

Φ -Value analysis and the nature of protein-folding transition states

Alan R. Fersht* and Satoshi Sato

Medical Research Council Centre for Protein Engineering, Hills Road, Cambridge CB2 2QH, United Kingdom

Contributed by Alan R. Fersht, April 16, 2004

Φ values are used to map structures of protein-folding transition states from changes in free energies of denaturation ($\Delta\Delta G_{D-N}$) and activation on mutation. A recent reappraisal proposed that Φ values for $\Delta\Delta G_{D-N} < 1.7$ kcal/mol are artifactual. On discarding such derived Φ values from published studies, the authors concluded that there are no high Φ values in diffuse transition states, which are consequently uniformly diffuse with no evidence for nucleation. However, values of $\Delta\Delta G_{D-N} > 1.7$ kcal/mol are often found for large side chains that make dispersed tertiary interactions, especially in hydrophobic cores that are in the process of being formed in the transition state. Conversely, specific local interactions that probe secondary structure tend to have $\Delta\Delta G_{D-N} \approx 0.5$ – 2 kcal/mol. Discarding Φ values from lower-energy changes discards the crucial information about local interactions and makes transition states appear uniformly diffuse by overemphasizing the dispersed tertiary interactions. The evidence for the 1.7 kcal/mol cutoff was based on mutations that had been deliberately designed to be unsuitable for Φ -value analysis because they are structurally disruptive. We confirm that reliable Φ values can be derived from the recommended mutations in suitable proteins with $0.6 < \Delta\Delta G_{D-N} < 1.7$ kcal/mol, and there are many reliable high Φ values. Transition states vary from being rather diffuse to being well formed with islands of near-complete secondary structure. We also confirm that the structures of transition-state ensembles can be perturbed by mutations with $\Delta\Delta G_{D-N} \gg 2$ kcal/mol and that protein-folding transition states do move on the energy surface on mutation.

barnase | protein A | nucleation–condensation | framework | Hammond

The Φ -value analysis is a particular set of protein-engineering methods that is used to map the structures of transition states and intermediates in protein-folding, catalysis, binding, and conformational transitions of proteins at the level of individual residues (1–5). Φ is the ratio of change of free energy of activation for folding, $\Delta\Delta G_{\ddagger-D}$, to the equilibrium free energy of folding, $\Delta\Delta G_{N-D}$,[†] and scores the extent of formation of structure on a scale of 0 to 1 at the level of individual residues. Φ is similar but not identical to the constants α or β of classical rate-equilibrium-free-energy relationships (REFERs) of covalent-bond chemistry. Linear free-energy relationships are the classical means of analyzing the structures of transition states. The structure of a reagent is subtly altered by small changes, and the consequent perturbations of the kinetics and equilibrium of the reaction are measured. Under certain circumstances, which often rely on the chemist's judgement in making sensible structural changes, there can be a linear relationship between ΔG_{\ddagger} , the change in activation energy, and ΔG^0 , the change in equilibrium free energy; i.e., $(\partial\Delta G_{\ddagger}/\partial\text{structure})/(\partial\Delta G^0/\partial\text{structure}) = \alpha$ in an REFER (6) (or $= \beta$ in the earlier Brønsted plots for catalysis). The α (or β) value is a measure of extent of covalent-bond making or breaking in the reaction, with $\alpha = 0$ implying no bond making (or breaking) and $\alpha = 1$ implying complete making (or breaking). Intermediate values of α imply partial bond making or breaking. Protein engineering allowed the equivalent of REFERs to be applied to changes in noncovalent interactions of protein side chains as true multipoint plots

and as a collection of two-point Φ values (1, 2). In general, protein engineering gives two-point, mutationally specific plots, and the existence of multipoint plots is a bonus when all mutations respond with the same α or β (2).

There are important differences between Φ and α or β . Many chemical processes respond smoothly to changes in structure that are remote from the seat of reaction, but Φ can depend on the specific interactions that are mutated and the change in energetics of the denatured state, including its solvation energy, on mutation (3, 4, 7). These energies affect the interpretation of Φ and require that Φ -value analysis is best applied to certain types of mutation, interpreted within various constraints, and can be difficult for fractional values of Φ . In the purest form of Φ -value analysis, mutations are made that delete interactions that stabilize the native state of the protein without disrupting the structure of the protein, introducing new interactions or changing stereochemistry: “nondisruptive deletion mutations” (2), preferably of hydrophobic moieties, although more radical mutations of surface residues are permitted (4).

Φ -value analysis should be applied with the following caveats. (i) In general, fractional values of Φ are not linear with the extent of bond formation, and only the extreme values of 0 and 1 are fully interpretable *per se* as being completely denatured-like and completely native-like, respectively [the term “denatured” is used because denatured states can have residual structure, and thus changes are measured relative to the residual structure (8)]. (ii) However, Φ for nondisruptive deletion mutations of hydrophobic side chains is approximately linear with the extent of formation of the bonds between denatured and native conformations, and thus fractional values are a good indicator of the extent of noncovalent-bond formation. (iii) Deletion of large side chains can alter many interactions between different substructures and thus give an average value of Φ for all of those interactions [a fine structure analysis is required to dissect the different components (9)]. (iv) Deletion of large side chains may also move the transition state by Hammond and anti-Hammond effects (10–12). (v) Because of the uncertainties in interpretation of Φ , as many Φ values as possible should be made and the results divided into classes of “weak,” “medium,” and “strong” (as with nuclear Overhauser effects in NMR spectroscopy) for purposes of combining with simulation to obtain atomic-level resolution of transition states (13).

Φ values *per se* can give extensive atomic-level information on structures of transition states. For example, Φ -value analysis on chymotrypsin-inhibitor 2 (CI2) provided the experimental evidence for the nucleation–condensation mechanism of folding in which secondary and tertiary interactions form together in the transition state, which appears to form around an extended nucleus that has moderate Φ values, with Φ falling off with distance from the nucleus (5, 14, 15). In general, Φ values are the

Abbreviations: REFER, rate-equilibrium-free-energy relationship; CI2, chymotrypsin-inhibitor 2.

