

Review

Cite this article: Fersht AR (2024). From covalent transition states in chemistry to noncovalent in biology: from β - to Φ -value analysis of protein folding. *Quarterly Reviews of Biophysics*, 57, e4, 1–28
<https://doi.org/10.1017/S0033583523000045>

Received: 26 November 2023
Accepted: 21 December 2023

Keywords:

folding; free-energy relationships; mutagenesis; protein engineering; structure-activity relationships

Email: arf25@cam.ac.uk

From covalent transition states in chemistry to noncovalent in biology: from β - to Φ -value analysis of protein folding

Alan R. Fersht^{1,2,3} 

¹MRC Laboratory of Molecular Biology, Cambridge, UK; ²Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, UK and ³Gonville and Caius College, University of Cambridge, Cambridge, UK

Abstract

Solving the mechanism of a chemical reaction requires determining the structures of all the ground states on the pathway and the elusive transition states linking them. 2024 is the centenary of Brønsted's landmark paper that introduced the β -value and structure-activity studies as the only experimental means to infer the structures of transition states. It involves making systematic small changes in the covalent structure of the reactants and analysing changes in activation and equilibrium-free energies. Protein engineering was introduced for an analogous procedure, Φ -value analysis, to analyse the noncovalent interactions in proteins central to biological chemistry. The methodology was developed first by analysing noncovalent interactions in transition states in enzyme catalysis. The mature procedure was then applied to study transition states in the pathway of protein folding – 'part (b) of the protein folding problem'. This review describes the development of Φ -value analysis of transition states and compares and contrasts the interpretation of β - and Φ -values and their limitations. Φ -analysis afforded the first description of transition states in protein folding at the level of individual residues. It revealed the nucleation-condensation folding mechanism of protein domains with the transition state as an expanded, distorted native structure, containing little fully formed secondary structure but many weak tertiary interactions. A spectrum of transition states with various degrees of structural polarisation was then uncovered that spanned from nucleation-condensation to the framework mechanism of fully formed secondary structure. Φ -analysis revealed how movement of the expanded transition state on an energy landscape accommodates the transition from framework to nucleation-condensation mechanisms with a malleability of structure as a unifying feature of folding mechanisms. Such movement follows the rubric of analysis of classical covalent chemical mechanisms that began with Brønsted. Φ -values are used to benchmark computer simulation, and Φ and simulation combine to describe folding pathways at atomic resolution.

Introduction

I have been fascinated with transition states for more than 60 years – a passion for understanding structure and mechanism which has directed my research at the borderlines of chemistry, physics and biology. Transition states of simple covalent reactions are traditionally studied by *structure-activity relationships* whereby perturbations of the energetics of kinetics and equilibria of reactions on small changes in the structure of reagents are correlated to give clues about the structure of the transition state. Much of biological chemistry is dominated by weak noncovalent interactions, especially those of proteins. The advent of protein engineering enabled structure-activity relationships to be applied to the noncovalent transition states of those biological processes. This invited review outlines the history of key steps by my research group and by others in translating those structure-activity methods of classical physical and organic chemistry to analyse noncovalent transition states. It begins with their introduction via protein engineering to the quantitative study of noncovalent interactions in enzyme catalysis and specificity and then their extension to protein folding to give Φ -value analysis. I discuss in particular how the combination of those methods and computer simulation has been used in solving problems of protein folding pathways.

It is a particularly appropriate time for this topic as it is the centenary of the publication of the landmark paper in the history of physical-organic chemistry that led to structure-activity studies, the discovery of general-base catalysis and its dependence on the strength of the base by Brønsted and Pedersen (1924). That discovery and the ensuing Brønsted β -value have inspired much of my research and the contents of this review. It is also the half-centenary of my paper that sent me down the slippery slope of analysing non-covalent interactions in transition states (Fersht, 1974). Pertinent also it is the centenary of the chess grandmaster S. G. Tartakower's *Die Hypermoderne Schachpartie* in which he wrote *Die Fehler sind dazu da, um gemacht zu werden* (Tartakower, 1924, p. 90). The usual translation 'The mistakes are all there, waiting to be made' should be the

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

watchword of every experimentalist and theoretician as well as chess player, especially in areas as complex and with pitfalls as protein folding.

Transition states in covalent chemistry

Transition states are the transient structures at the peaks of plots of free energy as a reaction progresses as opposed to intermediates that are in a basin (Figure 1). Simple transition state theory relates the rate constant for a reaction to the energy difference between the transition and ground states, ΔG^\ddagger , as if the two states were in equilibrium: the rate constant for the reaction going through the transition state, k , is given by:

$$k = \kappa \left(\frac{k_B T}{h} \right) \exp \left(- \frac{\Delta G^\ddagger}{RT} \right), \quad (1)$$

where: k_B is the Boltzmann, h is the Planck, R is the gas constants, T is the temperature, and κ is a transmission coefficient (Pelzer and Wigner, 1932; Evans and Polanyi, 1935; Eyring, 1935). Examination of the transition state structure relative to the ground states gives important clues as to what drives a reaction and how its rate or even its products may change by altering the structure of the reagents, the reaction conditions or employing catalysts. For example, the rate of attack of a negatively charged nucleophile on a reagent can be increased by introducing electron-withdrawing substituents. Transition states are essential structures in defining reaction pathways. To solve a reaction pathway, we must characterise all the ground states and the transition states linking them. Ground states and intermediates are best studied by direct observation. The only state between ground states that can be characterised experimentally is the elusive transition state and the only current experimental means is by using indirect evidence from structure-reactivity relationships.

Linear-free-energy relationships: LFER and REFERs – β - and α -values

The classical physical-organic chemist's approach to analysing the structure of a transition state of a reaction is to use quantitative measurements of the changes in reactivity and equilibria on small changes in the structure of reagents. For example, Brønsted and Pedersen began the analysis of the effects of strengths of bases and acids on their powers of catalysis of simple organic reactions in

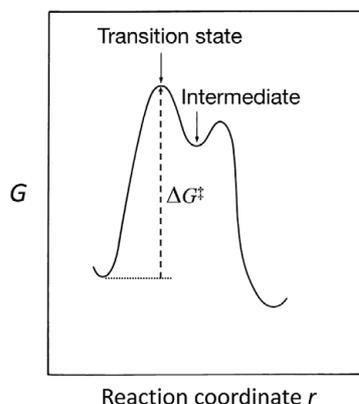


Figure 1. Transition state is at a maximum for free energy, G , versus reaction coordinate, r .

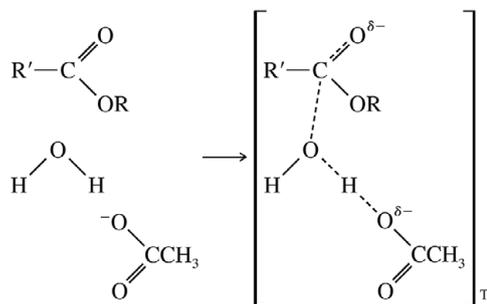


Figure 2. Transition state for the general-base-catalysed attack of water on an ester.

solution (Brønsted and Pedersen, 1924). They found, for example, that there is often a simple equation relating the second-order rate constant (k_2) for catalysis of a reaction by a general base to the pK_a of the conjugate acid (2).

$$\log k_2 = A + \beta pK_a. \quad (2)$$

This is an example of a *linear-free-energy relationship* (LFER) since it is equivalent to:

$$\Delta G^\ddagger = A' + \beta \Delta G_0, \quad (3)$$

where ΔG^\ddagger is the free energy of activation and ΔG^0 the equilibrium free energy change of a process. β is for base but we usually call it the Brønsted β . The equation can be formulated for a wider range of reactions, $\Delta G^\ddagger = A + \alpha \Delta G^0$, as described by Leffler (1953), and the description *rate-equilibrium-energy relationship* (REFER) alternatively used.

L. P. Hammett translated these LFERs to chemical reactions involving aromatic compounds by measuring the effects of chemical substituents in the *meta* and *para* positions of benzoic acid on its pK_a to assign a σ -value for each substituent (corresponding to the change it makes in the pK_a) and relating the sensitivity of the logarithms of rate constants for chemical reactions to σ by a parameter ρ , equivalent to the Brønsted β (Hammett, 1937). The *meta* and *para* positions are chosen to minimise direct steric interactions with the seat of reaction (Hammett, 1940).

The simple reasoning behind the magnitude of the β and ρ values in many chemical reactions is that they often result from electrostatic effects. For example, in the transition state of the general-base-catalysed attack by acetate ion of H_2O on an ester (Figure 2), an H^+ is in the process of being transferred from the H_2O to the $-CO_2^-$ catalyst, partly neutralising its negative charge. If a substituent that has an electron-withdrawing or donating propensity is put into the $-CH_3$ of acetic acid, it will perturb its pK_a by $\Delta \Delta G^0$ because of the electrostatic interactions with the negatively charged carboxylate relative to the neutral state. The electrostatic interaction of the substituent with the partly neutralised negative charge on the $-CO_2^-$ in the transition state, $\Delta \Delta G^\ddagger$, will be less than $\Delta \Delta G^0$ because of the H^+ being transferred so that:

$$\beta = \frac{\Delta \Delta G^\ddagger}{\Delta \Delta G^0}, \quad (4)$$

where β approximates to the extent of bond formation with the H^+ in the example of Eq. (4) or in other cases a covalent bond in the transition state. $\beta = 0$ means there is no transfer of the proton to the base and $\beta = 1$ means complete transfer, and fractional values are something in between. One possible generic basis of LFERs is explained in Figure 3 where the reagents are in two energy wells that

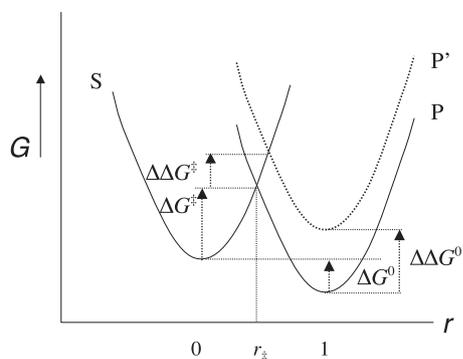


Figure 3. Illustration of one type of origin for a LFER. In the plot of G versus reaction coordinate, r , the energy function of the starting material S crosses that of the products P at the transition state. To an approximation, if the structure and energetics are perturbed such the energy of P is increased relatively by $\Delta\Delta G^0$ to S , the energy of the transition state will be increased by a value of $\Delta\Delta G^\ddagger$ that is less than $\Delta\Delta G^0$ and determined by the angles and so forth at the point of intersection. Apart from the extreme values of the position of the transition $r^\ddagger = 0$ or 1 , r^\ddagger does not generally = $\Delta\Delta G^\ddagger/\Delta\Delta G^0$ that is, $\neq \alpha$ or β (Fersht, 2004b). The small change in r^\ddagger with changes in energetics is the basis of the Hammond Postulate (Hammond, 1955) whereby as the energy of the high energy state increases, the transition state structure moves closer to it.

intersect at the transition state. Applying a simplified version of the treatment by Marcus (1968) of outer sphere electron transfer reactions, I assume the energy functions are simple harmonic wells. For the starting material S , $\Delta G_S = \lambda_1 r^2$ and for products $\Delta G_P = \lambda_2 (1 - r)^2 - \Delta G^0$, which gives for $\alpha = \Delta\Delta G^\ddagger/\Delta\Delta G^0$ (Fersht, 2004b):

$$\alpha = \lambda_1 r_\ddagger \left((\lambda_1 - \lambda_2) \Delta G^0 + \lambda_1 \lambda_2 \right)^{-1/2}. \quad (5)$$

For the special case of $\lambda_1 = \lambda_2$, $\alpha = r_\ddagger$. But, apart from the extreme values of the position of the transition state $r_\ddagger = 0$ or 1 , r_\ddagger does not generally = α (or β) (Fersht, 2004b). The situation is, of course, even more complicated than the above for fractional values. The reaction coordinate diagram is not two-dimensional and there can be movement in other dimensions with much complexity (Jencks, 1985).

LFERs have been found in many types of physical-chemical processes, and the interpretation is usually simply phenomenological with the value of α or β interpreted only qualitatively for mechanism and semi-quantitatively for predictive purposes. In the qualitative analysis of the effects of changes of structure on reactivity, it is just the changes in ΔG^\ddagger in the $\kappa(k_B T/h) \exp(-\Delta G^\ddagger/RT)$ term of the transition state theory that are examined, and the pre-exponential component cancelling out in the comparison of rate constants. $\Delta\Delta G^\ddagger$ and $\Delta\Delta G^0$ are the key quantities. My first published paper as a graduate student centred on using LFERs to analyse transition states in chemical mechanisms, with a series of substituted aspirins (Fersht and Kirby, 1967), and LFERs figure prominently in my textbook on enzymes (Fersht, 1977, 1985).

Transition states in noncovalent chemistry: biological catalysis and specificity

Classical chemistry is dominated by covalent bonds and strong ionic interactions. Much of chemistry in biology, on the other hand, is dominated by weak noncovalent interactions, such as van der Waals interactions, hydrogen bonds, salt bridges, and the hydrophobic

effect. Utilisation of these weak interactions is the hallmark of biological specificity in general and modulation of catalysis by enzymes.

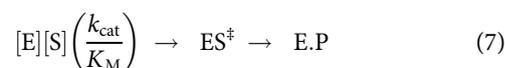
Enzyme catalysis and binding of the transition state

The rates of enzyme-catalysed reactions are many orders of magnitude greater than simple reactions catalysed in solution by acids and bases or nucleophiles. To answer why, Haldane proposed that enzymes might catalyse reactions by straining the structures of the substrates towards that of the products (Haldane, 1930). Pauling refined that concept by stating that an enzyme could have a structure complementary to that of the activated complex or transition state of the substrate, and hence stabilise it (Pauling, 1948). Classical studies varying the structures of substrates of α -chymotrypsin, for example, showed that binding energy could be distributed between tighter binding of substrate and higher rate constants (Jencks, 1975). Analogues mimicking the structure of transition states of substrates may also bind more tightly than the substrates themselves (Schramm, 1998). So, free energies of activation of the covalent chemical reaction, ΔG_{cov}^\ddagger , can be modulated by changes in binding energies, $\Delta G_{noncov}^\ddagger$.

The Michaelis–Menten equation (6) relates the reaction rate v of a substrate S to the total concentration of enzyme, $[E]_0$, an apparent first-order rate constant k_{cat} , and an apparent dissociation constant K_M .

$$v = k_{cat} \frac{[E]_0 [S]}{([S] + K_M)}. \quad (6)$$

In the simplest case, K_M is the dissociation constant for the E.S complex, K_S , and k_{cat} is the rate constant for its giving products. But, these apparent rate and equilibrium constants can hide a complexity of additional terms, from additional chemical steps to non-productive binding. Crucially, however, the ratio k_{cat}/K_M is an apparent second-order rate constant for the process of free enzyme, $[E]$, and free substrate, $[S]$ proceeding to the highest transition state on the reaction pathway to give products, and complicating factors are usually cancelled in the ratio k_{cat}/K_M , Eq. (7).



Applying simple transition state theory suggests two notional processes in the evolution of maximal rate (Fersht, 1974). The enzyme evolves to have a structure that is complementary to that of the transition state of the reaction, which maximises the value of k_{cat}/K_M . And, if rate is the prime concern, the enzyme will also evolve to increase K_M at constant k_{cat}/K_M until the K_M is higher than the physiological substrate concentration. This is because low-energy intermediates can be thermodynamic pits where there is a higher ΔG^\ddagger going from them to the transition state than there is from the initial state. The strain theories of Haldane and Pauling propose strong binding of the transition state and concomitant weak binding of the substrate, and the highest catalysis occurs when the binding energy in the E.S complex is sufficiently weak such that it is the complex is largely dissociated and intermediates do not accumulate on reaction pathways (Fersht, 1974).

Specificity depends on the relative binding of transition states

When two substrates A and B are competing for the active site of an enzyme, their relative rate of reaction at all concentrations of free $[A]$ and $[B]$ is given by (Fersht, 1974):

$$v_A/v_B = (k_{\text{cat}}/K_M)_A[A]/(k_{\text{cat}}/K_M)_B[B]. \quad (8)$$

As k_{cat}/K_M is for the process of unbound enzyme and unbound substrate proceeding to the transition state ES^\ddagger , the specificity is independent of the interactions in the enzyme-substrate complex and depends only on the relative binding of transition states. Accordingly, both the magnitude and specificity of enzyme catalysis depends upon the binding of transition states.

Equation (8) is very useful for measuring the apparent contributions to binding energy of parts of substrates by comparing modified versions of them. For example, a substrate, containing a particular radical can be compared with the substrate modified to have, say, an -H replacing that radical to give an empirical measure of the energetics of binding of that radical. The aminoacyl-tRNA synthetases have evolved to maximise the specificity of competing amino acids, for example, the isoleucyl-tRNA synthetase with isoleucine versus valine. We measured ratios of k_{cat}/K_M for cognate versus non-cognate amino acids with different aminoacyl-tRNA synthetases to explore the upper limits of binding energies under evolutionary pressure (Fersht, 1981).

