.

Using the three estimates of free energy changes described above (equations (1)–(3)), different Φ-value estimates were calculated:

$$\Phi_F^{\mathrm{kin}} = RT\ln\left(k_f^{wt(1.6M)}/k_f^{mut(1.6M)}\right)/\Delta\Delta G^{\mathrm{kin}}_{\mathrm{F-U}} \quad (4)$$

$$\Phi_F^{Cm} = RT\ln\left(k_f^{H2O(wt)}/k_f^{H2O(mut)}\right)/\Delta\Delta G^{Cm}_{\mathrm{F-U}} \quad (5)$$

$$\Phi_U = -RT\ln\left(k_u^{5.8M(wt)}/k_u^{5.8M(mut)}\right)/\Delta\Delta G^{5.8M}_{\mathrm{F-U}} \quad (6)$$

where k_f^{wt} and k_f^{mut} are the folding rates in water obtained by extrapolation from wt and mutants chevron curves, respectively; $k_u^{5.8\ M(wt)}$ and $k_u^{5.8\ M(mut)}$ are the folding rates in 5.8 M urea for wt and mutants, respectively. Φ_F^{ext} mentioned in the text corresponds to Φ_F calculated from fully extrapolated thermodynamic and kinetic data, i.e. k_f^{H2O} and $\Delta\Delta G_{F\text{-}U}$ (Supplementary Data, Table S3).

ΔΔ$G_{U\text{-}‡}$ and ΔΔ$G_{U\text{-}‡}$rel calculation

The $\Delta\Delta G^{1.6\ M}_{U\text{-}‡}$ is obtained from:

$$\Delta\Delta G^{1.6M}_{U-‡} = RT\ln\left(k_f^{1.6M(wt)}/k_f^{1.6M(mut)}\right) \quad (7)$$

$\Delta\Delta G^{1.6\ M}_{U\text{-}‡}$rel is the normalization of $\Delta\Delta G^{1.6\ M}_{U\text{-}‡}$ obtained for a mutant against the most TS destabilized mutant ($\Delta\Delta G^{1.6\ M}_{U\text{-}‡}$ max) obtained in the collection of mutants:

$$\Delta\Delta^{1.6M}_{U-‡}\mathrm{rel} = \Delta\Delta G^{1.6M}_{U-‡}/\Delta\Delta G^{1.6M}_{U-‡}\mathrm{max} \quad (8)$$

In the case of Raf RBD and ubiquitin, the maximum destabilization of the transition state is observed with I58A and L15A (equivalent to V70 in Raf RBD), respectively. For Supplementary Data, Figure S3, the $\Delta\Delta G_{U\text{-}‡}$ values were normalized for side-chain volume variation according to the volume scale of Richards[64] and by arbitrarily giving to mutation A→G, a correction factor of 1.

Manipulation of the data from previous studies: Φ-values and ΔΔ$G_{U\text{-}‡}$rel of other proteins discussed

The Φ-values and $\Delta\Delta G_{U\text{-}‡}$rel were directly derived from the data reported in Tables of the original studies. (1) Ubiquitin average Φ-values obtained from unfolding and folding experiments were used in the graph and structural scheme for comparison with Raf RBD. The $\Delta\Delta G_{U\text{-}‡}$rel values were calculated from k_f^{H2O} (26). In Supplementary Data, Figure S4(a), the 16 pairs or mutations of Raf RBD and ubiquitin compared are: I3A-I58A, V5A-V60A, T7A-L62A, L15A-V70A, V17A-V72A, T22A-S77A, I23A-L78A, V26-C81, K27-L82, I30A-A85G, Q41A-C96A, L43A-V98A, L50A-L112A, L56A-A118G, L67-L126 and L69-V128. (2) For protein-L the equivalent of Φ_F^{kin} was utilized. $\Delta\Delta G_{U\text{-}‡}$rel values were calculated from $k_f^{0.4\ M}$, a value used for the calculation of the previous parameter.[15] (3) For protein-G the equivalent of Φ_F^{kin} was utilized. $\Delta\Delta G_{U\text{-}‡}$rel values were calculated from $k_f^{0.5\ M}$, a value used for the calculation of the previous parameter.[16] (4) For the cold shock protein, mean Φ were derived from two Φ-value estimates: (i) refolding kinetics and thermodynamic parameters (alternatively, from unfolding kinetics data in exceptional cases), and (ii) solely from kinetic experiments. The $\Delta\Delta G_{U\text{-}‡}$rel values were calculated from k_f^{H2O}.[25]