*To whom correspondence should be addressed. E-mail: arf25@cam.ac.uk.

[†] Φ can be measured for folding or unfolding. $\Delta\Delta G_{N-D}$, the free energy of folding, is equal to $-\Delta\Delta G_{D-N}$, the free energy of denaturation.

© 2004 by The National Academy of Sciences of the USA

experimental data for benchmarking simulation and for reconstructing protein-folding transition states and pathways at atomic resolution by combining experiment and simulation, and Φ -value analysis is at its most powerful when in combination with simulation (16–21). The theoretical studies use a global analysis of the Φ values, not just a selected few.

A recent appraisal of the results of Φ -value analysis concluded that measurements should be restricted to those for $\Delta\Delta G_{D-N}$ values >1.7 kcal/mol (7 kJ/mol) (22). It also was concluded that most high values of Φ were artifacts of $\Delta\Delta G_{D-N}$ being <7 kJ/mol. On discarding the “artificially” high Φ values, it was concluded that “diffuse” protein-folding transition states are uniformly diffuse, which is counter to much of theory, experiment, and simulation (16, 23–27). The authors’ reasoning was based on the premise that Φ should be the same for all mutations at the same site, and thus deviations from this will show which Φ values are inaccurate (22). They tested the linearity hypothesis on a data set (28), which appeared to give the 1.7 kcal/mol cutoff point. However, in many cases, Φ is mutation-specific, because each mutation removes different interactions and has different effects on the denatured state (3, 4) and the data set used to calibrate Φ values had been deliberately designed by Davidson and coworkers (28) to be unsuitable for Φ -value analysis.

We formally demonstrate the flaws in the Φ -value reappraisal and confirm from experimental data that Φ values from non-disruptive mutations can be adequately reliable down to a $\Delta\Delta G_{D-N}$ value of ≈ 0.6 kcal/mol (2.5 kJ/mol) for suitable proteins. We show how the crucial probes for the formation of local secondary structure tend to have $\Delta\Delta G_{D-N}$ values of ≈ 0.6 –2 kcal/mol (2.5–8 kJ/mol), and thus discarding their values discards the evidence for nucleation sites. We also refute the proposal (22, 29) that structures of transition-state ensembles are not affected by mutation. We use the REFER methods that supposedly indicated the 1.7 kcal/mol cutoff point to show that the cutoff is closer to 0.6 kcal/mol for suitable proteins.

Formal Analysis of Φ Values.

Relationship Between Φ and α . The premise of Sanchez and Kiefhaber (22) is that Φ is identical to the Leffler α and that all mutations at a particular position should give the same value of Φ . However, there is a crucial difference between the Leffler α and Φ , which may be shown formally. Let the free energy of the native state N be G_N , that of the denatured state be G_D , that of the transition state be G_{TS} , and mutant states be denoted by a prime (Fig. 1). The free energy of denaturation $\Delta G_{D-N} = G_D - G_N$; the free energy of folding $\Delta G_{N-D} = G_N - G_D$; the activation energy of unfolding $\Delta G_{TS-N} = G_{TS} - G_N$; and the activation energy of folding $\Delta G_{TS-D} = G_{TS} - G_D$. Abbreviating structure to “S”:

$$\partial(G_{TS} - G_D)/\partial S = \partial G_{TS}/\partial S - \partial G_D/\partial S \quad [1]$$

$$\partial(G_N - G_D)/\partial S = \partial G_N/\partial S - \partial G_D/\partial S. \quad [2]$$

Thus,

$$\alpha = (\partial G_{TS}/\partial S - \partial G_D/\partial S)/(\partial G_N/\partial S - \partial G_D/\partial S). \quad [3]$$

The change in free energy of the denatured state with mutation, $\partial G_D/\partial S$, is an important component of α . It was the essential difference between the classical Leffler α or Brønsted β that led us to give the free-energy constant a new name, Φ (3).

The experimentally accessible quantities used in Φ -value analysis are equilibrium and kinetic data of denaturation measured on wild-type and mutant proteins separately, ΔG_{D-N} , $\Delta G_{D'-N'}$, ΔG_{TS-N} , $\Delta G_{TS'-N'}$, ΔG_{TS-D} , and $\Delta G_{TS'-D'}$ (Fig. 1). The observed difference in free energy of denaturation of wild-type protein and that of a mutant (denoted by ') $\Delta\Delta G_{D-N} = \Delta G_{D'-N'} - \Delta G_{D-N}$. The difference in free energy of activation

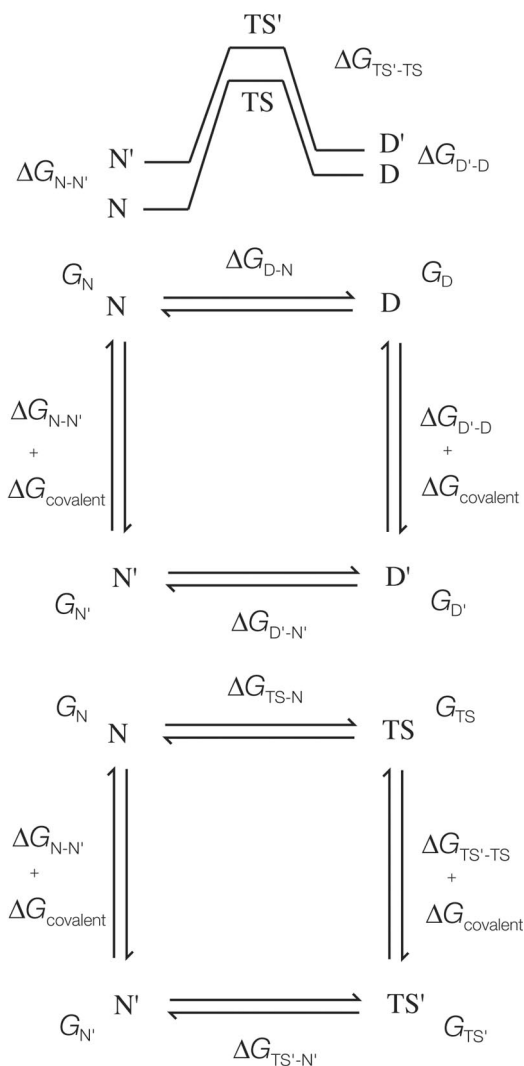


Fig. 1. Schematics of thermodynamic cycles and free-energy profiles.