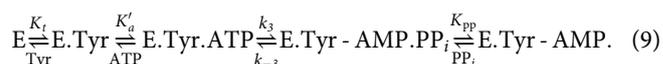
Noncovalent interactions in enzyme transition states: LFER analysis

We would like to know how the structures of proteins change in the transition states of biological processes and how it contributes to them. The way experimentally to characterise those details by analogy with covalent chemistry is by using similar systematic structure-reactivity relationships, which is something I had been wanting to do since starting in enzymology. The introduction of site-directed mutagenesis at the end of the 1970s to revert mutants of bacteriophage ϕ X174 (Hutchison *et al.*, 1978) made this possible and laid open the new field of protein engineering, which was left largely unploughed for 4 or 5 years.

The initial paradigm: protein engineering the tyrosyl-tRNA synthetase

Gregory Winter and I began a collaboration and published the first paper on protein engineering studies on a protein of known structure (Winter *et al.*, 1982). It may seem surprising that the practical application of the mutagenesis technology of the 1978 paper (Hutchison *et al.*, 1978) took so long. Site-directed mutagenesis was then very difficult to do on the genes of recombinant proteins; the necessary oligonucleotides were not commercially available; only a few protein chemists were using recombinant DNA technology; and some did not believe that site-directed mutagenesis was anything more than a new form of chemical modification (reported by Bryan, 2000). I spent a sabbatical in 1978–1979 in Arthur Kornberg's laboratory to learn recombinant DNA technology and worked on reverting mutants of ϕ X174 to study the fidelity of DNA replication (Fersht, 1979). Gregory Winter had sequenced the genes of aminoacyl-tRNA synthetases, and we chose to do protein engineering of the tyrosyl-tRNA synthetase from *Bacillus stearothermophilus*. His goal was to use it as an entry into making novel proteins, paralleling synthetic organic chemistry, and he subsequently pioneered antibody engineering. My goal was to use it for structure-activity studies to understand the chemistry of noncovalent interactions in biology, paralleling physical-organic chemistry. This thermophilic enzyme is an exceptional paradigm for this latter purpose: it may be expressed in *Escherichia coli*, and any activity

of contaminating mesophilic enzyme that could obscure steady-state kinetics removed by heating; it is amenable to study by pre-steady kinetics so intermediates can be directly observed; and as a bonus, it is an enzyme whose chemical pathway was known but nothing about what groups were involved in catalysis. The first step in the aminoacylation of tRNA is the nucleophilic attack of the carboxylate of the amino acid on the α -phosphate of ATP to generate an enzyme-bound aminoacyl-adenylate, which subsequently transfers the tyrosine to its cognate tRNA (9).



Tyrosyl-adenylate is highly reactive in solution but is sequestered and stable in the complex with the enzyme in the absence of tRNA. The crystal structure of the complex reveals a large number of protein side chains binding the intermediate, principally by making hydrogen bonds.

The strategy for structure-activity studies of transition states of proteins

The fundamental strategy for structure-activity studies is simple and taken straight from classical chemistry: make small rational changes in structure and measure the changes in the equilibrium free energies and activation free energies of the chemical steps. Here, the steps are: (1) truncate the side chains that are hydrogen bond donors or acceptors with the substrate to give quantitative information on the effective strengths and to provide the $\Delta\Delta G^0$ terms for the application of LFERs; and (2) do kinetics on mutants to measure the corresponding $\Delta\Delta G^\ddagger$ values. Step 1 is useful in general *per se* as it provides empirical quantitative data on biological interactions. The same strategy is applied analogously to other processes such as protein folding.

The first experiments measured the strengths of hydrogen bonds using Eq. (8) and the ratios of k_{cat}/K_M from steady-state kinetics for wild-type and mutants. The apparent energies spanned 0.5–1.5 kcal/mol (Fersht *et al.*, 1985). I usually refer to these as *apparent* binding energies because they measure the relative binding energies that are found in practice but not absolute energies – all binding reactions in water represent an exchange reaction with H_2O of solvation (Fersht *et al.*, 1985). In general, energies from mutagenesis experiments have complex components, which I have emphasised from the start, but sometimes overlooked (Fersht, 1987, 1988).

LFER analysis uncovers a novel enzyme mechanism just involving binding energy

The second step of the strategy was to determine $\Delta\Delta G^\ddagger$ and $\Delta\Delta G^0$ for individual steps in Eq. (9) using rapid reaction pre-steady state kinetics (Wells and Fersht, 1985). There is a progressive increase in the apparent binding energy of the hydrogen bonds, as illustrated in Figure 4 where Cys-35 and His-48 are truncated to Gly, and the energies of the mutant compared with wild-type plotted. These progressive curves were described in terms of *difference energies* (Wells and Fersht, 1986). Subsequently, the ratio of $\Delta\Delta G^\ddagger/\Delta\Delta G^0$ was used and called a β -value, in homage to Brønsted (Fersht *et al.*, 1987). This is effectively a series of two-point LFERs around the substrate for each interaction from a side chain. As seen in Figure 4, mutation of side chains that bind the sugar ring of ATP hardly weakens the binding of ATP in the E.Tyr.ATP complex but

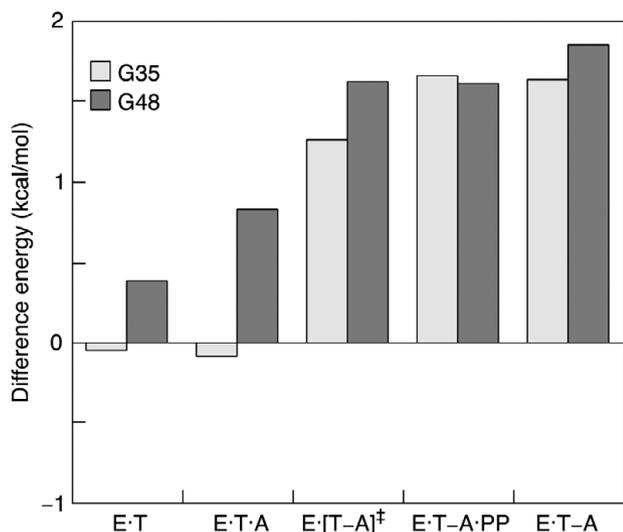


Figure 4. Difference energy plot for mutations of side chains of the tyrosyl-tRNA synthetase. The values of $\Delta\Delta G_{mut-wt}$ (mutant – wild type) for the ΔG of binding Tyr, ATP, [T-A][‡], T-A.PPi and T-A in the formation of tyrosyl-adenylate (Eq. (9)) on mutation of residues Cys35 and His48 (data from Wells and Fersht, 1986; Fersht *et al.*, 1987).

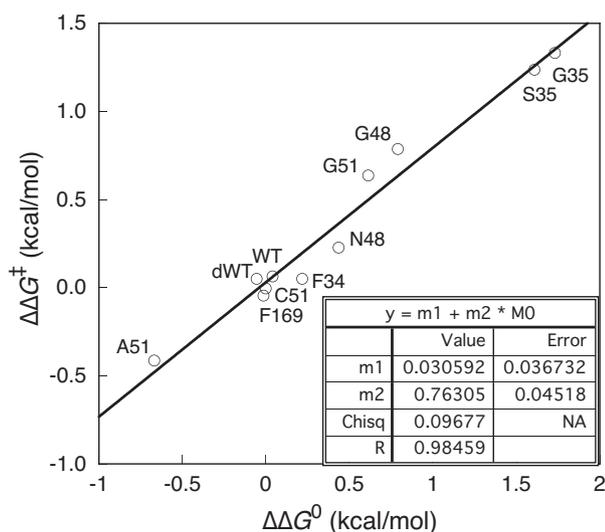


Figure 5. A linear free energy relationship for the reaction E.Tyr.ATP → E.Tyr.ATP.PPi of the tyrosyl-tRNA synthetase (k_3 and k_3/k_{-3} in Eq. (9)) (Fersht *et al.*, 1987).

develops in the E.[Tyr-ATP][‡] transition state (Leatherbarrow and Fersht, 1987). And, there is a further twist on this. The tyrosyl-adenylate is a high-energy compound, as well as being highly reactive, and the equilibrium constant for its formation from enzyme-bound tyrosine and ATP would normally be very low. But, the side chains bind the adenylyate tightest of all, and so displace the equilibrium to stabilise its formation as well as sequester it from solution (Wells and Fersht, 1989).

Interestingly, the individual values of $\Delta\Delta G^\ddagger$ and $\Delta\Delta G^0$ for the different mutations that bind the ribose of ATP could be combined to give sets of multi-point LFERs with β -value slopes, Figure 5 (Fersht *et al.*, 1986, 1987). These linear plots are not generally found in mutagenesis experiments as conformational changes are usually inhomogeneous, and so comparison of two-point plots and local clustering is the mainstay of the approach. The finding of subsets of LFERs in the sets of two-point measurements is a bonus here and in folding (Fersht and Sato, 2004). The presence of a multipoint

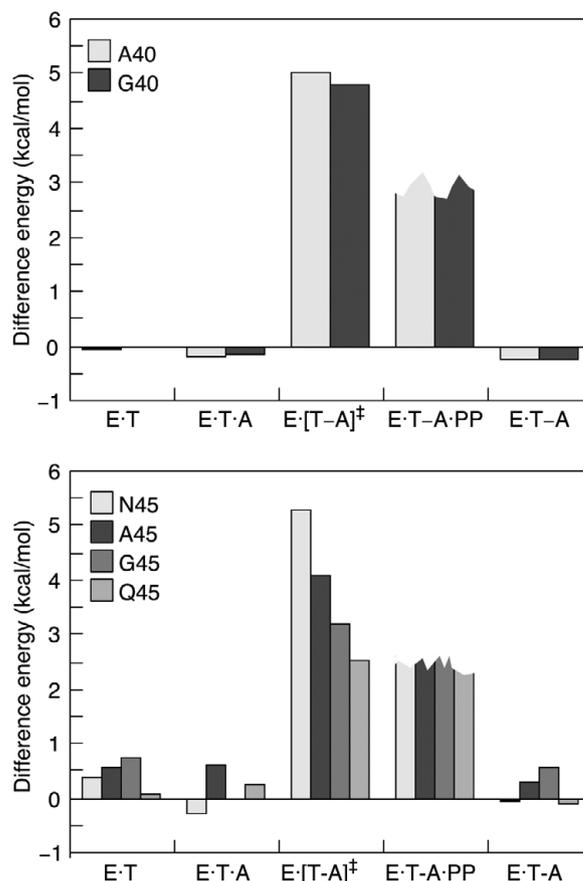


Figure 6. Difference energy diagrams for residues in the binding site of the tyrosyl-tRNA synthetase that bind to the charged oxygens of α -phosphate of ATP primarily in the pentacovalent transition state on the nucleophilic attack of the carboxylate of tyrosine (Fersht, 1987).

localised LFER for the residues that bind the sugar ring shows the enzyme generates a local pressure on the substrate to form the transition state, which validates Haldane’s: ‘Using Fischer’s lock and key simile, the key does not fit the lock perfectly, but exercises a certain strain on it’ (Haldane, 1930). The most dramatic mutational site, located by model building, has residues that barely affect the binding of the substrate or tyrosyl-adenylate product but just greatly stabilise charges developed on the α -phosphate in the transition state with $\beta \gg 1$, Figure 6 (Leatherbarrow *et al.*, 1985; Fersht, 1987), more consistent with Pauling’s general idea of transition state stabilisation (Pauling, 1948).

There are no chemical groups on the enzyme directly involved in catalysis. The carboxylate of the substrate tyrosine is a competent nucleophile and it appears that the mechanism of catalysis is the utilisation of binding energy to stabilise the transition state and displace an unfavourable equilibrium. By good fortune, the first application of protein engineering to study noncovalent interactions in enzyme catalysis discovered the first example of a natural enzymatic reaction being catalysed purely by transition state stabilisation without any of the classical mechanisms of chemical catalysis.

Basis for Φ -analysis for folding studies

Our 1987 paper provided the template for the analysis and choice of mutations for the analysis of folding pathways (Fersht *et al.*, 1987). In it, we introduced two-point β s for individual mutations from ratios of

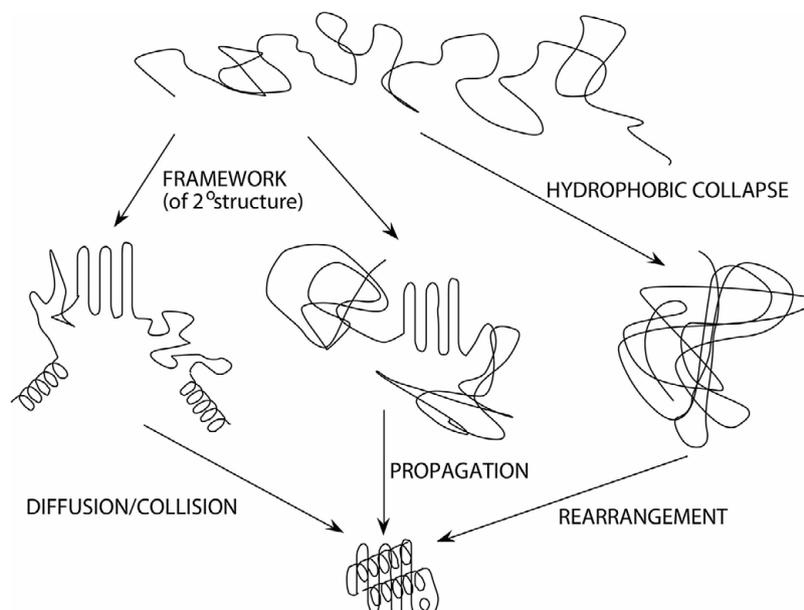


Figure 7. Classical mechanisms of folding. *Left:* the framework/diffusion-collision model; *middle,* nucleation-growth; *right,* hydrophobic collapse/molten globule.

$\Delta\Delta G^\ddagger/\Delta\Delta G^0$ in the difference energy plots, and elaborated on the possible groupings of them together to give true multipoint LFERs. We classified the mutations into six categories for choosing them: *Nondisruptive Deletion*, ‘a side chain is replaced by another that lacks a group involved in a specific interaction’; *Disruptive Deletion*, ‘replacement of a side chain may lead to a perturbation elsewhere in the structure’; *Conservative Substitution*, ‘a side chain is replaced by one that can substitute in the same interactions’; *Semiconservative Substitution*, ‘some of the function is conserved on replacement’; *Disruptive Substitution*, ‘substitution of a large size chain for a small one in a buried close packed region of a protein’; and *Nondisruptive addition*, ‘bulky groups may be added to the surface of proteins without necessarily causing perturbation of structure’. We documented the caveats about the effects of reorganisation of structure and effects of changes in solvation obscuring the analysis, which I discussed in more detail (Fersht, 1988). The protein-engineering β methodology that was developed for studying binding and catalysis was directly transferable to the problem of protein folding.

Naming the ratio $\Delta\Delta G^\ddagger/\Delta\Delta G^0$ as β , though well-intentioned, was misleading as the interpretation of protein engineering values differs in crucial ways from the Brønsted β of covalent chemistry because of the effects of mutation on denatured states among other details. β was renamed Φ in its first application to protein folding (Matouschek *et al.*, 1989) as Φ is not strictly a linear free energy quantity but approximates to one in certain circumstances. To avoid confusion, β is now reserved for the classical β of covalent catalysis and Φ for its counterpart in protein engineering (sections ‘From β - to Φ -value analysis’ and ‘Differences between β - and Φ -value analysis’ below).

Noncovalent transition states in protein folding: Φ -value analysis

The protein folding problem

The ‘protein folding problem’ consists of three closely related puzzles: (a) What is the folding code? (b) What is the folding

mechanism? (c) Can we predict the native structure of a protein from its amino acid sequence? (Dill *et al.*, 2008). Part (c), prediction of the three-dimensional structure of a protein from its linear amino acid sequence, goes back to Anfinsen (1973); and (b) the determination of the pathway to the folded structure from the unfolded to Levinthal (1968). The ‘code’ is how the information to fold is distributed along the structure. There is now a huge database of experimentally determined three-dimensional structures that has been the basis of very successful machine learning procedures for structure prediction, as embodied in AlphaFold (Jumper *et al.*, 2021). However, it is a black box that does not reveal the code or the pathway (Ooka and Arai, 2023). Determination experimentally of the pathway of folding of a protein is extremely difficult because a polypeptide chain progresses through a multitude of transient states as noncovalent interactions are formed and rearranged, and they are not amenable to direct experimental study.