Structure representation

Structural representations were created using MolMol software and the molecular coordinates of Raf RBD, ubiquitin, protein-L, protein-G and cold shock protein: 1RFA, 1UBI, 1HZ6, 2GB1 and 1CSP, respectively. The residues which side-chains are represented by spheres were identified using their regular residue numbering, except in the case of ubiquitin for which the Raf RBD residue numbers was used to facilitate comparison.

Acknowledgements

We are grateful to Dr Jeffrey W. Keillor for discussion and access to stopped-flow apparatus; Alexis Vallée-Belisle for discussions and data exchange; Dr Emily Manderson and Dr Allan Davidson for reading the manuscript and inputs; Dr William A. Eaton for communicating work ahead of print; Dr Jim Omichinski for sharing materials. We also want to highlight the contribution of the reviewers. The NSERC and CIHR funded this project. F.-X.C.V. was a scholar of CIHR, le programme de biologie moléculaire and the FES. S.W.M. is the Canada Research Chair in Integrative Genomics.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2006.10.079

References

1. Jackson, S. E. & Fersht, A. R. (1991). Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry*, **30**, 10428–10435.
2. Fersht, A. R., Matouschek, A. & Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224**, 771–782.

3. Serrano, L., Matouschek, A. & Fersht, A. R. (1992). The folding of an enzyme. III. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. *J. Mol. Biol.* **224**, 805–818.
4. Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995). The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.* **254**, 260–288.
5. Kragelund, B. B., Osmark, P., Neergaard, T. B., Schiodt, J., Kristiansen, K., Knudsen, J. & Poulsen, F. M. (1999). The formation of a native-like structure containing eight conserved hydrophobic residues is rate limiting in two-state protein folding of ACBP. *Nature Struct. Biol.* **6**, 594–601.
6. Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I. & Baker, D. (1999). Experiment and theory highlight role of native state topology in SH3 folding. *Nature Struct. Biol.* **6**, 1016–1024.
7. Martinez, J. C. & Serrano, L. (1999). The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nature Struct. Biol.* **6**, 1010–1016.
8. Northey, J. G., Di Nardo, A. A. & Davidson, A. R. (2002). Hydrophobic core packing in the SH3 domain folding transition state. *Nature Struct. Biol.* **9**, 126–130.
9. Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M. & Dobson, C. M. (1999). Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nature Struct. Biol.* **6**, 1005–1009.
10. Ternstrom, T., Mayor, U., Akke, M. & Oliveberg, M. (1999). From snapshot to movie: phi analysis of protein folding transition states taken one step further. *Proc. Natl Acad. Sci. USA*, **96**, 14854–14859.
11. Jager, M., Nguyen, H., Crane, J. C., Kelly, J. W. & Gruebele, M. (2001). The folding mechanism of a beta-sheet: the WW domain. *J. Mol. Biol.* **311**, 373–393.
12. Hamill, S. J., Steward, A. & Clarke, J. (2000). The folding of an immunoglobulin-like Greek key protein is defined by a common-core nucleus and regions constrained by topology. *J. Mol. Biol.* **297**, 165–178.
13. Fowler, S. B. & Clarke, J. (2001). Mapping the folding pathway of an immunoglobulin domain: structural detail from Phi value analysis and movement of the transition state. *Structure (Camb.)*, **9**, 355–366.
14. Wright, C. F., Lindorff-Larsen, K., Randles, L. G. & Clarke, J. (2003). Parallel protein-unfolding pathways revealed and mapped. *Nature Struct. Biol.* **10**, 658–662.
15. Kim, D. E., Fisher, C. & Baker, D. (2000). A breakdown of symmetry in the folding transition state of protein L. *J. Mol. Biol.* **298**, 971–984.
16. McCallister, E. L., Alm, E. & Baker, D. (2000). Critical role of beta-hairpin formation in protein G folding. *Nature Struct. Biol.* **7**, 669–673.
17. Capaldi, A. P., Kleanthous, C. & Radford, S. E. (2002). Im7 folding mechanism: misfolding on a path to the native state. *Nature Struct. Biol.* **9**, 209–216.