of unfolding $\Delta\Delta G_{TS-N} = \Delta G_{TS'-N'} - \Delta G_{TS-N}$. The difference in free energy of activation of folding $\Delta\Delta G_{TS-D} = \Delta G_{TS'-D'} - \Delta G_{TS-D}$, which for two-state kinetics has the same transition state. The observed experimental quantities are related to the changes in the individual states on mutation for virtual thermodynamic cycles based on Fig. 1 (3, §): $\Delta\Delta G_{D-N} = \Delta G_{D'-D} - \Delta G_{N'-N}$; $\Delta\Delta G_{TS-N} = \Delta G_{TS'-TS} - \Delta G_{N'-N}$; and $\Delta\Delta G_{TS-D} = \Delta G_{TS'-TS} - \Delta G_{D'-D}$.

The experimentally determined Φ value for folding, Φ_F ,[‡] is defined by

$$\Phi_F = \Delta\Delta G_{TS-D}/\Delta\Delta G_{N-D}. \quad [4]$$

In terms of the other changes in the cycle,

$$\Phi_F = (\Delta G_{TS'-TS} - \Delta G_{D'-D})/(\Delta G_{N'-N} - \Delta G_{D'-D}). \quad [5]$$

Eq. 5 is the two-point version of Eq. 3 for a finite change in structure.

[‡]Although of routine use in physical chemistry, the application of thermodynamic cycles to protein folding was described as incorrect because they have hypothetical steps (30).

[†]Folding can be more complicated than unfolding because of (unknown) residual structure in the denatured state and changes in rate-determining steps, and Φ -value analysis is applied more often to unfolding kinetics.

The change in free energy on mutating N to N', $\Delta G_{N'-N}$, can be split up into notional components: the change in energy of the covalent bond that is mutated, ΔG_{cov} , the change in noncovalent interactions at the site of mutation, ΔG_{noncov} (without any reorganization of the structure of the protein on mutation), plus any additional changes because of the reorganization of the protein, ΔG_{reorg} , and the change in solvation energy, ΔG_{solv} (4, 7). ΔG_{cov} is the same for all states and drops out of the equations.

$$\begin{aligned} \Phi_F = & [\Delta G_{\text{noncov}(\text{TS}'\text{-TS})} + \Delta G_{\text{reorg}(\text{TS}'\text{-TS})} + \Delta G_{\text{solv}(\text{TS}'\text{-TS})} \\ & - \Delta G_{\text{noncov}(\text{D}'\text{-D})} - \Delta G_{\text{reorg}(\text{D}'\text{-D})} \\ & - \Delta G_{\text{solv}(\text{D}'\text{-D})}] / [\Delta G_{\text{noncov}(N'-N)} \\ & + \Delta G_{\text{reorg}(N'-N)} + \Delta G_{\text{solv}(N'-N)} - \Delta G_{\text{noncov}(\text{D}'\text{-D})} \\ & - \Delta G_{\text{reorg}(\text{D}'\text{-D})} - \Delta G_{\text{solv}(\text{D}'\text{-D})}] \quad [6] \end{aligned}$$

The presence of the $\Delta G_{\text{D}'\text{-D}}$ term in Eq. 5 [= $\Delta G_{\text{reorg}(\text{D}'\text{-D})} + \Delta G_{\text{solv}(\text{D}'\text{-D})}$ in Eq. 6] can lead to Φ being uninterpretable, and the ΔG_{noncov} terms can be mutation-specific (2, 3, 7).

Φ is identical to the classical α or β under two extreme circumstances. (i) The target region is as unstructured in the transition state as in the denatured state. Under these circumstances, all the ΔG terms in TS'-TS are the same as those of D'-D, and Eqs. 5 and 6 reduce to $\Phi_F = 0$ (and for two-state folding, Φ for unfolding, $\Phi_U = 1$). (ii) The target region is as structured in the transition state as in the native state. Under these circumstances; all the ΔG terms in TS'-TS are the same as those of N'-N, and Eqs. 5 and 6 reduce to $\Phi_F = 1$ (and for two-state folding, $\Phi_U = 1$). The extreme values of 0 and 1 should be interpretable, therefore, for all mutations.

Fractional values of Φ are readily interpretable for the chemically sensible nondisruptive deletion mutations, especially of hydrophobic side chains, because the ΔG_{reorg} terms are minimized (2, 4) as are also the ΔG_{solv} for aliphatic to aliphatic mutations. For example, mutations of Ile \rightarrow Ala and Val have values of ΔG_{solv} in water for fully exposed side chains of only -0.21 and -0.16 kcal/mol, respectively (31). Thus, when $\Delta\Delta G_{\text{D-N}}$ is significant and ΔG_{reorg} and ΔG_{solv} are low, $\Phi_U = [\Delta G_{\text{noncov}(\text{TS}'\text{-TS})} - \Delta G_{\text{noncov}(N'-N)}] / [\Delta G_{\text{noncov}(\text{D}'\text{-D})} - \Delta G_{\text{noncov}(N'-N)}]$, which is analogous to α . Additionally, the energetics for deletion of hydrophobic elements of side chains is dominated by van der Waals' interactions, which are approximately additive, as is found experimentally (32). Thus, for nondisruptive deletions of hydrophobic side chains, $\Phi_F \approx (n_{\text{TS}} - n_{\text{D}}) / (n_{\text{N}} - n_{\text{D}})$, where n_{TS} is the number of van der Waals' interactions made by the target portion of the side chain in the transition state, n_{N} is in the native state, and n_{D} in the denatured state. For two-state kinetics, $\Phi_U \approx (n_{\text{TS}} - n_{\text{N}}) / (n_{\text{D}} - n_{\text{N}})$. Importantly, Φ reports back on the interactions made in the native protein.