The ‘Levinthal Paradox’ was that proteins could not fold in finite time in a random search. (See an interesting aside from Baldwin who was present at its initial presentation (Baldwin, 2017).) To solve this paradox, Wetlaufer proposed that one solution for the kinetics of folding was a nucleation-growth mechanism where a small local element of secondary structure slowly formed a nucleus and the structure rapidly grew around it (Wetlaufer, 1973). Ptitsyn proposed a framework (Ptitsyn, 1973) or diffusion-collision mechanism (Karplus and Weaver, 1976), whereby a framework of elements of secondary structure formed an intermediate rapidly in which they diffused and collided to dock on each other. Another proposal was hydrophobic collapse where non-specific tertiary interactions are rapidly made to form a molten-globule, which rearranges to give the final folded structure (Ptitsyn, 1991; Figure 7). Simple theoretical models, usually based on simulations on lattices, showed that the paradox arose because the original assumption was for an unbiased search for the folded state on a flat energy surface. In contrast, mechanisms utilising the gradual or otherwise acquisition of native interactions funnelling folding to the desired state obviated the paradox (Sali *et al.*, 1994a; Bryngelson

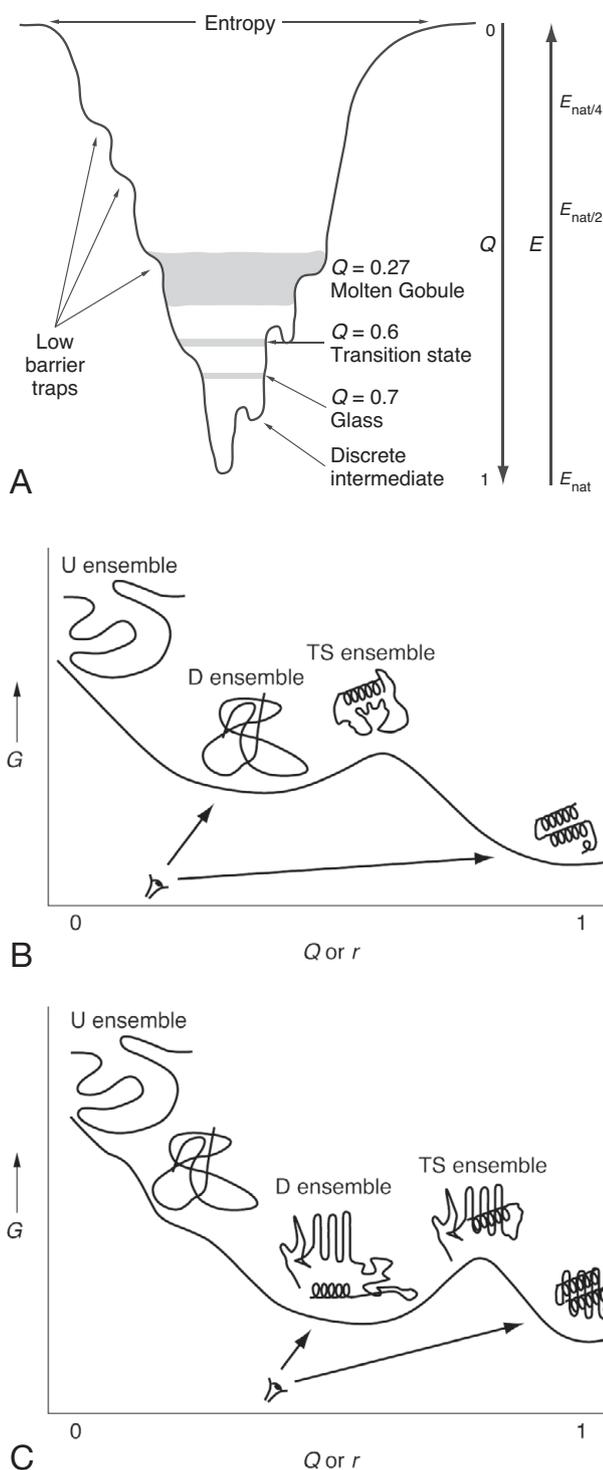


Figure 8. Reduction of an energy landscape to a conventional reaction coordinate diagram. This reconciles the classical view of a pathway with the ‘new view’ of an energy landscape with an ensemble of conformations (after Eaton *et al.*, 1996). Q is the relative number of pairwise native contacts in the landscape description and r is the conventional overall reaction coordinate. The number and heterogeneity of individual states decreases as the protein folds. (A, cross-section through a folding funnel (courtesy of P.G. Wolynes); B, reducing the landscape to a collection of ensembles moving along a pathway for the folding of a two-state protein such as C12; and C, folding of a protein with a more structured denatured state.

et al., 1995; Dill *et al.*, 1995; Onuchic *et al.*, 1995; Karplus, 2011; Takada, 2019; Finkelstein *et al.*, 2022). There was, however, an apparent conflict between the ‘classical view’ of protein folding proceeding along defined pathways with intermediates and a supposed ‘new’ view of folding on an energy landscape (Baldwin, 1995). From these theoretical studies, we now envisage proteins folding on multi-dimensional energy landscapes with a large number of conformations in the denatured state ensemble with high entropy converging on decreasingly smaller ensembles in transition states and intermediates to the final structure, with the gain in enthalpy from native interactions compensating the loss of entropy. We can represent these ensembles as states along a two-dimensional energy diagram, Figure 8 (Eaton *et al.*, 1996). It must be emphasised that what the experimentalist sees as the denatured state, D , under conditions that favour folding, D_{phys} , is not usually a random coil, U , but a more structured state varying from having flickering interactions (Figure 8a) to a fairly structured on- or off-pathway intermediate (Figure 8b). The basics of protein folding studies are discussed in more detail in Fersht (1999, 2017, 2018, Chaps. 17–19).

Nucleation mechanisms went out of favour because the early experimental examples of protein folding were found to proceed via intermediates on the pathway (Ptitsyn, 1987; Kim and Baldwin, 1990), and nucleation is characterised by not having intermediates that would accumulate.

From β - to Φ -value analysis

Studies on the effects of point mutations on folding kinetics had begun in the late 1980s with Matthews analysing natural mutants of the α -subunit of tryptophan synthase (Matthews, 1987). Goldenberg protein engineered mutants of bovine pancreatic trypsin inhibitor (Goldenberg *et al.*, 1989). We began applying the technology and Φ -strategy developed on the tyrosyl-tRNA synthetase to the folding of a small RNase, Barnase (Kellis *et al.*, 1988; Sali *et al.*, 1988; section ‘Barnase: the test bed’). The two-point LFER approach used for the mapping the progress of noncovalent interactions in enzyme catalysis is directly applicable to studying transition states and transient intermediates in folding. But, there are crucial refinements, which were laid out in the initial LFER paper (Matouschek *et al.*, 1989 and subsequently expanded in more depth (Fersht *et al.*, 1992; Fersht and Sato, 2004), relying on the thermodynamic cycles in Figure 9, which are essential to the analysis (the use of such alchemical cycles was perhaps not obvious and queried at the time (Buchner and Kiefhaber, 1990). Accordingly we used the same strategy as before: (1) make chemically sensible mutations in a suitable protein by truncating side chains to remove stabilising interactions (avoid mutations that cause stereochemical clashes or unstable charges within the protein – the nondisruptive deletions, especially of hydrophobic side chains); (2) measure the change in the free energy of folding of the protein on mutation, $\Delta\Delta G_{N-D} (= \Delta G_{N'-D'} - \Delta G_{N-D})$, where N = native state, D denatured state, and N' and D' refer to mutants); and (3) measure the rate constants of folding, k_f , of the wild-type and mutant proteins to determine the changes in the free energies of activation $\Delta\Delta G_{\ddagger-D} (= \Delta G_{\ddagger-D'} - \Delta G_{\ddagger-D} = -RT\ln(k'_f/k_f))$ and rate constants for unfolding, k_u , to give $\Delta\Delta G_{\ddagger-N} (= \Delta G_{\ddagger-N'} - \Delta G_{\ddagger-N} = -RT\ln(k'_u/k_u))$.

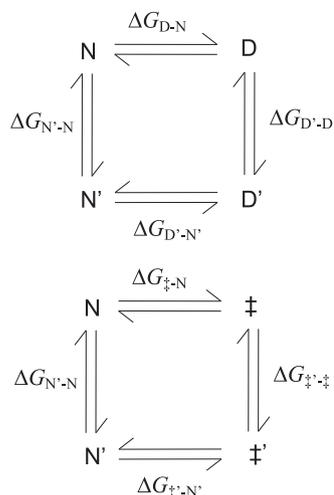


Figure 9. Thermodynamic cycles for the basis of Φ -value analysis (relabelled from Matouschek *et al.*, 1989).

We then defined a parameter Φ for folding. In the direction of folding:

$$\Phi_F = \Delta\Delta G_{‡-D} / \Delta\Delta G_{D-N} = (\Delta G_{‡'-D'} - \Delta G_{‡-D}) / (\Delta G_{N'-D'} - \Delta G_{N-D}). \quad (10)$$

And for unfolding:

$$\Phi_U = \Delta\Delta G_{‡-N} / \Delta\Delta G_{D-N} = (\Delta G_{‡'-N'} - \Delta G_{‡-N}) / (\Delta G_{D'-N'} - \Delta G_{D-N}). \quad (11)$$

We can derive from the thermodynamic cycles in Figure 9 that $\Delta\Delta G_{D-N} = \Delta G_{D'-D} - \Delta G_{N'-N}$; $\Delta\Delta G_{‡-N} = \Delta G_{‡'-N'} - \Delta G_{N'-N}$; and $\Delta\Delta G_{‡-D} = \Delta G_{‡'-D'} - \Delta G_{D'-D}$. Accordingly,

$$\Phi_F = (\Delta G_{‡'-D'} - \Delta G_{D'-D}) / (\Delta G_{D'-D} - \Delta G_{N'-N}), \quad (12)$$

$$\Phi_U = (\Delta G_{‡'-N'} - \Delta G_{N'-D}) / (\Delta G_{N'-D} - \Delta G_{D'-D}). \quad (13)$$

Ignoring the changes in covalent energies on mutation as they cancel out in subsequent calculations, the term $\Delta G_{N'-N} = \Delta G_{(N'-N)\text{noncovalent}} + \Delta G_{(N'-N)\text{reorg}}$, where $\Delta G_{(N'-N)\text{noncovalent}}$ is the change in noncovalent interactions from the mutation and $\Delta G_{(N'-N)\text{reorg}}$ is any energetics of reorganisation of the structure of the folded protein. There are similar equations involving ΔG_{reorg} for the change in energetics of the denatured and transition states including changes in solvation, ΔG_{solv} . For denatured states that are highly unfolded, ΔG_{solv} is the major term in ΔG_{reorg} but often for the interior in folded proteins $\Delta G_{\text{solv}} = 0$.

Building on our classification of mutations (Fersht *et al.*, 1987) and thermodynamic analysis (Fersht, 1988), it was spelled out clearly in the first paper what type of mutations to make in the light of incursion of ΔG_{solv} , and how the choice affects the observed values of Φ (Matouschek *et al.*, 1989). Assuming that the effects of mutation on the noncovalent interactions are localised to the site of the side chain, the two extreme situations are readily interpretable (Figure 10). If the side chain is as unstructured in the transition state as in the denatured state, $\Delta G_{‡'-‡} = \Delta G_{D'-D}$ and so $\Phi_F = 0$ and $\Phi_U = 1$. Conversely, if the side chain is as structured in the transition state as in the native state, $\Delta G_{‡'-‡} = \Delta G_{N'-N}$, and so $\Phi_F = 1$ and $\Phi_U = 0$. This is the same as the extreme cases of the Brønsted β . For mutations of larger to smaller aliphatic side chains, which are the most suitable as we cannot emphasise enough, $\Delta G_{D'-D}$ (i.e. ΔG_{reorg}) should be small. For

example, mutation of Ile→Ala and Ile→Val have $\Delta G_{\text{solv}} = -0.21$ and -0.16 kcal/mol, respectively. The deletion of a $-\text{CH}_2-$ group will lead to minimal G_{reorg} . Accordingly, Φ_F is related to the extent of local structure formation in the native and transition states (Matouschek *et al.*, 1989; Fersht *et al.*, 1992; Fersht and Sato, 2004). This is especially so for Ala→Gly scanning in helices (section 'Ala→Gly scanning of secondary structure').

Differences between β - and Φ -value analysis

In many ways, the interpretation of Φ -values is analogous to that of β , but there are important differences that must be minimised for the successful application of Φ . In the classical chemical LFERs, the structural changes made in the reagents are at positions separated from the reacting bonds and the effects of the substituents transmitted through the molecule. ΔG_{reorg} terms for β in covalent chemistry are ignored because they are relatively small or non-existent. Basically, β (or α) = $\partial\Delta G_{‡-S} / \partial\Delta G_{P-S}$ in Figure 3. In the protein engineering LFERs, the very groups making the bonds are changed and there can be a significant ΔG_{reorg} in the native state and possibly in a structured denatured state. There can also be ΔG_{solv} terms for both states. To acknowledge these differences, as mentioned previously, β was renamed Φ , and Φ -analysis experiments designed to minimise or accommodate those ΔG terms (Matouschek *et al.*, 1989). When this is done, Φ is very similar to β .

(Water molecules surrounding the reactants and catalyst in classical chemical LFER experiments may rearrange on changing a substituent and cause significant changes of $\Delta H_{\text{reorg}}^\ddagger$ and $-T\Delta S_{\text{reorg}}^\ddagger$ but those changes tend to compensate and cancel out in $\Delta\Delta G$, although they do complicate attempts to measure the ΔH and ΔS components of the actual chemical steps.)

REFERS: β Tanford (β_T), Leffler/Brønsted plots and Φ

Protein folding has other important differences such as the difficulty in choosing a suitable reaction coordinate. A global average may be defined for overall folding but the formation of structure is not homogeneous and the local reaction coordinates for substructures are what define the formation of transition states and intermediates. The interpretation of Φ -values is more complicated than that of β and extra procedures may be involved. A simple overall reaction coordinate was introduced by Tanford (1968, 1970). All parts of a protein are stabilised by denaturant, Den, and its free energy increases linearly with [Den] and the solvent accessible surface area (SASA). There is a decrease in SASA on going from $D \rightarrow \ddagger \rightarrow N$, so $\Delta G_{‡-D} = \Delta G_{‡-D}^0 - m_{‡-D}[\text{Den}]$; $\Delta G_{‡-N} = \Delta G_{‡-N}^0 + m_{N-D}[\text{Den}]$; and $\Delta G_{D-N} = \Delta G_{D-N}^0 - (m_{N-D})[\text{Den}]$; where for 2-state kinetics $m_{N-D} = m_{‡-D} + m_{‡-N}$ (all m -values +ve). The relative change in surface area in the transition state, which I renamed β_T in homage to Tanford (Matouschek *et al.*, 1995), is given by: $\beta_T = m_{‡-D} / m_{D-N}$. The Tanford plot is a true REFER.

Leffler plots, which are also called Brønsted plots, of $\Delta G_{‡-N}$ versus ΔG_{D-N} or $\Delta G_{‡-D}$ versus ΔG_{N-D} also give an indication of the overall change in energetics. However, they can exhibit scatter depending on the inhomogeneity of structure formation in the transition state (see later in the discussion of Figure 15, section 'Chymotrypsin inhibitor 2: computer simulations'). Just as the finding of multipoint LFERs/REFERS for the tyrosyl-tRNA synthetase is a bonus, resulting from concerted movement of parts of the binding site relative to the substrate, the same can sometime be found for Φ -analysis. Part of a helix in barnase, for example, is uniformly present in the transition state and its formation can be

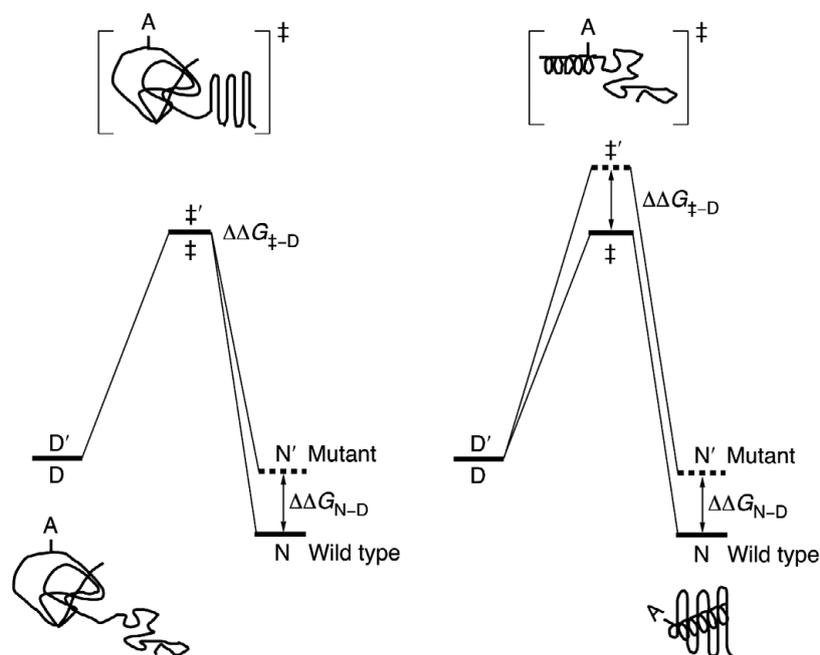


Figure 10. Free energy profiles for mutations giving $\Phi = 0$ when the mutated residue A is in disordered region (left) or 1 in a fully native (right). The energy profiles are simplified with the energies of the denatured states D for wild-type and D' for mutant being set at the same level.

benignly probed by truncating surface exposed side chains to Ala and then Gly to give a series of overlapping 3-point Leffler/Brønsted plots (Matthews and Fersht, 1995; Fersht and Sato, 2004). Accordingly, Φ is a true REFER for those mutations.