18. Gianni, S., Guydosh, N. R., Khan, F., Caldas, T. D., Mayor, U., White, G. W. *et al.* (2003). Unifying features in protein-folding mechanisms. *Proc. Natl Acad. Sci. USA*, **100**, 13286–13291.
19. Sanchez, I. E. & Kiefhaber, T. (2003). Origin of unusual phi-values in protein folding: evidence against specific nucleation sites. *J. Mol. Biol.* **334**, 1077–1085.
20. Los Rios, M. A., Muralidhara, B. K., Wildes, D., Sosnick, T. R., Marqusee, S., Wittung-Stafshede, P. *et al.* (2006). On the precision of experimentally determined protein folding rates and phi-values. *Protein Sci.* **15**, 553–563.
21. Sosnick, T. R., Dothager, R. S. & Krantz, B. A. (2004). Differences in the folding transition state of ubiquitin indicated by phi and psi analyses. *Proc. Natl Acad. Sci. USA*, **101**, 17377–17382.
22. Feng, H., Vu, N. D., Zhou, Z. & Bai, Y. (2004). Structural examination of phi-value analysis in protein folding. *Biochemistry*, **43**, 14325–14331.
23. Settanni, G., Rao, F. & Caflisch, A. (2005). Phi-value analysis by molecular dynamics simulations of reversible folding. *Proc. Natl Acad. Sci. USA*, **102**, 628–633.
24. Sanchez, I. E. & Kiefhaber, T. (2003). Hammond behavior versus ground state effects in protein folding: evidence for narrow free energy barriers and residual structure in unfolded states. *J. Mol. Biol.* **327**, 867–884.
25. Garcia-Mira, M. M., Boehringer, D. & Schmid, F. X. (2004). The folding transition state of the cold shock protein is strongly polarized. *J. Mol. Biol.* **339**, 555–569.
26. Went, H. M. & Jackson, S. E. (2005). Ubiquitin folds through a highly polarized transition state. *Protein Eng. Des. Sel.* **18**, 229–237.
27. Daggett, V. & Fersht, A. (2003). The present view of the mechanism of protein folding. *Nature Rev. Mol. Cell Biol.* **4**, 497–502.
28. Geierhaas, C. D., Paci, E., Vendruscolo, M. & Clarke, J. (2004). Comparison of the transition states for folding of two Ig-like proteins from different superfamilies. *J. Mol. Biol.* **343**, 1111–1123.
29. Guerois, R. & Serrano, L. (2000). The SH3-fold family: experimental evidence and prediction of variations in the folding pathways. *J. Mol. Biol.* **304**, 967–982.
30. Nauli, S., Kuhlman, B. & Baker, D. (2001). Computer-based redesign of a protein folding pathway. *Nature Struct. Biol.* **8**, 602–605.
31. Zarrine-Afsar, A., Larson, S. M. & Davidson, A. R. (2005). The family feud: do proteins with similar structures fold *via* the same pathway? *Curr. Opin. Struct. Biol.* **15**, 42–49.
32. Emerson, S. D., Madison, V. S., Palermo, R. E., Waugh, D. S., Scheffler, J. E., Tsao, K. L. *et al.* (1995). Solution structure of the Ras-binding domain of c-Raf-1 and identification of its Ras interaction surface. *Biochemistry*, **34**, 6911–6918.
33. Campbell-Valois, F. X., Tarassov, K. & Michnick, S. W. (2005). Massive Sequence Perturbation of a Small Protein. *Proc. Natl Acad. Sci.* **102**, 14988–14993.
34. Campbell-Valois, F. X., Tarassov, K. & Michnick, S. W. (2006). Massive sequence perturbation of the Raf *ras* binding domain reveals relationships between sequence conservation, secondary structure propensity, hydrophobic core organization and stability. *J. Mol. Biol.* **362**, 151–171.
35. Vallee-Belisle, A., Turcotte, J. F. & Michnick, S. W. (2004). raf RBD and ubiquitin proteins share similar folds, folding rates and mechanisms despite having unrelated amino acid sequences. *Biochemistry*, **43**, 8447–8458.
36. Maxwell, K. L., Wildes, D., Zarrine-Afsar, A., Los Rios, M. A., Brown, A. G., Friel, C. T. *et al.* (2005). Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci.* **14**, 602–616.
37. Matouschek, A. & Fersht, A. R. (1993). Application of physical organic chemistry to engineered mutants of proteins: Hammond postulate behavior in the transition state of protein folding. *Proc. Natl Acad. Sci. USA*, **90**, 7814–7818.
38. Matouschek, A., Otzen, D. E., Itzhaki, L. S., Jackson, S. E. & Fersht, A. R. (1995). Movement of the position