Fractional values can arise from either genuinely weakened interactions in a single transition-state ensemble or from a mixture of states in parallel pathways. The folding of the barley CI2, for example, has predominantly fractional values of Φ (14, 33). The conforming to an REFER shows that there are not parallel pathways (34).

Fractional values may also arise if a side chain makes interactions with multiple elements of substructures that have varying degrees of structure formation. The overall Φ value is then the weighted mean value of all the interactions. This possibility was noted for barnase and CI2, and a fine structure analysis was performed by making systematic mutations in the side chains concerned: e.g., Ile \rightarrow Val \rightarrow Ala \rightarrow Gly, which gave the individual Φ values (9, 32, 35).

Accordingly, our preferred strategy for Φ -value analysis is to (i) mutate buried hydrophobic moieties by nondisruptive deletions (preferably Ile \rightarrow Val \rightarrow Ala \rightarrow Gly; Leu \rightarrow Ala \rightarrow Gly;

Thr \rightarrow Ser; and Phe \rightarrow Ala \rightarrow Gly, avoiding Phe \rightarrow Leu because of the change in stereochemistry); (ii) make a wider range of surface mutations, because larger side chains may be acceptable; (iii) use double-mutant cycles in which changes in solvation and reorganization energies tend to cancel out (4); and (iv) mutate Ala \rightarrow Gly at solvent-exposed positions in secondary structural regions ("Ala \rightarrow Gly scanning"), especially in α -helices, because they provide an exquisite probe of secondary structure (12). (v) Perform additional fine structure analyses by deleting different parts of larger side chains.

The magnitude of $\Delta\Delta G_{\text{D-N}}$ on mutation is a compromise: small values represent the smallest perturbation to the structure but have more attendant errors; larger values can be measured more accurately but often involve large changes that have more artifacts from reorganization energy changes on mutation and removing dispersed interactions. Our lower limit of acceptability for $\Delta\Delta G_{\text{D-N}}$ has generally been ≈ 0.6 kcal/mol (2.5 kJ/mol) for small, nondisruptive deletions (9, 14, 36).

Experimental Evidence for Lower Limit of $\Delta\Delta G_{\text{D-N}}$ for Φ Values.

Comparing individual Φ values for mutations at a particular site with a multipoint Leffler plot would be a good means of detecting deviations for particular mutations (22), but for the problem of the ΔG_{solv} terms involved in $\Delta G_{\text{D}'\text{-D}}$ and ΔG_{reorg} in the native state, unless carefully chosen, the different mutants will have different values of ΔG_{solv} and could have very different values of ΔG_{reorg} in the native and transition-state structures. Additionally, there must also be a relatively linear function of ΔG_{noncov} with formation of structure over the range of structural transition.

An earlier attempt to calibrate two-point Φ plots against a multipoint Leffler plot (37) used mutations that are specifically not recommended, with only 5 of 47 being nondisruptive hydrophobic deletions: Ser \rightarrow Asp, Glu; Ile \rightarrow Thr, Tyr; Val \rightarrow Ala, Lys; Leu \rightarrow Ala, Thr, Ile; Ala \rightarrow Cys, Pro, Ser, Gly, Thr, Leu, Asn; Gln \rightarrow Ala, Leu, Ser, Thr, Lys; Ser \rightarrow Ala, Gly, Trp, Cys, Thr, Ile, Tyr, Val, Asn, Gln, Lys; Val \rightarrow Ala, Cys, Leu, Thr; Leu \rightarrow Ser, Gln, Pro, Phe, Asn, Glu, Tyr; Leu \rightarrow Ala, Ser; and Ser \rightarrow Ala, Leu. Sanchez and Kiefhaber (22) showed that two-point Φ values deviated from a multipoint Leffler plot for $\Delta\Delta G_{\text{D-N}} < 1.7$ kcal/mol by using data from a study by Davidson and coworkers (28). However, the title of that study was "Protein Folding Kinetics Beyond the Φ -Value: Using Multiple Amino Acid Substitutions to Investigate the Structure of the SH3 Domain Folding Transition State," and the rationale was described by the authors in the abstract as: "In contrast to most other folding kinetic studies which have focused primarily on nondisruptive substitutions with Ala or Gly, here we have examined the effects of substitutions with diverse amino acid residues." They mutated Glu \rightarrow Asp, Gln, His, Lys, Ala, Ser, Val, Pro, Gly, Arg, Ile, Leu, and Ser \rightarrow Lys, Arg, Leu, Ala, His, Val, Ile, Asn, Asp, Gly, Phe, Tyr, which are virtually all disruptive mutations of a mainly buried hydrogen bond. These mutations are by intent and definition unsuitable for calibrating Φ -value analysis (28).

Leffler Plots and Ala \rightarrow Gly Scanning for Exposed Surface Regions.

The Davidson and coworkers (28) mutations are not suitable, because the side chains are buried, which complicates both the specific interactions involved for each chain as well as the changes in solvation on mutation. In contrast, surface-exposed residues, especially of α -helices, provide an opportunity for testing REFERS of more than two points. If the solvent-exposed end of the residue in the N state does not make any specific interactions with the rest of the protein, then it should make similar interactions with solvent in the TS and D states. Thus, the energetics of solvation of polar moieties of common surface residues such as Arg, Lys, Glu, Asp, Gln, and Asn, for

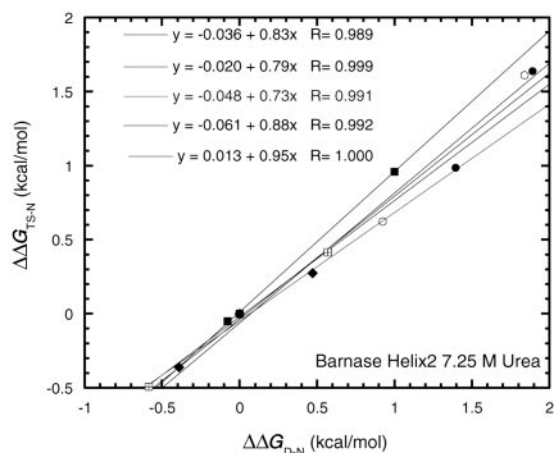


Fig. 2. Three-point Leffler plots for the unfolding of barnase. Mutations are as follows: Thr-26 → Ala → Gly, Lys-27 → Ala → Gly, Ser-28 → Ala → Gly, Glu-29 → Ala → Gly, and Gln-31 → Ala → Gly in helix 2.