ψ -value analysis

Disulphide crosslinks tie together residues in both the transition states and denatured states as well as native states, with predictable effects on kinetics that can detect when the linked elements of structure are formed during the folding pathway of wild-type protein (Clarke and Fersht, 1993). This is a highly specific procedure and very limited in applicability. Sosnick has pioneered a more general mutational procedure for this crosslinking approach for surface residues, ψ -value analysis (Krantz and Sosnick, 2001; Baxa and Sosnick, 2022). Pairs of histidine residues as metal-binding sites are introduced on the surface typically close to each other in the folded state, for example, at positions $i, i+4$ in an α -helix or at neighbouring strands in a β -sheet ('nondisruptive additions'). A metal ion can then crosslink the pair. This contrasts with Φ -value analysis in that ψ - adds new interactions to the protein and analyses their effects on the mutants whereas Φ -analysis uses non-disruptive deletions that probe the extent of formation of interactions present in the wild-type structure. ψ -value analysis is not an REFER but the values of 1 or 0 should be interpretable (Fersht, 2004a; Bodenreider and Kiefhaber, 2005). Indeed, simulation of the transition state for the folding of ubiquitin is consistent with ψ -values of 1 or 0 but not the fractional ones. It is a useful tool for those values (Varnai *et al.*, 2008).

Interpretation of Φ -values

Weak, medium, and strong categorisation of Φ

The values of $\Phi = 0$ or 1 may be interpreted with confidence. Mutations such as Ile→Val, Ala→Gly, and Thr→Ser are particularly suitable and Ile→Ala can be good – see section 'Experimental

approach to Φ -value analysis'. In general, the Φ -values should be interpreted only semi-quantitatively and with caution: $0 < \Phi_F < 0.2$, 'low' or 'weak', little or no structure in transition state; $0.3 < \Phi_F < 0.6$, 'medium' significant to strong; and $0.7 < \Phi_F < 1$, 'high' or 'strong', very significant structure (with flexibility as to the boundaries) – like weak, medium and strong NOEs used as distance constraints in molecular dynamics (MD) calculations in structure determination by NMR (Fersht and Sato, 2004; Garcia-Mira *et al.*, 2004) and such classification has been applied with success in computer simulations of the structure of transition states (Geierhaas *et al.*, 2008). As discussed later, Φ -values may be powerfully combined with computer simulations of unfolding and folding trajectories to give true atomic-level descriptions of protein folding pathways. It is important to make many mutations and over-sample to find consistent results that then give reliable information. Φ -values by themselves can give gross and near atomic resolution details on the structures of transition states. There are some areas that are more problematic, which I next describe and how they may be resolved.

Φ and non-native interactions: $\Phi < 0$ or $\Phi > 1$

Φ , like β , is predicated on a single bond or set of bonds being formed, with limits of 0 for no formation and 1 for complete. It parallels in some ways the Gō model in simulation that assumes that only native contacts are involved in the folding process and they consolidate (Taketomi *et al.*, 1975; Takada, 2019). If there are non-native interactions in transition states or intermediates, then unnatural values of Φ of < 0 or > 1 may be observed, and they are a useful signal for that. Residual structure in the denatured state can give rise to non-classical values (Cho and Raleigh, 2006). Small two-state single-domain proteins are the most likely not to involve non-native interactions (Best and Hummer, 2016), and Gō model simulations can fit well with Φ measurements (Clementi *et al.*, 2000; Wu *et al.*, 2008; Naganathan and Orozco, 2011).

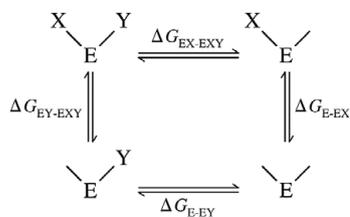


Figure 11. Double-Mutant cycles. X and Y are mutated individually and as a pair, and the values of ΔG_{D-N} or $\Delta G_{\ddagger-N}$ measured. Interaction energies of X and Y with other residues cancel in the $\Delta\Delta G_{\text{int}}$ cycles and are perturbed only by $\Delta\Delta G_{\text{reorg}}$ terms in the folded state. For the denatured state, $\Delta\Delta G_{\text{int}} = 0$ when the residues X and Y do not interact with each other. Accordingly, the measured values of $\Delta\Delta G_{\text{int}} = \Delta\Delta G_{\text{EY-EY}} - \Delta G_{\text{E-EX}} = \Delta G_{\text{EX-EXY}} - \Delta G_{\text{E-EY}}$ give the interaction energies between X and Y in the native state at equilibrium or in the transition state for kinetics.

Double-mutant cycles to identify native partners in interactions

Φ -value analysis interprets changes in energy to changes in structure and assumes that the native interactions are involved, and there can be complications from non-native interactions. Strong evidence about which residues interact can be found by the procedure of double-mutant cycles (Figure 11), first introduced for the tyrosyl-tRNA synthetase (Carter *et al.*, 1984). Two residues that interact in the native state of the protein are mutated individually, and then pairwise. An interaction energy between just those two residues $\Delta\Delta G_{\text{int}}$ is measured without complications from an unfolded denatured state (Fersht *et al.*, 1992). The same is true for the interaction in the transition state, $\Delta\Delta G_{\text{int}}^{\ddagger}$. Values of $\Phi_{\text{int}} = \Delta\Delta G_{\text{int}}^{\ddagger} / \Delta\Delta G_{\text{int}}$, show with high certainty whether or not and by how much those interactions are formed in the transition state (Horovitz and Fersht, 1990, 1992; Horovitz *et al.*, 1991; Fersht *et al.*, 1992; Pagano *et al.*, 2021). They can be used to provide constraints for computer simulations of transition state structure (Salvatella *et al.*, 2005). Multi-mutant cycles can also be performed (Horovitz and Fersht, 1990, 1992).

Parallel pathways and fractional Φ -values

A fractional Φ -value is usually interpreted as arising from a single transition state ensemble that has weakened interactions. But

there could be parallel pathways, as in, Figure 12, with some having full structure at the point of mutation and others disordered and these could give an apparent fractional value (Baldwin, 1994; Sali *et al.*, 1994b). This can be tested, however, by making a series of additional mutants that would have different and predictable effects on the disordered and structured pathway states, and the fractional values of Φ for the protein CI2 (below) being consistent with a single pathway through the transition state (Fersht *et al.*, 1994).

Residual structure in denatured states

Denatured states can have residual structure even at high concentrations of denaturants (Dill and Shortle, 1991; Cho and Raleigh, 2006) and especially at low concentrations where the most stable denatured state may be a folding intermediate or an off-pathway state with non-native interactions. Residual structure is melted out less slowly by denaturants and temperature as there are smaller changes in surface area. These states can severely affect folding kinetics of all types. But, unfolding kinetics and $\Delta G_{\ddagger-N}$ from the folded state are unaffected as the denatured states are after the rate-determining transition state. ΔG_{D-N} is measured at higher concentrations of denaturant but there could be significant $\Delta G_{(D-N)\text{reorg}}$ terms with mutations affecting structure in the denatured state. Values of Φ_U close to 0 will be relatively unaffected but values closer to 1 may have artefacts. For these reasons, we gave up the terminology ‘U’ = unfolded for the denatured state and call it D or D_{phys} under physiological conditions.

Experimental approach to Φ -value analysis

ΔG_{reorg} and choice of mutation

The presence of $\Delta G_{(N'-N)\text{reorg}}$ and similar terms dictates the choice of mutation. To recapitulate earlier points, a mutation of a buried side chain to a larger one will likely cause a significant $\Delta G_{(N'-N)\text{reorg}}$ as will changes in buried charges. Accordingly, mutations that preferably delete interactions, non-disruptive deletions, or are

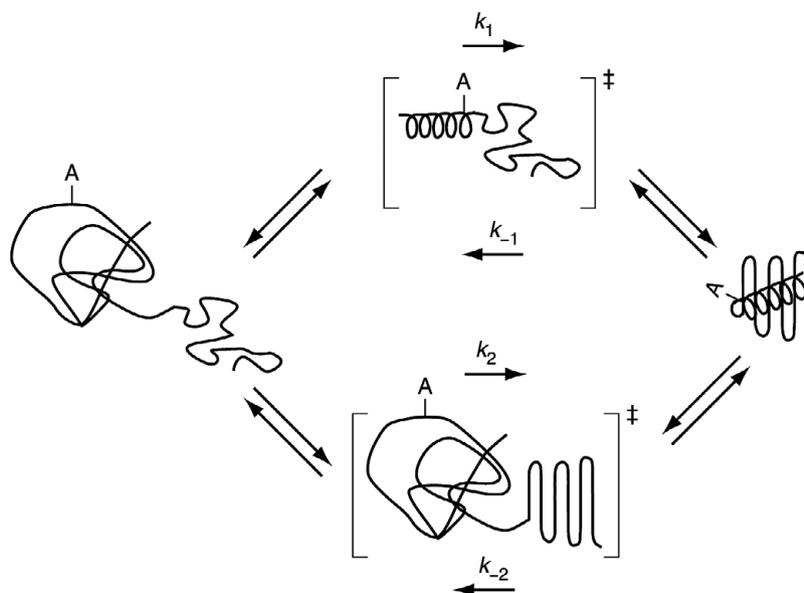


Figure 12. Possible parallel pathways of folding (Fersht *et al.*, 1994).

isosteric are most suitable. The changes in energetics must be sufficiently large to be able to be measured accurately but not too large, otherwise the position of the transition state may be perturbed or there will be a local rearrangement of structure on making a too-large deletion.

Our preferred strategy is: (1) to mutate the buried hydrophobic moieties Ile→Val→Ala→Gly; Leu→Ala→Gly; Thr→Ser; and Phe→Ala→Gly. Deletion of a $-\text{CH}_2-$ has minimal effects on the solvation energies of the denatured state and low ΔG_{reorg} in all states; (2) make a wider range of surface mutations; (3) mutate Ala→Gly positions in secondary structural regions ('Ala→Gly scanning', see 7.2), especially in α -helices, because they provide an exquisite probe of secondary structure in the helix since mutation perturbs mainly intra-helical interactions; and (4) use, sparingly, double-mutant cycles in which changes in solvation and reorganisation energies tend to cancel out. Mutation of a long aliphatic side chain in the hydrophobic core, such as that of isoleucine, can give information on the degree of consolidation of the core on mutation to Ala, and then the structure of the helix during that process on subsequent mutation to Gly. Successive deletion of different parts of larger side chains may give multiple probes of structure (Serrano *et al.*, 1992b). These types of mutation tend to give values of $\Delta\Delta G_{\text{D-N}}$ in the range of 0.6–2 kcal/mol, which can be measured with adequate precision and typical of the interactions that report on secondary structure as well as local interactions in hydrophobic cores (Friel *et al.*, 2003; Fersht and Sato, 2004; Garcia-Mira *et al.*, 2004; Sato *et al.*, 2006). Larger changes can lead to a movement of the transition state on the energy landscape (Fersht and Sato, 2004).

Ala→Gly scanning of secondary structure

Mutation of Ala→Gly in helices is a particularly clean tool (Matthews and Fersht, 1995). The CH_3- side chain of Ala stabilises an α -helix relative to the H- of Gly mainly by burial of the hydrophobic surface area, from 0.4 to 2 kcal/mol, and mutation has minimal structural perturbation (Serrano *et al.*, 1992a,c). Further, unfolded alanine- and glycine-containing peptides are approximately isoenergetic in noncovalent interactions (Scott *et al.*, 2007) and so mutation of Ala→Gly has minimal ΔG_{reorg} terms in both states. Accordingly, $\Phi_{\text{Ala} \rightarrow \text{Gly}}$ is the most reliable measure of structure formation of all Φ -values.

Experimental determination of ΔG s

The changes in ΔG^\ddagger and $\Delta G_{\text{D-N}}$ are mostly measured from variation of the rate constants of folding and unfolding and the equilibrium constant with concentration of a denaturant such as urea or guanidinium chloride. Usually, logarithms of the rate and equilibrium constants for unfolding increase linearly with concentrations of denaturant under the accessible experimental conditions, but sometimes with small deviations at very low concentrations (Tanford, 1968, 1970). For two-state kinetics, the logarithm of the rate constants for folding decrease linearly with denaturant concentration (Tanford *et al.*, 1973) and plots of the combinations of $\log k_{\text{u}}$ and $\log k_{\text{f}}$ give so-called chevron plots as in Figure 13 (Jackson and Fersht, 1991a). For multi-state systems, the refolding limb is usually characterised by 'rollover' where the folding rate constant tends to plateau at low denaturant concentration as there are changes in rate-determining steps, Figure 13, inset (Matouschek *et al.*, 1989). The proteins in Figure 13 refold on the tens of ms time scale, the kinetics measured by rapid-mixing

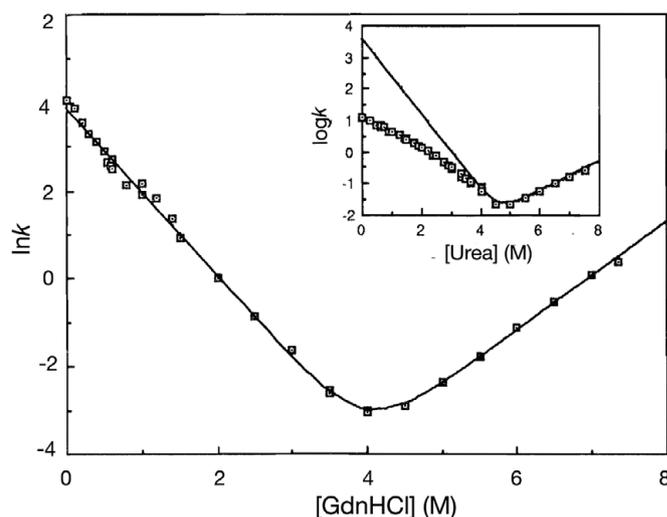


Figure 13. Chevron plot for the folding of C12 determined by stopped-flow kinetics (Jackson and Fersht, 1991a) and, inset, barnase (Matouschek *et al.*, 1990). Rate constants are in units of s^{-1} . For C12, the plot is for a perfect two-state transition and the arms are linear. For barnase, there is deviation at low denaturant concentration from the perfect theoretical two-state (solid line) because of a change in the structure of the denatured state or presence of a folding intermediate.

stopped-flow methods. Smaller single-domain proteins can fold even faster on the μs time scale as for the 37-residue Formin-Binding Protein, FBP28, a canonical three-stranded β -sheet WW domain, Figure 14 (Petrovich *et al.*, 2006). Its kinetics of folding and unfolding are too fast for rapid mixing but are readily and accurately measured using temperature-jump apparatus. The unfolding of such small proteins exposes only a relatively small amount of buried surface area and so the transition is spread out over a wide range of concentration of denaturant. The FBP28 domain has a very polarised transition state as readily seen directly from the chevron plots. Some plots have the folding limbs nearly superposed, showing $\Delta G_{\ddagger-\text{D}} \sim 0$ and so $\Phi_{\text{F}} \sim 0/\Delta\Delta G_{\text{N-D}}$, that is ~ 0 for non-zero values of $\Delta\Delta G_{\text{N-D}}$. Conversely, other plots have the unfolding limbs nearly superposed, showing $\Delta G_{\ddagger-\text{N}} \sim 0$ and so $\Phi_{\text{U}} \sim 0/\Delta\Delta G_{\text{D-N}}$, that is ~ 0 . As, $\Phi_{\text{U}} + \Phi_{\text{F}} = 1$ for two-state kinetics, these chevrons of $\Phi_{\text{U}} \sim 0$ have $\Phi_{\text{F}} \sim 1$. These values of $\Phi_{\text{F}} \sim 0$ or 1 are also determined with the highest confidence as the errors around $\Delta\Delta G_{\ddagger-\text{N}}$ and $\Delta\Delta G_{\ddagger-\text{D}} \sim 0$ are small. An error of, say, ± 0.1 for a mean of $\Phi_{\text{U}} = 0.05$ is a very high percentage error in the absolute value of Φ_{U} but in the context of where Φ_{U} is on the scale of 0 to 1 is sufficiently accurate for the purposes of interpretation. Accordingly, the most readily interpretable values of Φ , 0 and 1, are the ones most amenable to confident measurement.