of the transition state in protein folding. *Biochemistry*, **34**, 13656–13662.
39. Matthews, J. M. & Fersht, A. R. (1995). Exploring the energy surface of protein folding by structure-reactivity relationships and engineered proteins: observation of Hammond behavior for the gross structure of the transition state and anti-Hammond behavior for structural elements for unfolding/folding of barnase. *Biochemistry*, **34**, 6805–6814.
40. Otzen, D. E. & Oliveberg, M. (2002). Conformational plasticity in folding of the split beta-alpha-beta protein S6: evidence for burst-phase disruption of the native state. *J. Mol. Biol.* **317**, 613–627.
41. Jackson, S. E., Moracci, M., elMasry, N., Johnson, C. M. & Fersht, A. R. (1993). Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. *Biochemistry*, **32**, 11259–11269.
42. Aurora, R. & Rose, G. D. (1998). Helix capping. *Protein Sci.* **7**, 21–38.
43. Block, C., Janknecht, R., Herrmann, C., Nassar, N. & Wittinghofer, A. (1996). Quantitative structure-activity analysis correlating Ras/Raf interaction *in vitro* to Raf activation *in vivo*. *Nature Struct. Biol.* **3**, 244–251.
44. Fersht, A. R., Itzhaki, L. S., elMasry, N. F., Matthews, J. M. & Otzen, D. E. (1994). Single versus parallel pathways of protein folding and fractional formation of structure in the transition state. *Proc. Natl Acad. Sci. USA*, **91**, 10426–10429.
45. Wong, K. B., Clarke, J., Bond, C. J., Neira, J. L., Freund, S. M., Fersht, A. R. & Daggett, V. (2000). Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. *J. Mol. Biol.* **296**, 1257–1282.
46. Kazmirski, S. L., Wong, K. B., Freund, S. M., Tan, Y. J., Fersht, A. R. & Daggett, V. (2001). Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proc. Natl Acad. Sci. USA*, **98**, 4349–4354.
47. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. & Fersht, A. R. (2005). Solution structure of a protein denatured state and folding intermediate. *Nature*, **437**, 1053–1056.
48. Munoz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. (1997). Folding dynamics and mechanism of beta-hairpin formation. *Nature*, **390**, 196–199.
49. Zerella, R., Evans, P. A., Ionides, J. M., Packman, L. C., Trotter, B. W., Mackay, J. P. & Williams, D. H. (1999). Autonomous folding of a peptide corresponding to the N-terminal beta-hairpin from ubiquitin. *Protein Sci.* **8**, 1320–1331.
50. Munoz, V., Ghirlando, R., Blanco, F. J., Jas, G. S., Hofrichter, J. & Eaton, W. A. (2006). Folding and aggregation kinetics of a beta-hairpin. *Biochemistry*, **45**, 7023–7035.
51. Hedberg, L. & Oliveberg, M. (2004). Scattered Hammond plots reveal second level of site-specific information in protein folding: phi′ (beta++). *Proc. Natl Acad. Sci. USA*, **101**, 7606–7611.