example, will cancel out in the equations defining Φ . We have accumulated data for several helices in which surface-exposed residues are mutated to Ala and Gly. Helix 2 of barnase provides a good test for three-point Leffler plots, because the two-point Φ values show that the helix becomes highly unfolded in the transition state for unfolding (9) and thus the whole of the helix could constitute a multipoint REFER. Indeed, the energetics of mutation of Thr-26 → Ala → Gly, Lys-27 → Ala → Gly, Ser-28 → Ala → Gly, Glu-29 → Ala → Gly, and Gln-31 → Ala → Gly (Fig. 2) fit a good linear plot, with each three-point plot for an individual position having a correlation coefficient (R) between 0.99 and 1.0 and the slopes varying between 0.73 and 0.95 (mean = 0.84 ± 0.04 standard error). Individual values of Φ for all mutants (relative to wild type; Fig. 3) give a mean value of Φ of 0.86 ± 0.04 in a spread of 0.6 to 1.1. Ala → Gly scanning at each position gives a mean of 0.95 ± 0.08 . The Φ values were derived from values of $\Delta\Delta G_{D-N}$ that have a standard error of ± 0.06 kcal/mol and values of $\Delta\Delta G_{TS-N}$ at 7.25 M urea that have a standard error of ± 0.01 – 0.03 kcal/mol (12). Even with a $\Delta\Delta G_{D-N}$ value of 0.6 kcal/mol, the expected error in Φ should be only $\approx 10\%$. The linear plots in Fig. 2 and the values of Φ (Fig. 3) are nearly all obtained from values of $\Delta\Delta G_{D-N}$ below the 1.7 kcal/mol (7 kJ/mol) cutoff proposed by Sanchez and Kiefhaber (22).

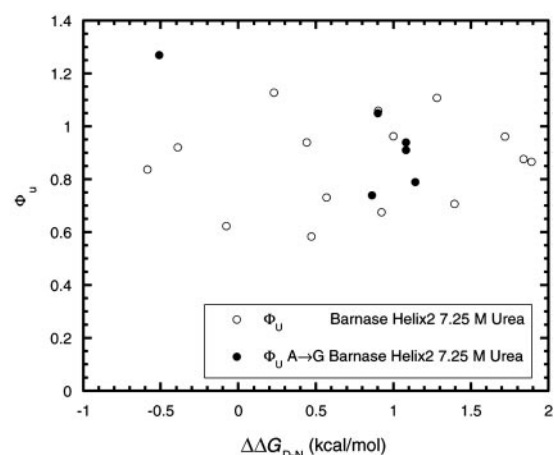


Fig. 3. Plot of Φ_U versus $\Delta\Delta G_{D-N}$ for mutations in helices 1 and 3 of barnase.

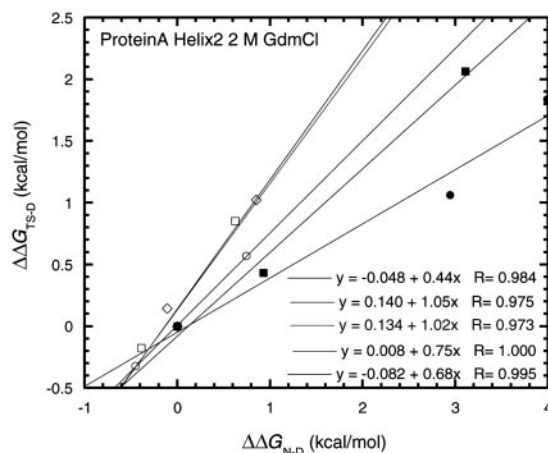


Fig. 4. Three-point Leffler plots for the following mutations of helix 2 of the B-domain of protein A: Gln-27 → Ala → Gly, Arg-28 → Ala → Gly, Asn-29 → Ala → Gly, Gln-33 → Ala → Gly, and Ser-34 → Ala → Gly.

Another suitable protein for testing the accuracy of values of Φ derived from low $\Delta\Delta G_{D-N}$ values is the B-domain of protein A from *Staphylococcus aureus*, a three-helix bundle protein. The data are slightly less accurate ($\Delta\Delta G_{D-N} \pm 0.1$ and $\Delta\Delta G_{TS-N} \pm 0.08$ kcal/mol), so that Φ for $\Delta\Delta G_{D-N} = 0.6$ kcal/mol should have an error of $\pm 20\%$, dropping to $\pm 10\%$ at $\Delta\Delta G_{D-N} = 1.2$ kcal/mol (36). Individual Φ values show that the first and third helices are in the process of being formed in the transition state, whereas the second is nearly fully formed. There are good three-point REFERs (Fig. 4) for the mutations in helix 2 of Gln-27 → Ala → Gly, Arg-28 → Ala → Gly, Asn-29 → Ala → Gly, Gln-33 → Ala → Gly, and Ser-34 → Ala → Gly. The only low value of the slope is for Gln-27 → Ala → Gly, which results from specific interactions. Gln-27 makes a hydrogen bond with Asn-24, and there are large changes of $\Delta\Delta G_{D-N}$ (3 and 4 kcal/mol) on its mutation. REFERs for the energetics of Ala → Gly scanning (Fig. 5) show that helix 2 is $\approx 80\%$ formed in the transition state, whereas the other two are not. The data for Ala → Gly scanning are clearly acceptable down to a $\Delta\Delta G_{D-N}$ value of ≈ 0.6 kcal/mol.