I advocate for optimising precision measuring differences in ΔG^\ddagger and $\Delta G_{\text{D-N}}$ directly under the same reaction conditions (same concentration of denaturant, [Den]) and not extrapolating to the absence of denaturant. In our laboratory, we can measure $\Delta\Delta G_{\text{D-N}}$ with adequate precision down to ~ 0.6 kcal/mol from the differences in the midpoints of equilibrium denaturation curves of wild-type and mutants (Clarke and Fersht, 1993) or from the unfolding and folding rate constants (Fersht and Sato, 2004) as do other (Friel *et al.*, 2003; Garcia-Mira *et al.*, 2004). First-order rate constants for unfolding and refolding can be determined with high precision. Attention to detail is important. We make up stock solutions of denaturant for each concentration, using volumetric flasks rather than diluting one concentrated stock solution into buffer. I avoid using phosphate buffer with guanidinium chloride as it lowers the

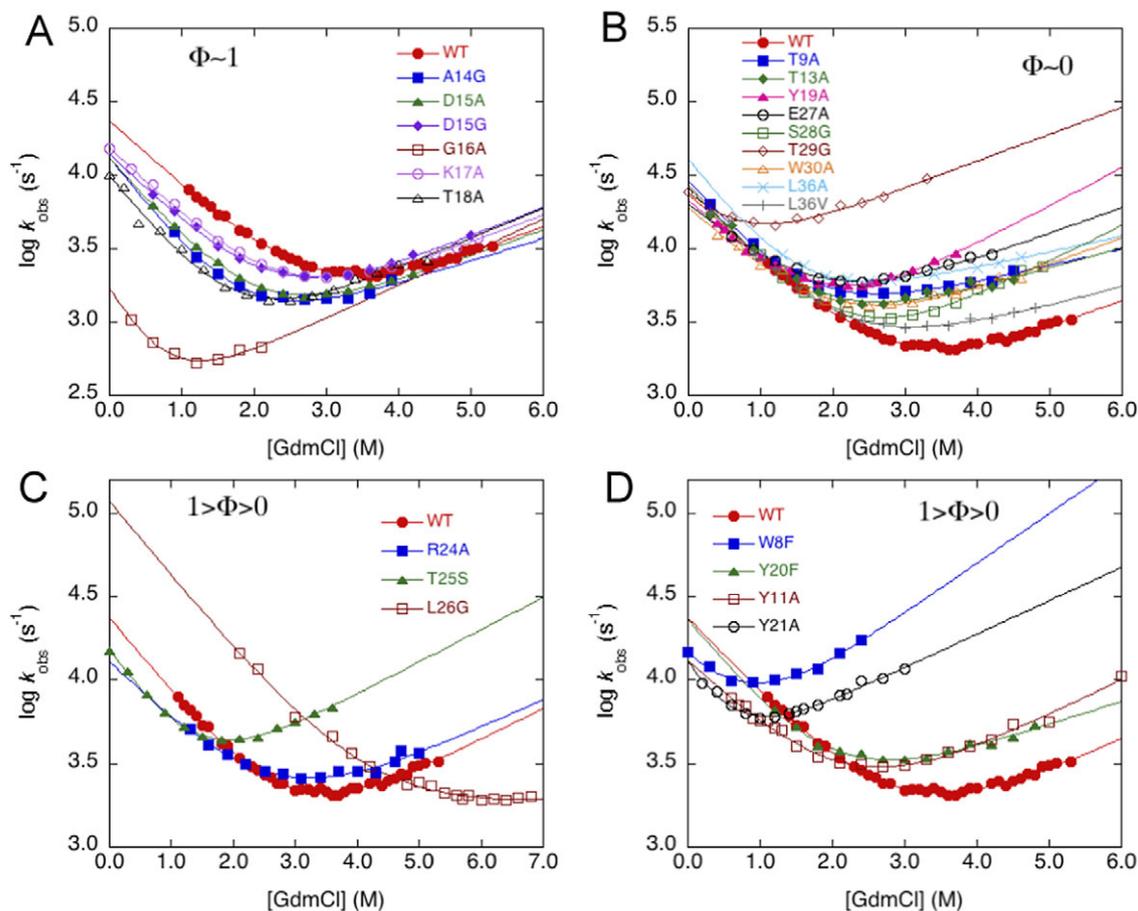


Figure 14. Chevron plots for folding of FBP28, which nicely illustrate $\Phi = 0$, B, where the refolding limbs overlap, or 0, A, where the unfolding limbs overlap, and C and D for fractional values. T-jump was required for the rate constants in the range of $10,000 \text{ s}^{-1}$ (Petrovich *et al.*, 2006).

pK_a greatly with increasing [Den] because its ionic component displaces the ionisation equilibrium $\text{H}_2\text{PO}_4^- = \text{HPO}_4^{2-} + \text{H}^+$ as according to the Debye–Hückel equation the activity coefficient of an ion depends on the charge squared (Debye and Huckel, 1923). (The application to kinetics was implemented in the Brønsted–Bjerrum equation.). Instead, I prefer an amine buffer at neutrality or at lower pH acetate because their ionizations parallel more closely the principal protein ionizations at those pHs; histidine/ α -amino groups, and aspartate/glutamate (Fersht and Petrovich, 2013). Urea does not have this problem. To minimise problems from changes of pH with temperature and denaturant concentrations and so forth, measurements are best made at pHs where free energies and kinetics are pH independent.

Combining Φ -values with and benchmarking computer simulation

The complete conscription of folding pathways of proteins can be achieved only by computer simulation. This is possible *de novo* only when the energy potentials are sufficiently reliable, or a black box machine learning is applicable. The role of the experimentalist has been to provide the structures of all the states along the pathway as a starting basis for simulation and to benchmark simulation within the limitations of current energy functions. Φ -values since their initial introduction have provided the crucial benchmark for interactions in the transition state for the folding of the small domains, the most easily studied computationally because of the limitations

on computing power. They are being used for testing more complex folding of large proteins (Ooka and Arai, 2023). There are methods for calculating Φ -values directly (Best and Hummer, 2016).

Barnase: the test bed

Φ -value analysis was pioneered on the 110-residue RNase, Barnase, from *Bacillus amyloliquefaciens*. It is a most suitable small protein for structure-activity studies using protein engineering, readily expressed from *E. coli* and does not have complications from disulphide bridges or *cis*-prolines in the folded state. The strategy for studying it has two steps as for the tyrosyl-tRNA synthetase studies; (1) mutate the protein sensibly and extensively to build up a library of the common interactions that stabilise proteins; and (2) select suitable mutants for kinetic analysis.

Step 1: library of interaction energies that stabilise proteins

The magnitudes of the hydrophobic effect and other interactions were usually measured from simple free energies of transfer from organic solvents to water (Fersht, 1999, 2017, 2018 Ch. 11) or more appropriately for α -helices the stabilities of synthetic peptides in water (Padmanabhan *et al.*, 1990). We made the first systematic measurements of the common interactions that stabilise proteins directly in a protein from the values of ΔG_{D-N} of wild-type barnase versus mutants whose side chains had been truncated by non-disruptive deletions. The deletion of $-\text{CH}_2-$

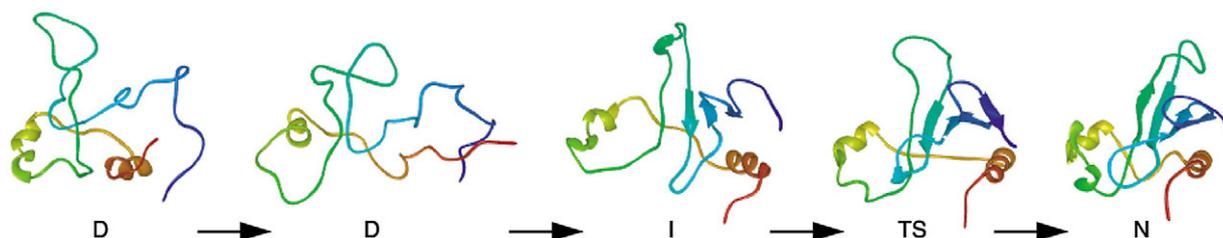


Figure 15. Barnase folding from experiment and simulation. An MD unfolding simulation from the native state N to the denatured state D at 225 C, is shown in reverse. The structures are coloured from red at the N-terminus to blue at the C-terminus. The denatured state is an ensemble of structures whose overall topology resembles that of the native state. The hairpin at the centre of the antiparallel β -sheet is present in the denatured state, albeit with some non-native interactions. The N-terminal helix is partly structured, stabilised by hydrophobic interactions. The final transition state consists of the largely formed N-terminal helix docked onto the β -sheet, which is strongly formed in the central regions, with the hydrophobic core in the process of being formed and other interactions consolidated (Fersht and Daggett, 2002).

group from a residue in the hydrophobic core lowers stability by up to 1.6 kcal/mol compared with 0.68 kcal/mol in the simple chemical models (Kellis *et al.*, 1988, 1989). The mutation of Ala \rightarrow Gly in the exposed surface of helices lowers stability of 0.4–2 kcal/mol and depends on the amount of surface area of the CH_3 - group of Ala buried (Serrano *et al.*, 1992a, 1992b). Mutants from these studies with suitable values of $\Delta G_{\text{D-N}}$ were chosen for the kinetic studies.

Step 2: kinetics

The initial study was on the unfolding of the protein as it starts from the best-characterised state on the pathway and the folding direction can be beset by problems of residual structure in the denatured state or even intermediates (Matouschek *et al.*, 1989). Unfolding kinetics provides in general the most reliable data and is very relevant to biology because many diseases are initiated by protein unfolding. The folded state is the best-characterised starting point also for computer simulation. The unfolding transition state for folding is generally the highest energy state on the folding pathway.

Barnase is a multimodular protein, having regions that make more interactions within themselves than with the rest of the protein, with three hydrophobic cores and a mixed $\alpha + \beta$ architecture. Some of the regions have Φ -values near 1, others have values of 0, and some regions are intermediate. The centre of the sheet and the C-terminal portion of helix 1 have Φ -values of approximately 1. There are fractional Φ -values for the edges of the sheet and for the packing of the N-terminal α -helix on the β -sheet, which constitutes the major hydrophobic core. The second domain, containing helix 2, and the loops have Φ -values ~ 0 . The multimodular barnase has a polarised major transition state, which occurs late on the reaction pathway with much of the secondary structure being formed and the hydrophobic core between the major α -helix and β -sheet in the process of being consolidated (Matouschek *et al.*, 1989; Serrano *et al.*, 1992a).

Folding intermediate or structured D_{phys} ?

The downward curvature in the refolding limb of the $\log k_{\text{obs}}$ versus [Urea] plot (Figure 13) was the initial evidence that there is either a folding intermediate or structured denatured state, D_{phys} , whose concentration or properties change with concentration of denaturant (Matouschek *et al.*, 1990). A structured D_{phys} that progressively unfolds in a non-cooperative transition could give rise to a variable two-state process. Φ -values probe the structure of this state

(Matouschek *et al.*, 1992), which has been extensively studied by a variety of methods (Khan *et al.*, 2003) and simulation (Caflich and Karplus, 1995; Li and Daggett, 1998; Wong *et al.*, 2000; Galano-Frutos and Sancho, 2019). The biophysics is consistent with a cooperative unfolding of the state (Dalby *et al.*, 1998a, 1998b). There are probably two intermediates on the pathway (Khan *et al.*, 2003; Sanchez and Kiefhaber, 2003). Φ_{F} -values measured from ill-defined folding intermediates must be interpreted with caution because there may be non-native interactions involved. Time-resolved small-angle X-ray scattering indicates an expanded state (Konuma *et al.*, 2011). The evidence is consistent with some fraction of the denatured ensemble containing residual, non-random structure, especially in helix 1 and the turn ($\beta 3$ – $\beta 4$) in the centre of the sheet consistent with MD simulation of the denatured state (Bond *et al.*, 1997; Wong *et al.*, 2000). The folding pathway is simulated atomistically by running the unfolding pathway in reverse, Figure 15 (Fersht and Daggett, 2002; Daggett and Fersht, 2003).

Chymotrypsin inhibitor 2: two-state kinetics and nucleation-condensation

Our second protein studied, Chymotrypsin Inhibitor (CI2), is a 64-residue single-domain protein, unlike most of the previous proteins then studied which were multi-domain. It has a single α -helix, docked onto β -sheet, a single-module protein. In contrast to those other proteins then studied (Ptitsyn, 1987; Kim and Baldwin, 1990), CI2 was found to fold by two-state kinetics without an intermediate and, for that time, relatively fast on the 10 ms time scale (Jackson and Fersht, 1991a, 1991b). Intermediates do not detectably accumulate in its folding and the ratio of rate constants for folding and unfolding give the correct equilibrium constant for denaturation, again unlike for the previously studied proteins. The chevron plot has perfectly linear arms, Figure 13. Its single rate-determining transition state for folding can be studied in both directions to show unfolding and folding are the reverse pathways of each and so microscopic reversibility is obeyed. More examples of two-state folding were quickly found (Jackson, 1998) and 89 proteins are now reported with two-state folding kinetics (Manavalan *et al.*, 2019). The small single-domain proteins are very suitable for gaining insights into the early stages of folding before their assembly into more complex tertiary structures in larger multi-domain proteins. They often fold and unfold sufficiently fast that their denatured and native states are in rapid equilibrium *in vivo* and so the *in vitro* studies are also directly relevant to biology. There could of course be high-energy intermediates, such as in Figure 1, which are cryptic. Two-state folding

without accumulating intermediates resurrected the possibility of nucleation mechanisms.

Chymotrypsin inhibitor 2: nucleation-condensation mechanism

We always perform a large number of mutations, but the Φ -value analysis of CI2 was exhaustive: 100 mutations at 45 of the 64 residues and a network of 11 double-mutant cycles (Itzhaki *et al.*, 1995). It revealed not only nucleation but discovered a new mechanism: the *nucleation-condensation* mechanism (Fersht, 1995; Itzhaki *et al.*, 1995). The single observed transition state for folding and unfolding consists of a structure in which an extended nucleus is formed, built around the single α -helix, which is being formed at the same time as the rest of the structure is condensing around it. Apart from one residue, all the Φ -values are fractional, approaching closer to 0, the further away from the diffuse nucleus. The physical-chemistry reasoning behind this is quite simple. None of the elements of regular secondary structure, such as the α -helix, are stable in the absence of the rest of the protein structure – as is generally found for proteins – and so those regions when separate from the rest of the structure are largely random in solution (Épand and Scheraga, 1968). For most proteins, the secondary structure needs to be stabilised by long-range interactions. Protein folding is, accordingly, such a cooperative process that the major transition state for folding of a domain is one in which the structure is largely formed. Nucleation-condensation is now a well-established general mechanism for the folding of single domains (Nolting and Agard, 2008; Kucic *et al.*, 2017).

The important features of nucleation-condensation are not just that the nucleus is large and extended but its structure is like a distorted form of the native structure where interactions are not uniform but weaken away from the nucleus. A generally useful pointer to the nucleation-condensation mechanism or a diffuse transition state is a Leffler/Brønsted plot of $\Delta\Delta G_{\ddagger-N}$ versus $\Delta\Delta G_{D-N}$ (Figure 16). As the Φ -values are mainly fractional, the plot is scattered around a linear regression of slope 0.7 with deviations for the higher and lower values of Φ . In contrast, the plot for barnase with its polarised transition state and Φ spread from 0 to 1 has the points scattered between lines of slope 0 and 1 (Itzhaki *et al.*, 1995).