52. Sanchez, I. E. & Kiefhaber, T. (2003). Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* **325**, 367–376.
53. Cox, J. P., Evans, P. A., Packman, L. C., Williams, D. H. & Woolfson, D. N. (1993). Dissecting the structure of a partially folded protein. Circular dichroism and nuclear magnetic resonance studies of peptides from ubiquitin. *J. Mol. Biol.* **234**, 483–492.
54. Zhang, J., Qin, M. & Wang, W. (2005). Multiple folding mechanisms of protein ubiquitin. *Proteins: Struct. Funct. Genet.* **59**, 565–579.
55. Krantz, B. A., Dothager, R. S. & Sosnick, T. R. (2004). Discerning the structure and energy of multiple transition states in protein folding using psi-analysis. *J. Mol. Biol.* **337**, 463–475.
56. Kim, D. E., Yi, Q., Gladwin, S. T., Goldberg, J. M. & Baker, D. (1998). The single helix in protein L is largely disrupted at the rate-limiting step in folding. *J. Mol. Biol.* **284**, 807–815.
57. Srinivasan, R. & Rose, G. D. (1999). A physical basis for protein secondary structure. *Proc. Natl Acad. Sci. USA*, **96**, 14258–14263.
58. Harding, M. M., Williams, D. H. & Woolfson, D. N. (1991). Characterization of a partially denatured state of a protein by two-dimensional NMR: reduction of the hydrophobic interactions in ubiquitin. *Biochemistry*, **30**, 3120–3128.
59. Kitahara, R., Yamada, H. & Akasaka, K. (2001). Two folded conformers of ubiquitin revealed by high-pressure NMR. *Biochemistry*, **40**, 13556–13563.
60. Calloni, G., Taddei, N., Plaxco, K. W., Ramponi, G., Stefani, M. & Chiti, F. (2003). Comparison of the folding processes of distantly related proteins. Importance of hydrophobic content in folding. *J. Mol. Biol.* **330**, 577–591.
61. Lindberg, M. O., Haglund, E., Hubner, I. A., Shakhnovich, E. I. & Oliveberg, M. (2006). Identification of the minimal protein-folding nucleus through loop-entropy perturbations. *Proc. Natl Acad. Sci. USA*, **103**, 4083–4088.
62. Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. (2001). Three key residues form a critical contact network in a protein folding transition state. *Nature*, **409**, 641–645.
63. Li, L. & Shakhnovich, E. I. (2001). Constructing, verifying, and dissecting the folding transition state of chymotrypsin inhibitor 2 with all-atom simulations. *Proc. Natl Acad. Sci. USA*, **98**, 13014–13018.
64. Richards, F. M. (1974). The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.* **82**, 1–14.

Edited by F. Schmid

(Received 23 May 2006; received in revised form 5 September 2006; accepted 24 October 2006)
Available online 28 October 2006

Note added in proof: The authors note that the following works should be consulted in addition to Gianni et al.[18]: (1) Abkevich, V. I., Gutin, A. M. & Shakhnovich, E. I. (1994). *Biochemistry*, **33**, 10026–10036; (2) Shakhnovich, E. (2006). *Chem. Rev.* **106**, 1559–1588.