Schmid and coworkers (38), in a careful study of the folding of CspB protein, independently used a $\Delta\Delta G_{D-N}$ value of ≈ 0.6 kcal/mol as the lower limit for Φ -value analysis and also divided their results into weak, medium, and strong values. Radford and coworkers (25) used a cutoff of 0.7 kcal/mol for analyzing the

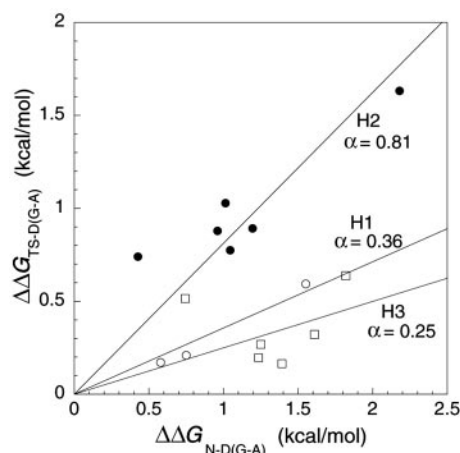


Fig. 5. Leffler plots of Ala → Gly scanning mutations in the three helices of the B-domain of protein A.

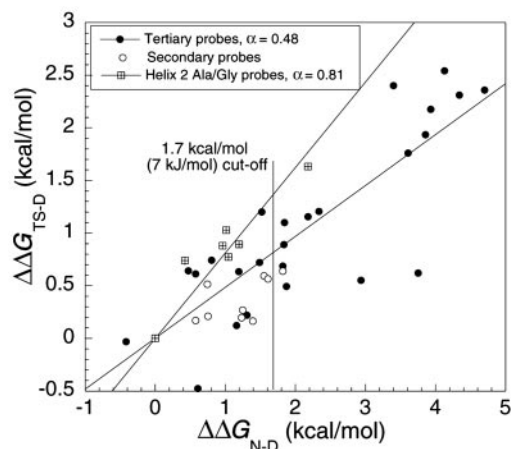


Fig. 6. Leffler plot for all mutations in the B-domain of protein A. Filled circles are used for the tertiary probes, and open circles are used for secondary structural probes.

folding of the immunity proteins Im7 and Im9, with satisfactory agreement. [Sanchez and Kiefhaber (22) recalculated Φ values from ref. 25, but the data were incomplete; several values from larger destabilizing mutations, some of which result in Φ values >0.3 , are omitted, and other values are shown that were not calculated in ref. 25 because the $\Delta\Delta G_{D-N}$ was below the cut off of 3 kJ/mol (0.7 kcal/mol) used (S. E. Radford, personal communication)].

Values of $\Delta\Delta G_{D-N}$ for Secondary and Tertiary Structure Probes. The fine structure probes that test specific interactions tend to be those that delete a small interaction; e.g., that of a single methylene group, or a single hydrogen bond, which both typically have $\Delta\Delta G_{D-N}$ values in the range of 1.5 ± 0.5 kcal/mol (39–41). Large changes of $\Delta\Delta G_{D-N}$ tend to be associated with the deletion of large side chains, especially in the hydrophobic core, or the disruption of buried salt bridges (32), as shown in Fig. 6 for the B-domain of protein A. By discarding the Φ values derived from a $\Delta\Delta G_{D-N}$ value of <1.7 kcal/mol, Sanchez and Kiefhaber (22) discarded most of the secondary structure probes and thus constructed REFER plots of mainly tertiary interactions, especially in the hydrophobic core. The formation of the core is always part of the rate-determining process and has fractional Φ values (3, 27, 36, 42, 43). Thus, by concentrating on such data, they concluded incorrectly that transition states are uniformly diffuse.

Movement of Transition-State Structure on Mutation. Sanchez and Kiefhaber (22, 29) claim that the structures of transition states do not change on mutation. They suggest that the observed movements of transition states along a reaction coordinate (10–12) are not a consequence of a change in transition-state structure via a Hammond effect but instead result from (partial) changes in rate-determining steps between formation and breakdown of intermediates or have complications from effects of mutation on the denatured state (44). It is very difficult to distinguish between true Hammond behavior and changes in the rate-determining step. However, there are well documented examples of anti-Hammond behavior (movement perpendicular to the reaction coordinate) that cannot be accounted for by changes in the rate-determining step along a reaction coordinate (12). A Leffler plot of successive mutations in helix 1 of barnase (Fig. 7) has a slope for unfolding of -0.09 for mutations with $\Delta\Delta G_{D-N} < 2$ kcal/mol, showing that it is $\approx 90\%$ folded in the transition state, but for $\Delta\Delta G_{D-N} > 3$ kcal/mol, the slope gets

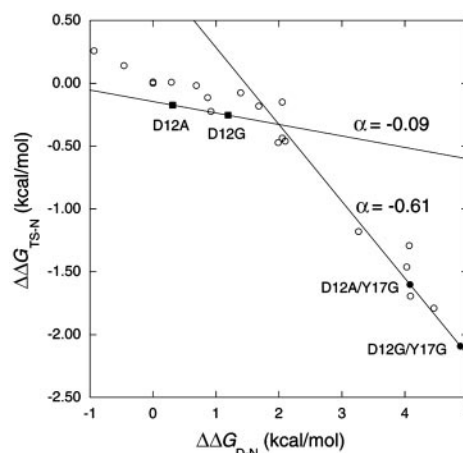


Fig. 7. Leffler plots for single, double, and triple mutations in helix 1 of barnase plus Ala \rightarrow Gly scanning at position 12 in the mutant Tyr-17 \rightarrow Gly. The mutants are as follows: Asp-8 \rightarrow Ala, Asp-12 \rightarrow Gly, Asp-12 \rightarrow Ala, Tyr-13 \rightarrow Ala, Tyr-13 \rightarrow Ala/Thr-16 \rightarrow Ser, Tyr-13 \rightarrow Ala/Tyr-17 \rightarrow Ala, Tyr-13 \rightarrow Ala/Thr-16 \rightarrow Ser/Tyr-17 \rightarrow Ala, Gln-15 \rightarrow Ile, Thr-16 \rightarrow Ala, Thr-16 \rightarrow Gly, Thr-16 \rightarrow Ser, Thr-16 \rightarrow Ser/Tyr-17 \rightarrow Ala, Thr-16 \rightarrow Arg, Tyr-17 \rightarrow Ala, His-18 \rightarrow Lys, His-18 \rightarrow Gln, Tyr-17 \rightarrow Ala, Tyr-17 \rightarrow Gly, His-18 \rightarrow Ala, His-18 \rightarrow Gly, Asp-12 \rightarrow Ala/Tyr-17 \rightarrow Gly, and Asp-12 \rightarrow Gly/Tyr-17 \rightarrow Gly.