Chymotrypsin inhibitor 2: computer simulations

CI2 is such a well-behaved system, small, and with so much experimental Φ -value data available that it stimulated and became a major test bed for computer simulation. I have had a long collaboration, beginning in 1994 (Fersht *et al.*, 1994; Li and Daggett, 1994), with Valerie Daggett, who had performed the first all-atom simulation of the unfolding of the bovine pancreatic trypsin inhibitor (Daggett and Levitt, 1992). Our collaboration agreement was that all her simulations were done blind without foreknowledge of our experimental data. Li and Daggett simulated the unfolding of CI2 at 498K, the simulated high temperature being necessary for the unfolding to be on the then accessible timescale of 2.2 ns (Li and Daggett, 1994) (the pathway does not change over a range of temperature (Day *et al.*, 2002)). The Φ -values from MD and experiment were very similar in the first study. As more experimental Φ -values became available, the good agreement remained. A simulation (Daggett *et al.*, 1996) gave a complete atomic-level description of the transition state and recapitulated all the experimental Φ -values (Itzhaki *et al.*, 1995). These simulations were then combined with further studies on the denatured state, including one of the first atomic views of a ‘random

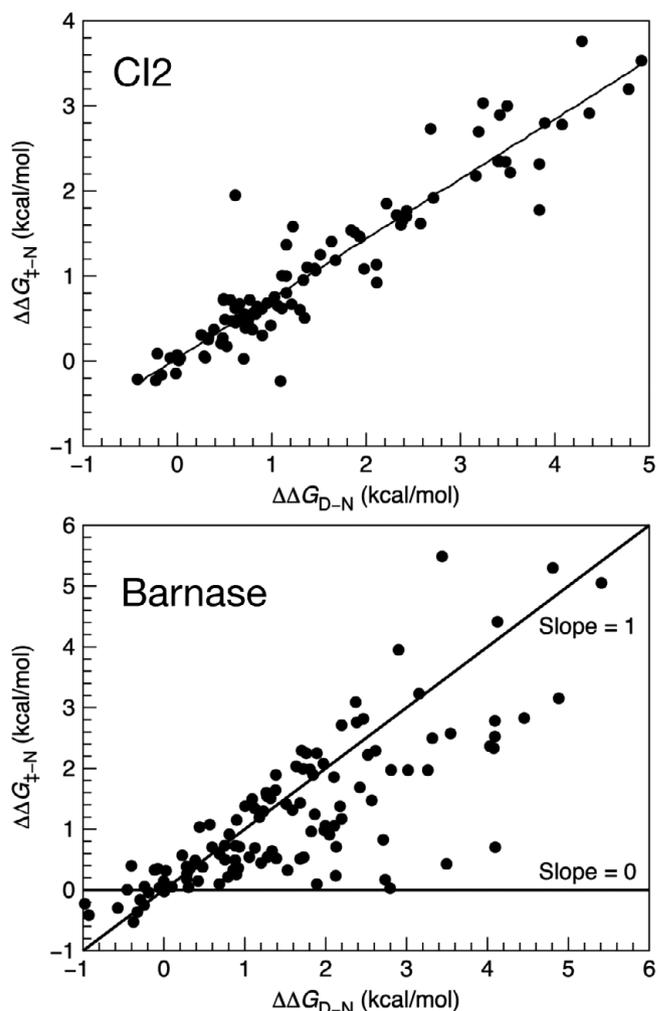


Figure 16. Brønsted (Leffler) plots of $\Delta\Delta G_{\ddagger-N}$ versus $\Delta\Delta G_{D-N}$ for CI2 which has a diffuse transition state and barnase which has a polarised one.

coil’ denatured state (Kazmirski *et al.*, 2001), and transition states (Li and Daggett, 1996; Kazmirski *et al.*, 2001), to give more detailed descriptions, reviewed by Fersht and Daggett (2002) and Daggett and Fersht (2003), Figure 17.

In multiple simulations of unfolding, single trajectories are distributed around an average ‘ensemble’ path (Day and Daggett, 2005). Simulations of folding and unfolding at the melting temperature showed that microscopic reversibility indeed holds (Day and Daggett, 2007). Overall, they found conformations in the transition state ensemble (TSE) have a probability of 0.5 to refold to the native state, with approximately 50% of the structures taken from the TSE refolding and the other 50% progressing to the denatured state (Day and Daggett, 2007). Further, simulations pointed to mutations that could speed up folding by relieving strain in the transition state, and one, Arg38→Phe48, was found that speeds up folding 40x to a $t_{1/2}$ of 400 μ s (Ladurner *et al.*, 1998). Thus, the MD-derived TSE consists of true transition states, validating the use of transition state theory underlying all Φ -value analyses, and also showing the power of simulation.

The results of multiple simulations of unfolding reconciled the ‘new view’ of folding on an energy landscape and the classical view of protein folding with a defined pathway – there is a statistically preferred pathway on a funnel-like average energy surface

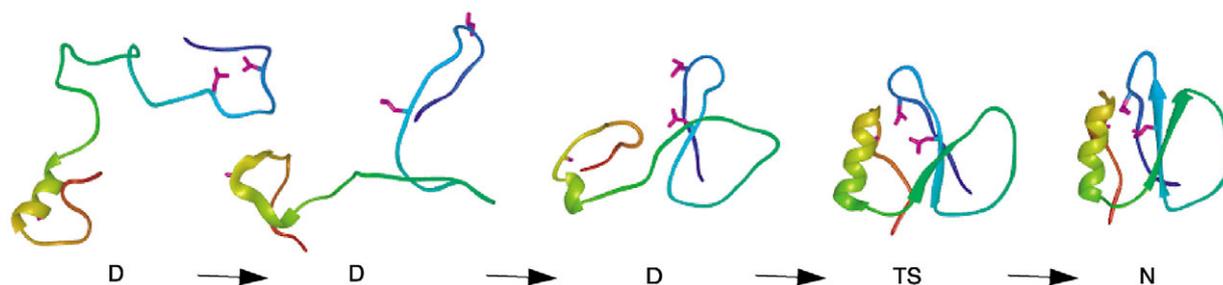


Figure 17. CI2 folding from experiment and simulation. An MD unfolding simulation from the native state N to the denatured state(s) D at 225°C shown in reverse. The structures are coloured from red at the N terminus to blue at the C terminus. The transition state is built around an extended nucleus, in which L49 and I57 pack against Ala16 (shown in magenta), towards the N terminus of the α -helix. There is flickering structure around Ala-16 in the denatured state.

(Lazaridis and Karplus, 1997). The funnelled nature of the energy landscape arising from Wolynes' minimal frustration principle (strong native bias) is consistent with unusual Φ -values being infrequent and that the transition state is a distorted version of the native state. Also, because the energy landscape is funnelled mutations, are not prone to change the structure of the native state (Oliveberg and Wolynes, 2005). CI2 Φ -values helped the theoreticians to clarify their views (Pande *et al.*, 1998).

CI2 occupies an important position in the development of protein folding studies because it was the first example of a single-domain protein showing two-state kinetics, the Φ -value analysis discovered the nucleation-condensation mechanism, and it stimulated so much theoretical advance.

Movement of TS on the energy landscape: Hammond and anti-Hammond effects

The transition state lies on a saddle point in the energy landscape and can move in a direction along the reaction coordinate, Hammond effect, or perpendicular to it, anti-Hammond, as the energetics are perturbed, Figures 3 and 18 (Jencks, 1985). We found both Hammond and anti-Hammond in folding transition states (Matouschek and Fersht, 1993; Matouschek *et al.*, 1995; Matthews and Fersht, 1995; Dalby *et al.*, 1998c) by comparing the extent of overall folding using Leffler/Brønsted plots of ΔG_{\ddagger}^0 versus ΔG_{D-N} or β_T with Φ -values for local structure (Matthews and Fersht, 1995; Fersht and Sato, 2004). A Leffler/Brønsted plot of successive mutations in helix 1 of barnase has a slope for unfolding of -0.09 for mutations with $\Delta G_{D-N} < 2$ kcal/mol, showing that it is $\sim 90\%$ folded in the transition state, but for $\Delta G_{D-N} > 3$ kcal/mol, the slope steepens to -0.6 , so that the helix is only $\sim 60\%$ folded. The overall position of the transition state moves closer to that of the native structure as it becomes less stable, measured by β_T , the Hammond effect, but the helix itself follows anti-Hammond behaviour and moves away from native. The anti-Hammond could result from a changing balance in parallel pathways (Matthews and Fersht, 1995) or true movement perpendicular. Simulation supports the latter (Daggett *et al.*, 1998). Movement of the transition state on large destabilising mutations signals caution in interpreting changes in Φ for them. Importantly, it points to how a series of mutations in a family of homologous proteins can lead to changes of mechanism.

Engrailed homeodomain: framework mechanism

The Engrailed homeodomain (EnHD) is a 61-residue 3-helix bundle protein. (Mayor *et al.*, 2000; Banachewicz *et al.*, 2011). In

addition, a combination of NMR, X-ray-crystallography, xX-ray-scattering, and various spectroscopic techniques on wild-type and mutant protein have also been in the rare position of being able to describe the structures of the denatured state, D_{phys} , and an intermediate at atomic resolution. These structural studies combined with Φ -values and molecular dynamics simulations provide a detailed description of its folding pathway from ns to μ s (Mayor *et al.*, 2000, 2003a, 2003b; Stollar *et al.*, 2003; DeMarco *et al.*, 2004; Religa *et al.*, 2005; Huang *et al.*, 2008; McCully *et al.*, 2008; Neuweiler *et al.*, 2010; Banachewicz *et al.*, 2011; Nasedkin *et al.*, 2015). Simulations of folding and unfolding pathways obey microscopic reversibility (McCully *et al.*, 2008).

The protein folds from the intermediate via a framework mechanism. EnHD has a very stable helix 1 which is up to $\sim 40\text{--}50\%$ α -helical in the absence of the rest of the protein, and helices 2 and 3 together form a helix-turn-helix motif which is not only structured in that folding intermediate (Mayor *et al.*, 2003a) but also stable as an independent sequence (Religa *et al.*, 2007). This intermediate is the most stable denatured state under conditions that favour folding, the more unfolded form being less stable, and its structure has been determined by NMR (Religa *et al.*, 2005). Φ -values show the final rate-determining transition state is the docking of helix 1 onto to the structure helices 2 and 3 to form the hydrophobic core (Figure 19; Mayor *et al.*, 2003a).

Homeodomain family: pointer to a unifying underlying mechanism

Slide from nucleation-condensation to framework across a family

Members of the same family of proteins having the same overall fold but with different sequences and secondary structural propensities can provide important information, especially from Φ -analysis (Im7, Im9 (Friel *et al.*, 2003); Ig-like (Geierhaas *et al.*, 2004; Lappalainen *et al.*, 2008); SH3 domains (Martinez and Serrano, 1999; Guerois and Serrano, 2000); protein L (Kim *et al.*, 2000), and more general discussions (Zarrine-Afsar *et al.*, 2005; Brunori *et al.*, 2012).

Three members of the homeodomain-like protein family that share the same overall topology with EnHD: human TRF1 Myb domain (hTRF1); human RAP1 Myb domain (hRAP1); and c-Myb-transforming protein (c-Myb) have decreasing propensity for α -helix formation in helix 1 (Figure 20) and helices 2 and 3 do not form independently stable helix-turn-helix motifs. These proteins vary widely in sequence, just having fold homology. There is a spectrum of folding processes that spans the complete transition

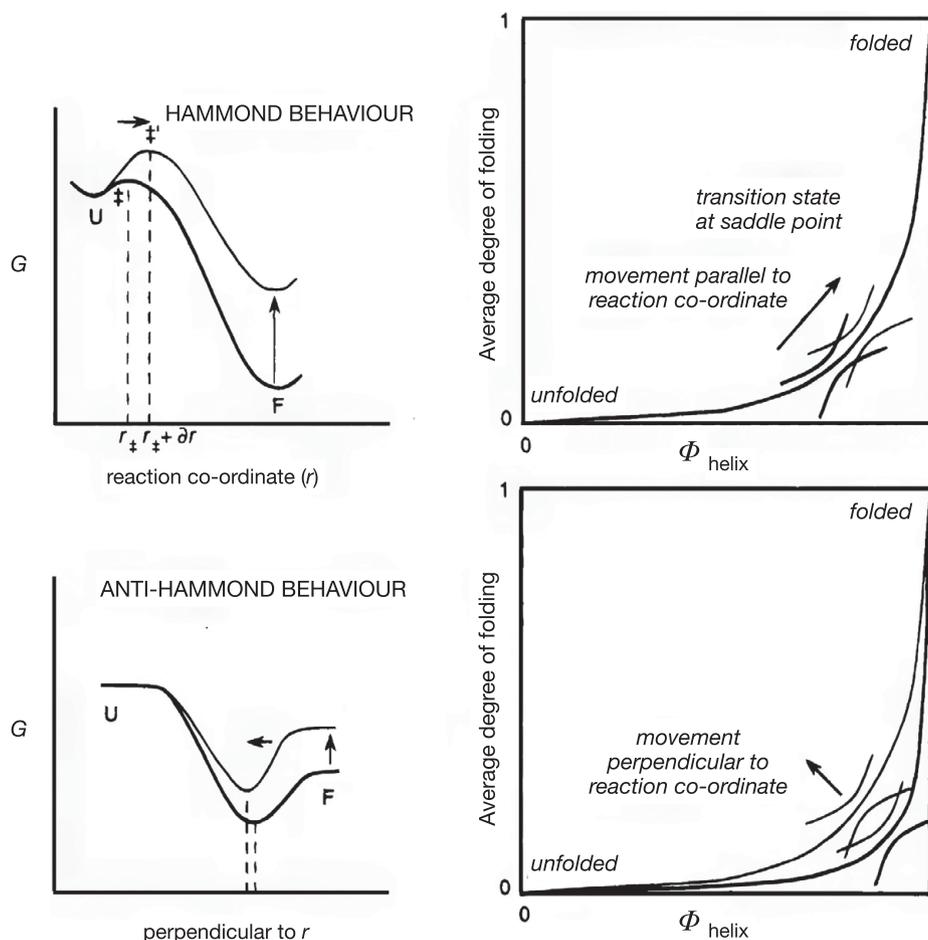


Figure 18. Hammond and anti-Hammond behaviour for the folding of a protein. *Left top:* Conventional Hammond behaviour as the transition state moves closer to the folded state (F) along the reaction coordinate with increasing destabilisation of F. *Left bottom:* Cross-section of the energy profile perpendicular to the reaction coordinate at the transition state. Anti-Hammond behaviour as the transition state moves closer to the unfolded state in a direction perpendicular to the reaction coordinate on destabilisation of F see (Jencks, 1985). *Right:* Correlation diagrams of the average degree of folding, say β_T , for the whole protein and Φ , the degree of formation of the helix, in the transition state. *Top right:* Average degree of folding in the transition state increases as the transition state moves along the reaction coordinate closer to F as the protein is destabilised by a mutation. *Bottom right:* Concurrent with the movement of the transition state along the reaction coordinate in the direction of F as the protein is destabilised by a mutation, there is anti-Hammond movement perpendicular to the reaction coordinate that leads to the helix becoming less folded and Φ decreases (Matthews and Fersht, 1995).

from framework to nucleation-condensation mechanism as the helical propensity decreases, Figure 21 (Gianni *et al.*, 2003). The common factor in their mechanisms is that the transition state for (un)folding is expanded and very native-like, with the proportion and degree of formation of secondary and tertiary interactions varying. It appears that framework and nucleation-condensation are different manifestations of an underlying common mechanism, Figure 21 (Daggett and Fersht, 2003; Gianni *et al.*, 2003).

Folding close to the speed limit

Pit1, the 63-residue homeodomain from pituitary-specific transcription factor, folds via an intermediate in wider separated phases than EnHD of $t_{1/2}$ 2.3 and 46 μs (Banachewicz *et al.*, 2011), allowing Φ -values to be measured for both phases (Banachewicz *et al.*, 2011). Its helix-turn-helix motif does not independently fold but is folded in the intermediate, docked to a misfolded helix 1, which rearranges to fold correctly. Pit1 is on the slide from framework in the EnHD folding to nucleation-condensation for Myb, TRF1 and RAP1.

The folding rate constant of $3 \times 10^5 \text{ s}^{-1}$ for the fast phase decreases with increasing viscosity and is only slightly sensitive to

mutation or denaturant concentration. The formation of the intermediate is partly rate-limited by chain diffusion and partly by an energy barrier to give a very diffuse transition state. The process is rather like the association of barnase with its protein inhibitor barstar which proceeds via an encounter complex that is diffusion-limited, relatively insensitive to mutations and then precisely docks and makes specific interactions in a slower step (Schreiber and Fersht, 1995, 1996). The folding is approaching the downhill-folding scenario of energy landscape theory (Gelman and Gruebele, 2014).