steeper at -0.61 , indicating it is only 40% folded and follows anti-Hammond behavior (12). The slopes are measured by Ala \rightarrow Gly scanning at position 12 in wild-type protein and the same position in the mutant Tyr-17 \rightarrow Gly, which is destabilized by 4.1 kcal/mol (12). Whereas the helix becomes less folded in the transition state on destabilization, the overall transition state follows Hammond and becomes more folded, with its relative surface exposure decreasing from 55% to 37% over a change of 5 kcal/mol in $\Delta\Delta G_{D-N}$ (Fig. 8). The data are for unfolding and are independent of mutations on the denatured state. The anti-Hammond behavior can be explained by either a gradual movement of the transition state or by a switch between parallel pathways (12). Simulation favors the gradual movement (45).

Sanchez and Kiefhaber propose that the larger the value of

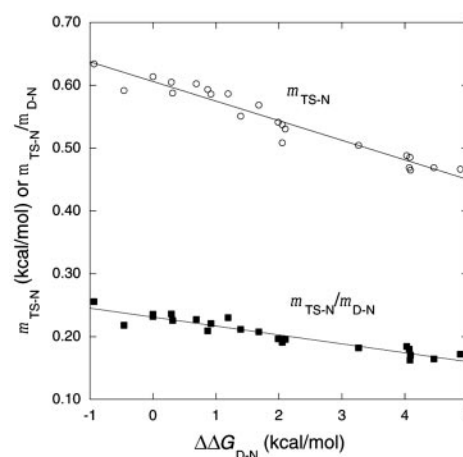


Fig. 8. Plots of data relating surface exposure on denaturation kinetics of barnase. m_{TS-N} is the slope of the plot of ΔG_{TS-N} versus [urea], and m_{D-N} is the slope of the plot of ΔG_{D-N} versus [urea]. m_{TS-N} is the value at 7.25 M urea, measured for unfolding data acquired between 6 and 8.5 M urea, and it is accurate to $\pm 2\%$. The ratio of m_{TS-N}/m_{D-N} is a measure of the relative solvent exposure of the transition state to the denatured state. The value of m_{TS-N} is a function of just the difference in solvent-accessible surface area of TS and N and does not depend on the properties of the denatured state.

$\Delta\Delta G_{D-N}$, the better for Φ -value analysis (22). However, Fig. 7 shows clearly that large changes of $\Delta\Delta G_{D-N}$ can lead to radical changes in structure in transition states. In general, the fine structural information requires specific probes, with energies of 0.6–2 kcal/mol. The more energy-disruptive probes can perturb the transition-state structure, which complicates a simple Φ -value analysis but can give information about the energy surface around the transition state.

Nature of Folding Transition States. Sanchez and Kiefhaber (22) use the classification of transition states being either diffuse, whereby most of the Φ values are fractional and polarized where there are regions that are fully formed or fully denatured. They suggest that there are never regions of high Φ value in the diffuse states, and thus all transition states are similar to that found for CI2. However, the transition state for the folding of the B-domain of protein A is not polarized, and there are regions, especially involving helix 2, that have Φ values approaching 1 (36). The Φ analyses of the Engrailed homeodomain family, although not extensive, show transition states that are basically structured all over and with regions of Φ values of 1 (27, 43). Transition states have a spectrum of structures, varying from the diffuse of pure nucleation–condensation to the more compact that approach the classical framework mechanism of folding, in which the repeating secondary structure is nearly fully formed and the core is in the process of consolidation (23–25, 38, 44–50).

In any case, it is not possible to divine from the transition-state structure *per se* whether there are nucleation sites, because the

structure does not reveal *per se* the starting point of folding or the route by which the structure is formed (51). The experimental evidence for nucleation in CI2 folding, for example, came from ancillary studies that examined the denatured state of the protein and its fragments (52), and the evidence for framework for Engrailed homeodomain came from analyzing the structures of ground states as well as simulation (43). Additionally, high Φ values need not be associated with a nucleus, and low Φ values can be found in nuclei (19).

Conclusions

There are strong analogies between the determination of solution structures of proteins by NMR combined with simulated annealing and the determination of structures of transition states by Φ values and simulation. Just as there are nuclear Overhauser effects in NMR spectra of spurious intensity, there are undoubtedly some misleading Φ values, especially when $\Delta\Delta G_{D-N}$ is small. However, provided mutations are made within the prescribed rules and that a sufficient number are analyzed, then reliable results will be obtained down to changes in $\Delta\Delta G_{D-N}$ of ≈ 0.6 kcal/mol under optimal conditions. Higher values of $\Delta\Delta G_{D-N}$ do give statistically more precise data, but much larger values of $\Delta\Delta G_{D-N}$ may give less precise information, because they arise from dispersed interactions and may have higher contributions from perturbations of structure. Just as special methods are continually introduced to refine NMR methods, so ancillary methods such as Ala \rightarrow Gly scanning are required for refining by Φ -value analysis.