The free energy barrier that separates the native and denatured states ensembles in the energy landscape model may disappear under extreme conditions that greatly energetically favour the native state (Bryngelson *et al.*, 1995), similar to extreme Hammond behaviour for the movement of transition states in covalent chemistry, Figure 18, where the transition state moves closer in structure to the denatured state as the product becomes more stable (Hammond, 1955). Under these conditions, the protein folds downhill energetically. The transition-state energy barrier reappears as conditions change to stabilise the denatured state ensemble, such as going through the thermal or denaturant

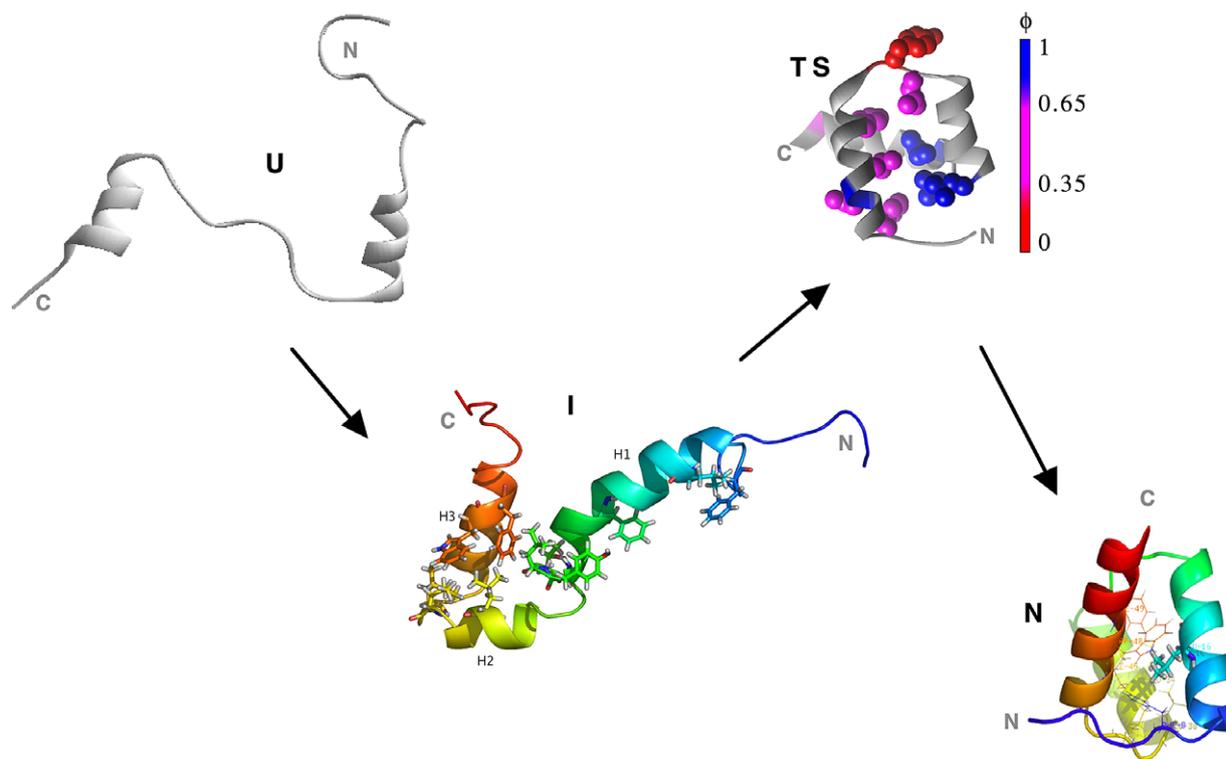


Figure 19. Folding pathway of Engrailed Homeodomain (EnHD) from experiment and simulation. From *right to left*: native state (NS) structure solved by nuclear magnetic resonance and X-ray crystallography; transition state (TS) by Φ -analysis of secondary structure (colour-coded from $\Phi = 0$, red, to $\Phi = 1$, blue); the folding intermediate (I) stably generated by protein engineering and solved by NMR; the denatured state (U), under conditions that favour folding, simulated using molecular dynamics; and the entire unfolding pathway was simulated by molecular dynamics.

unfolding transitions. The finding of very fast folding small domains, ‘miniproteins’ that fold on the μs time scale or faster led to increased interest as what happens to pathways at folding close to the speed limit (Kubelka *et al.*, 2004; Gelman and Gruebele, 2014). Barriers of $<3k_{\text{B}}T$ ($<1.8 \text{ kcal mol}^{-1}$ at 298 K) are suggested to be consistent with this type of downhill folding (Carter *et al.*, 2013; Prigozhin and Gruebele, 2013). However, ‘downhill folding on a rough energy landscape versus rapid folding through very shallow intermediates is in the eye of the beholder’ (Gelman and Gruebele, 2014). All the states along the pathway/landscape are ensembles of structures (Figure 8). There is a residual native and non-native structure in the denatured state, and this coexists with folding intermediates and the native structure in varying proportions with changing conditions. The folded state is dynamic, with regions locally unfolding as demonstrated by hydrogen-deuterium exchange (Englander *et al.*, 1997; Englander, 2023). The energy landscape has many local minima, which can contribute to kinetics when the transition state energy barrier is low. These problems are exacerbated for the small fast-folding domains because their folding equilibrium and activation energies are often low and the structure of domains taken from their parent is sensitive to the choice of domain boundaries.

Transition states across PSBD family: nucleation-condensation in very fast folding

The more thermostable two-helix bundle PSBD from *B. stearothermophilus* (E3BD) folds cooperatively and very rapidly, and its separated constituent α -helical regions have little helical tendency, showing fast folding does not require the docking of

preformed elements (Spector *et al.*, 1998, 1999a, 1999b; Spector and Raleigh, 1999). Φ -value analysis at 325K by *T*-jump relaxation kinetics (Ferguson *et al.*, 2005) and at 298K by rapid mixing and some *T*-jump (Ferguson *et al.*, 2006) show a nucleation-condensation mechanism, which has a very diffuse transition state but with helix 2 the most structured. There is good consistency with calculated values from MD simulation.

Comparison of Φ -values with two other members of the PSBD family that have significant sequence identity but different helix-forming propensities, POB, from *Pyrobaculum aerophilum* (Sharpe *et al.*, 2008) and BBL (Neuweiler *et al.*, 2009), Figure 22, provides information about conservation of folding mechanism in closely related, very fast folding, proteins. They all fold via nucleation-condensation, with Φ -values summarised in Figure 23. There are differences in that folding of E3BD and POB nucleates in Helix 2 but interactions in the folding transition state of BBL is more evenly dispersed across the structure, perhaps because of the high helical propensity of its Helix 1 (Neuweiler *et al.*, 2009). The folding rate constants for E3BD, BBL, and POB at 298 K are $27,500 \pm 500$, $124,000 \pm 5000 \text{ s}^{-1}$, and $210,000 \pm 5000 \text{ s}^{-1}$, respectively, and follow the predicted helical propensities sites in the second helix. An increased helical propensity at the nucleation site appears to stabilise the folding nucleus and results in an increased folding rate constant.

Other examples with Φ -values

Φ -analysis has now been applied by many groups to a large number of proteins to illuminate a range of processes and structures in the folding, assembly, and activity of proteins. Alm *et al.* (2002) used

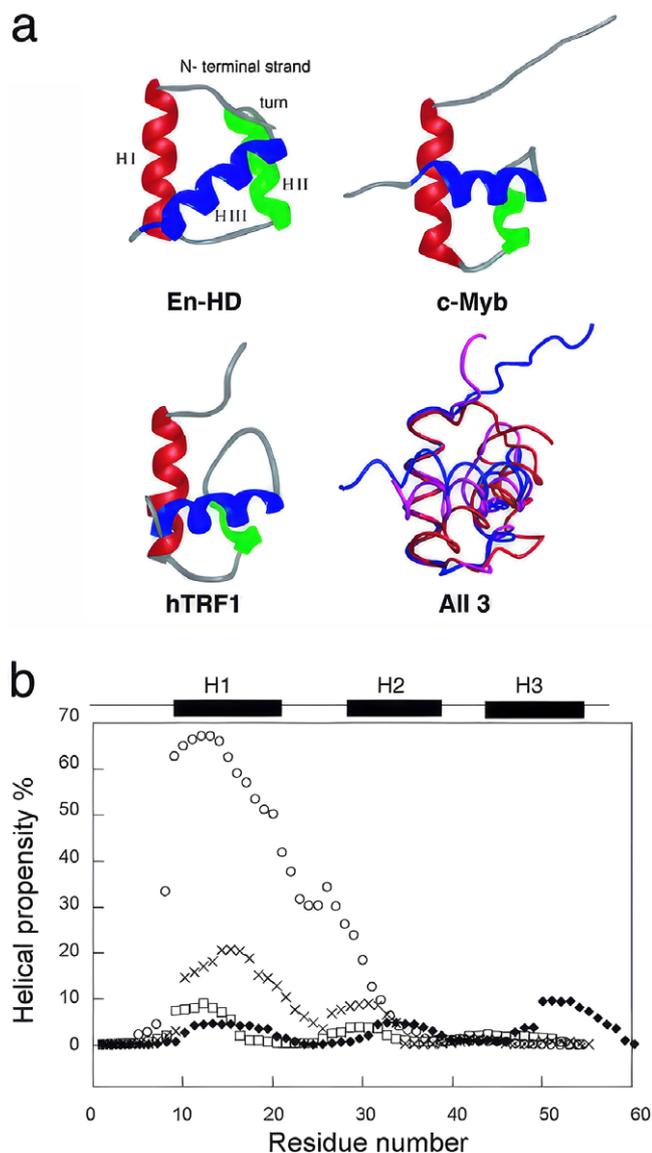


Figure 20. (a) Structures and (b) secondary structure prediction for En-HD (o), c-Myb (x), hRAP1 (◆), and hTRF1 (□) (Gianni *et al.*, 2003).

published data on 19 proteins with Φ -values to devise a simple model for folding. For over half of these, the theory reproduced Φ with correlation coefficients between 0.41 and 0.88. They classified transition-state structures into three categories. (1) *Small proteins with polarised transition states* include Protein L; Protein G; src; spectrin; and Sso7d SH3 domains. (2) *Large proteins with compact subdomains* include barnase; cheY; tenascin; titin; fibronectin (the tenth type III domain repeat of fibronectin); and U1A spliceosomal protein. (3) *Proteins with diffuse transition states*, which include: CI2; FKBP12 (FK501-binding protein); λ repressor; Suc1; muscle acylphosphatase; procarboxypeptidase; ribosomal protein S6; and villin headpiece. The examples with diffuse transition states correspond to the CI2 end of the nucleation-condensation mechanism, which slides to the polarised end for some of the polarised states.

Proteins, including some of the above with their sources, that have been subjected to Φ -analysis are in this by no means complete list. I have indicated (nc) for some that appear to fold by nucleation condensation and (fw) by framework. Monomeric λ -repressor

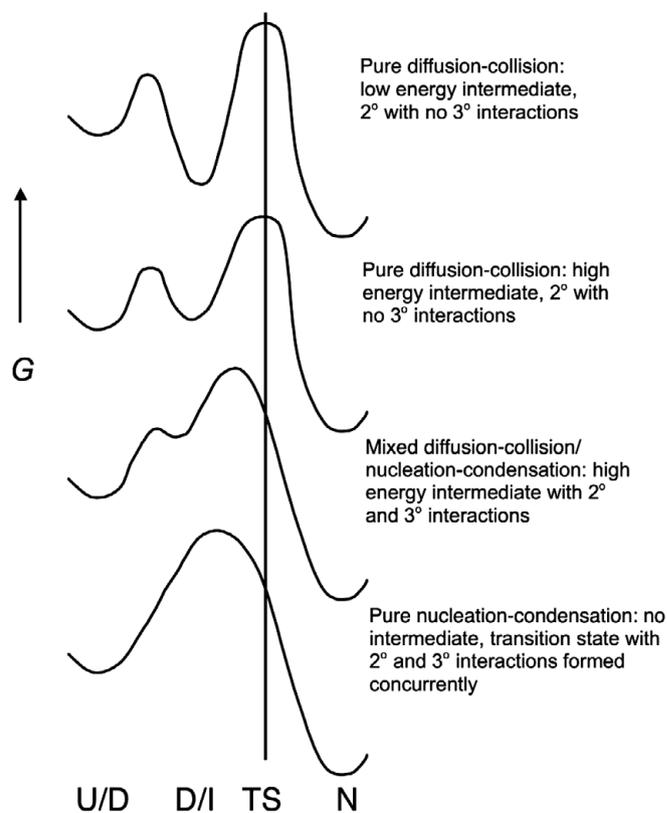


Figure 21. The slide from framework to nucleation-condensation (Gianni *et al.*, 2003).

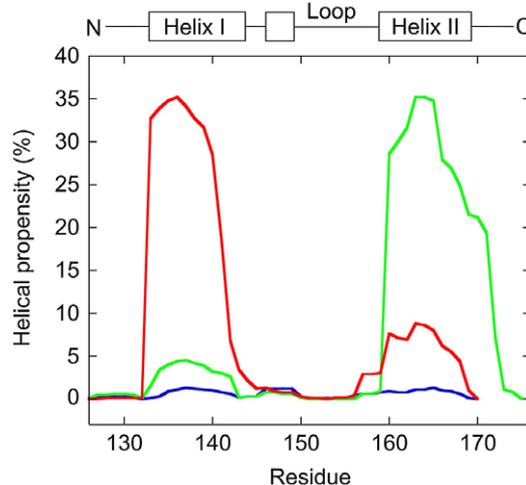


Figure 22. Calculated helical propensities of BBL (red), E3BD (blue), and POB (green) sequences (Neuweiler *et al.*, 2009).

(Burton *et al.*, 1998), ADA2H (nc) (Villegas *et al.*, 1998; Kucic *et al.*, 2017), acyl-coA binding protein (nc) (Kragelund *et al.*, 1999), SH3 domains (α -spectrin (nc), src) (Grantcharova *et al.*, 1998; Martinez and Serrano, 1999; Guerois and Serrano, 2000), SH3 domain from Grb2 (nc) (Troilo *et al.*, 2018), acylphosphatase (nc) (Chiti *et al.*, 1999), Im7 and Im9 (Friel *et al.*, 2003; Paci *et al.*, 2004; Bartlett and Radford, 2010), NTL9 domain of L9 (nc) (Anil *et al.*, 2005; Sato *et al.*, 2017), cheY (1 domain nc) (Lopez-Hernandez and Serrano, 1996), Sod1 (Yang *et al.*, 2018),

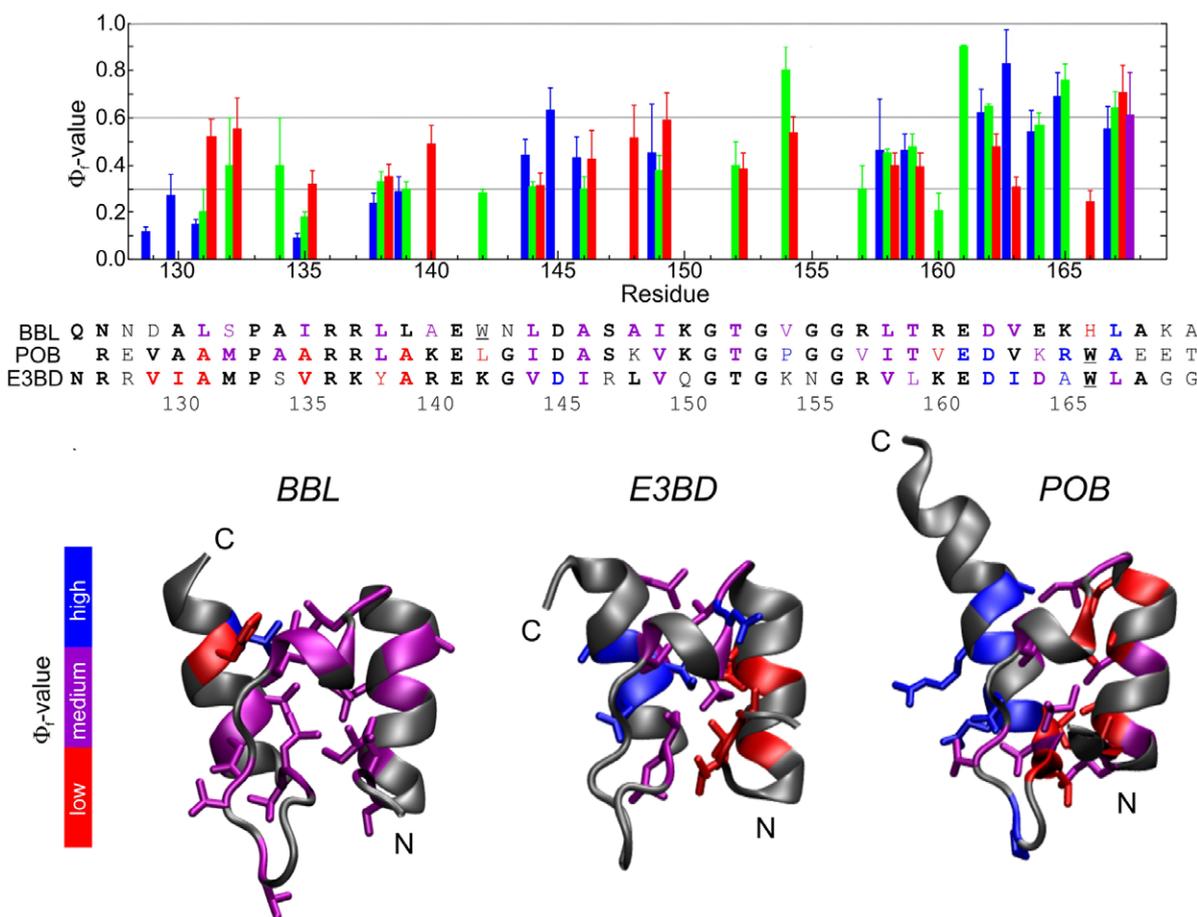


Figure 23. Φ -value analysis of PSBD family members. *Top:* Φ_i -values for BBL (red bars), E3BD (blue bars) and POB (green bars). *Middle:* Sequences aligned with similar residues in boldface. Φ_i -values are indicated using the colour code at bottom left grouped into 'low' ($0.0 < \Phi_i < 0.3$), 'medium' ($0.3 < \Phi_i < 0.6$), and 'high' ($0.6 < \Phi_i \leq 1.0$), and *bottom* mapped onto the sequences and native-state structures (modified from Neuweiler *et al.*, 2009).