- Fersht, A. R., Leatherbarrow, R. J. & Wells, T. N. C. (1986) *Nature* **322**, 284–286.
- Fersht, A. R., Leatherbarrow, R. & Wells, T. N. C. (1987) *Biochemistry* **26**, 6030–6038.
- Matouschek, A., Kellis, J. T., Jr., Serrano, L. & Fersht, A. R. (1989) *Nature* **340**, 122–126.
- Fersht, A. R., Matouschek, A. & Serrano, L. (1992) *J. Mol. Biol.* **224**, 771–782.
- Fersht, A. R. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 10869–10873.
- Leffler, J. E. (1953) *Science* **117**, 340–341.
- Fersht, A. R. (1988) *Biochemistry* **27**, 1577–1580.
- Matouschek, A., Kellis, J. T., Jr., Serrano, L., Bycroft, M. & Fersht, A. R. (1990) *Nature* **346**, 440–445.
- Serrano, L., Matouschek, A. & Fersht, A. R. (1992) *J. Mol. Biol.* **224**, 805–818.
- Matouschek, A. & Fersht, A. R. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 7814–7818.
- Matouschek, A., Otzen, D. E., Itzhaki, L. S., Jackson, S. E. & Fersht, A. R. (1995) *Biochemistry* **34**, 13656–13662.
- Matthews, J. M. & Fersht, A. R. (1995) *Biochemistry* **34**, 6805–6814.
- Fersht, A. R. (1995) *Curr. Opin. Struct. Biol.* **5**, 79–84.
- Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995) *J. Mol. Biol.* **254**, 260–288.
- Fersht, A. R. (1997) *Curr. Opin. Struct. Biol.* **7**, 3–9.
- Fersht, A. R. & Daggett, V. (2002) *Cell* **108**, 573–582.
- Paci, E., Vendruscolo, M., Dobson, C. M. & Karplus, M. (2002) *J. Mol. Biol.* **324**, 151–163.
- Klimov, D. K. & Thirumalai, D. (2002) *J. Mol. Biol.* **317**, 721–737.
- Hubner, I. A., Shimada, J. & Shakhnovich, E. I. (2004) *J. Mol. Biol.* **336**, 745–761.
- Weikl, T. R., Palassini, M. & Dill, K. A. (2004) *Protein Sci.* **13**, 822–829.
- Onuchic, J. N. & Wolynes, P. G. (2004) *Curr. Opin. Struct. Biol.* **14**, 70–75.
- Sanchez, I. E. & Kiefhaber, T. (2003) *J. Mol. Biol.* **334**, 1077–1085.
- Daggett, V. & Fersht, A. (2003) *Nat. Rev. Mol. Cell Biol.* **4**, 497–502.
- Daggett, V. & Fersht, A. R. (2003) *Trends Biochem. Sci.* **28**, 18–25.
- Friel, C. T., Capaldi, A. P. & Radford, S. E. (2003) *J. Mol. Biol.* **326**, 293–305.
- Li, L. & Shakhnovich, E. I. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 13014–13018.
- Gianni, S., Guydosh, N. R., Khan, F., Caldas, T. D., Mayor, U., White, G. W. N., DeMarco, M. L., Daggett, V. & Fersht, A. R. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 13286–13291.
- Northey, J. G. B., Maxwell, K. L. & Davidson, A. R. (2002) *J. Mol. Biol.* **320**, 389–402.
- Sanchez, I. E. & Kiefhaber, T. (2003) *J. Mol. Biol.* **327**, 867–884.
- Buchner, J. & Kiefhaber, T. (1990) *Nature* **343**, 601–602.
- Wolfenden, R., Anderson, L., Cullis, P. M. & Southgate, C. C. B. (1981) *Biochemistry* **20**, 849–855.
- Serrano, L., Kellis, J. T., Cann, P., Matouschek, A. & Fersht, A. R. (1992) *J. Mol. Biol.* **224**, 783–804.
- Otzen, D. E., Itzhaki, L. S., Elmasry, N. F., Jackson, S. E. & Fersht, A. R. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 10422–10425.
- Fersht, A. R., Itzhaki, L. S., Elmasry, N., Matthews, J. M. & Otzen, D. E. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 10426–10429.
- Matouschek, A., Serrano, L. & Fersht, A. R. (1992) *J. Mol. Biol.* **224**, 819–835.
- Sato, S., Religa, T. E. & Fersht, A. R. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 6952–6956.
- Cymes, G. D., Grosman, C. & Auerbach, A. (2002) *Biochemistry* **41**, 5548–5555.
- Garcia-Mira, M. M., Bohringer, D. & Schmid, F. X. (2004) *J. Mol. Biol.*, in press.
- Fersht, A. R., Shi, J. P., Knill-Jones, J., Lowe, D. M., Wilkinson, A. J., Blow, D. M., Brick, P., Carter, P., Wayne, M. M. Y. & Winter, G. (1985) *Nature* **314**, 235–238.
- Kellis, J. T. J., Nyberg, K. & Fersht, A. R. (1989) *Biochemistry* **28**, 4914–4922.
- Serrano, L., Sancho, J., Hirshberg, M. & Fersht, A. R. (1992) *J. Mol. Biol.* **227**, 544–559.
- Jackson, S. E., Elmasry, N. & Fersht, A. R. (1993) *Biochemistry* **32**, 11270–11278.
- Mayor, U., Guydosh, N. R., Johnson, C. M., Grossmann, J. G., Sato, S., Jas, G. S., Freund, S. M. V., Alonso, D. O. V., Daggett, V. & Fersht, A. R. (2003) *Nature* **421**, 863–867.
- Sanchez, I. E. & Kiefhaber, T. (2003) *J. Mol. Biol.* **325**, 367–376.
- Daggett, V., Li, A. J. & Fersht, A. R. (1998) *J. Am. Chem. Soc.* **120**, 12740–12754.
- Kragelund, B. B., Osmark, P., Neergaard, T. B., Schiodt, J., Kristiansen, K., Knudsen, J. & Poulsen, F. M. (1999) *Nat. Struct. Biol.* **6**, 594–601.
- Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M. & Dobson, C. M. (1999) *Nat. Struct. Biol.* **6**, 1005–1009.
- Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I. & Baker, D. (1999) *Nat. Struct. Biol.* **6**, 1016–1024.
- Martinez, J. C. & Serrano, L. (1999) *Nat. Struct. Biol.* **6**, 1010–1016.
- Lindberg, M., Tangrot, J. & Oliveberg, M. (2002) *Nat. Struct. Biol.* **9**, 818–822.
- Fersht, A. R. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 1525–1529.
- Itzhaki, L. S., Neira, J. L., Ruiz-Sanz, J., Gay, G. D. & Fersht, A. R. (1995) *J. Mol. Biol.* **254**, 289–304.