S6 (Otzen and Oliveberg, 2002; Lindberg *et al.*, 2006), U1a (nc) (Ternstrom *et al.*, 1999), azurin (Wilson and Wittung-Stafshede, 2005; Zong *et al.*, 2006), apo-flavodoxin (Campos *et al.*, 2004; Muralidhara *et al.*, 2005; Bueno *et al.*, 2006; Lopez-Llano *et al.*, 2006; Homouz *et al.*, 2009; Stagg *et al.*, 2010; Galano-Frutos *et al.*, 2022), WW domains (polarised) (Jager *et al.*, 2001; Petrovich *et al.*, 2006; Dave *et al.*, 2016), villin headpiece (Cho *et al.*, 2010), SH2 domains (Visconti *et al.*, 2019; Toto *et al.*, 2022), HYPA/FBP11 FF domain (bn) (Jemth *et al.*, 2005), PTP-BL, PDZ2 and PSD-95 PDZ3 (3-state) (Calosci *et al.*, 2008), ubiquitin (nc) (Went and Jackson, 2005; Varnai *et al.*, 2008), RNaseA (nc) (Font *et al.*, 2006), γ ACBP and 3 variants of bACBP (Teilum *et al.*, 2005), B domain of Protein A (nc) (Sato *et al.*, 2006), α -lactalbumin (Chedad *et al.*, 2005), R15 (nc) R16 (fw) R17 (fw) (domains of chicken brain α -spectrin) (Wensley *et al.*, 2009), TI I27 (Fowler and Clarke, 2001), TNfn3 (Geierhaas *et al.*, 2004), LysM domain (Nickson *et al.*, 2008), FKB12 (nc) (Main *et al.*, 2001), apocytchrome b562 (Zhou *et al.*, 2005), SAP (Dodson and Arbely, 2015), knotted proteins (Mallam *et al.*, 2008; Jackson *et al.*, 2017), raf (polarised) (Campbell-Valois and Michnick, 2007), four-helix HYPA/FPB11 (Jemth *et al.*, 2005), tumour suppressor P16 (Tang *et al.*, 2003), barstar (nc) (Nolting *et al.*, 1997), p13Suc1 (nc) (Schymkowitz *et al.*, 2000), arc repressor (Srivastava and Sauer, 2000), BPTI (Bulaj and Goldenberg, 2001), TAPLLR (Kelly *et al.*,

2014), tumour suppressor p53 (Wang and Fersht, 2015), and KIX domain (Troilo *et al.*, 2017).

Φ -analysis has been applied successfully to the folding of transmembrane proteins (Otzen, 2011; Booth, 2012; Paslawski *et al.*, 2015) and includes processes with receptors and gating (Cymes *et al.*, 2002; Mitra *et al.*, 2004; Cadugan and Auerbach, 2007; Aleksandrov *et al.*, 2009; Edelstein and Changeux, 2010). Φ -analysis is particularly useful for studying the folding of intrinsically disordered proteins on binding to folded partners (Karlsson *et al.*, 2012; Dogan *et al.*, 2013, 2014; Rogers *et al.*, 2014; Shammas *et al.*, 2016; Karlsson *et al.*, 2019; Toto *et al.*, 2019; Karlsson *et al.*, 2020; Malagrino *et al.*, 2020; Toto *et al.*, 2020; Karlsson and Jemth, 2021) and for proteins where mechanical force mimics their function *in vivo* (Best *et al.*, 2002; Best and Clarke, 2002; Fowler *et al.*, 2002). It has been extended to RNA folding (Silverman and Cech, 2001; Young and Silverman, 2002; Kim and Shin, 2010; Pereyaslavets and Galzitskaya, 2015) and DNA aptamers (Lawrence *et al.*, 2014).

The robustness and validity of Φ -analysis: Φ - Φ plots

The above examples show the wide and successful application of Φ -analysis. There have been criticisms of Φ -analysis, which have been critiqued by Gianni and Jemth (Gianni and Jemth, 2014).

They have a nice argument on how plots of Φ versus Φ for processes in common demonstrate the robustness of Φ -analysis. Such plots on homologous proteins are used to compare folding transition states (Calosci *et al.*, 2008; Wensley *et al.*, 2009; Wensley *et al.*, 2010). Sequences of identical proteins, such as circular permutants and circularised proteins, or homologous proteins with high sequence identity are aligned and values of Φ at the same position plotted for one against in the other, as in Figure 24. The probability that the pairs in each are not linearly related, P , is infinitesimal, consistent with their containing structural information. Provided that mutations are chosen as described and analysed in the first Φ -value paper (Matouschek *et al.*, 1989 and earlier (Fersht *et al.*, 1987), and too high or too low changes in ΔG_{D-N} not used (Fersht and Sato, 2004), Φ -value analysis is robust. The weak, medium, and strong categorisation provides adequate constraints for simulation.

Φ -value analysis has stood the test of time over three decades and we have gone from knowing virtually nothing about the fine structure of transition states for folding in the late 1980s to having a wealth of detailed information about many individual proteins. But can we draw generalisations?

The expanded transition state as a unifying mechanism for domain folding

Proteins have evolved for optimal function *in vivo* and not the greatest stability or fastest folding. Protein activity often requires flexibility and dynamics for function, a stability that is high enough but not too high to prevent turnover where necessary, a rate of unfolding for some that is sufficiently slow to inhibit aggregation via unfolding, and a trade-off between overall stability and local instability of binding and active sites. For example, simple mutations can change the rate constants for the folding of CI2 over three orders of magnitude: wild-type folds at 25°C at 56 s^{-1} , the double mutant A16G/I57A in the folding nucleus at 2.4 s^{-1} , and R48F at 2300 s^{-1} . The active site of barnase is a source of instability (Meiering *et al.*, 1992) and mutations elsewhere can greatly stabilise it without loss of activity (Serrano *et al.*, 1993). Those factors will conspire to complicate the formulation of simple models for folding and its kinetics and cause exceptions to mechanisms.

'In their search for order, chemists invented Brønsted and Hammett correlations and other free energy relationships' so begins Jencks in his review of the movement of transition states across energy landscapes (Jencks, 1985). So, here is an attempt to bring some order, bearing in mind that there will be many exceptions. The unifying feature across the folding of most domains that comes from Φ -value analysis is that the highest energy transition state is an expanded, distorted form of the native structure, Figure 25 (Fersht, 2000). It varies from the pure nucleation-condensation mechanism at one extreme with mainly low to mid-range Φ -values to framework mechanisms at the other extreme with highly polarised transition states and Φ -values from 0 to 1. The expanded nature of the transition state and its observed malleability both naturally across protein families and unnaturally on protein engineering accommodates the slide from pure nucleation condensation to framework mechanism, Figure 26.

Envoi

My research career has spanned seven decades that have seen ground-breaking innovations, beginning in the 1960s with the first

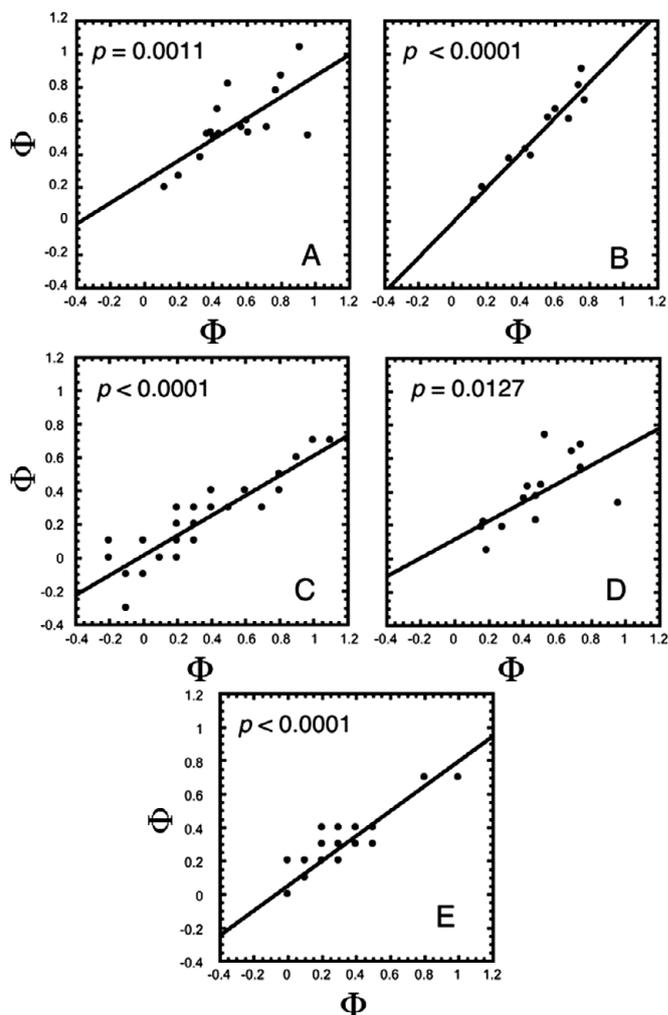


Figure 24. Φ - Φ plots demonstrating the robustness of Φ -values (Gianni and Jemth, 2014). (a) PDZ domains (Gianni *et al.*, 2007; Calosci *et al.*, 2008), (b) Circularly permuted PDZ domain (Ivarsson *et al.*, 2009), (c) circularization of LysM domain (Nickson *et al.*, 2008), (d) tryptophan as a fluorescence probe inserted in turn into each of the three helices of the B-domain of Protein A (Sato *et al.*, 2006), and (e) the spectrin R16 domain with different neighbouring domains (Batey and Clarke, 2008). The P -value is the probability that the two variables are not correlated.

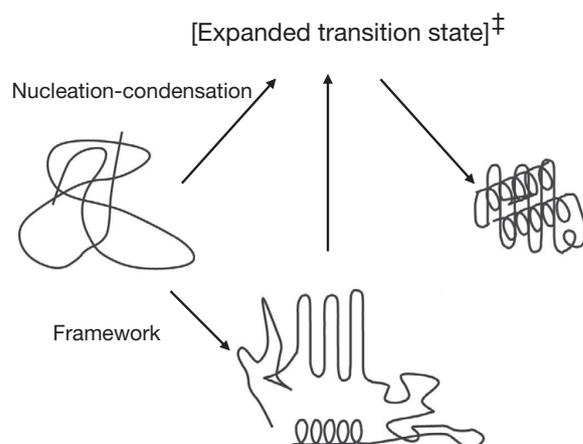


Figure 25. A transition state that is an expanded, distorted, native structure being common to framework and nucleation-condensation mechanisms.

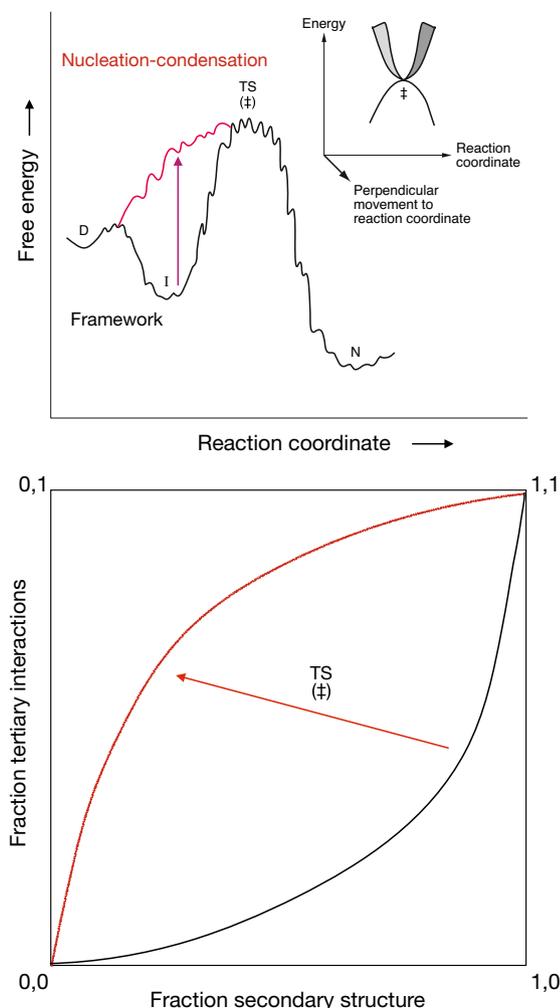


Figure 26. Combining elements of Figures 19 and 21 to illustrate how movement of the expanded transition state on an energy landscape according to the classical principles of physical-organic chemistry unifies the slide between a diffuse nucleation-condensation state and the framework mechanism via a polarised transition state. *Top:* Reaction coordinate diagram for a framework mechanism with preformed secondary structure in a low energy intermediate that slides to nucleation-condensation as the secondary structure becomes less stable and requires tertiary interactions to stabilise it. The transition state can move along and perpendicular to the reaction coordinate according to Hammond and anti-Hammond effects, respectively. Both mechanisms involve an extended network of long-range native-like tertiary interactions in the expanded transition state. *Bottom:* Correlation diagram of formation of native secondary and tertiary interactions illustrating the above.

high-resolution structures of proteins from X-ray crystallography, followed by recombinant DNA technology, DNA sequencing, new enabling biological and biophysical technologies, and advances in computation methods from simulation to machine learning today. It has been my privilege and pleasure to have been a participating protein scientist using directly or indirectly all these advances as they were introduced (Fersht, 2008, 2021). Over the same period, we have gone from being just observers of the properties of proteins to being able to manipulate their structures and activities. We have progressed from the pathway of protein folding being a mysterious unknown to using those methodologies to solve the folding pathways of small domains at atomic resolution. There is much more experimental work to be done on more complex systems, where Φ -values will continue to provide otherwise inaccessible information. I hope that the Φ -values gathered by us all will be used as benchmarks for computation far into the future. It has been a

marvellous time to have been a protein scientist. The best is still to come as we progress to unravelling the folding and mechanisms of complex protein systems and combine our acquired experimental knowledge with improved computation to design novel, functional proteins.

References

- Aleksandrov AA, Cui LY and Riordan JR (2009) Relationship between nucleotide binding and ion channel gating in cystic fibrosis transmembrane conductance regulator. *Journal of Physiology-London* **587**(12), 2875–2886. <https://doi.org/10.1113/jphysiol.2009.170258>.
- Alm E, Morozov AV, Kortemme T and Baker D (2002) Simple physical models connect theory and experiment in protein folding kinetics. *Journal of Molecular Biology* **322**(2), 463–476. [https://doi.org/10.1016/s0022-2836\(02\)00706-4](https://doi.org/10.1016/s0022-2836(02)00706-4).
- Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* **181**, 4096.
- Anil B, Sato S, Cho JH and Raleigh DP (2005) Fine structure analysis of a protein folding transition state; Distinguishing between hydrophobic stabilization and specific packing. *Journal of Molecular Biology* **354**(3), 693–705. <https://doi.org/10.1016/j.jmb.2005.08.054>.
- Baldwin RL (1994) Protein folding. Matching speed and stability. *Nature* **369** (6477), 183–184. <https://doi.org/10.1038/369183a0>.
- Baldwin RL (1995) The nature of protein folding pathways: The classical versus the new view. *Journal of Biomolecular NMR* **5**(2), 103–109. <https://doi.org/10.1007/BF00208801>.
- Baldwin RL (2017) Clash between energy landscape theory and foldon-dependent protein folding. *Proceedings of the National Academy of Sciences USA* **114**(32), 8442–8443. <https://doi.org/10.1073/pnas.1709133114>.
- Banachewicz W, Johnson CM and Fersht AR (2011) Folding of the Pit1 homeodomain near the speed limit. *Proceedings of the National Academy of Sciences USA* **108**(2), 569–573. <https://doi.org/10.1073/pnas.1017832108>.
- Banachewicz W, Religa TL, Schaeffer RD, Daggett V and Fersht AR (2011) Malleability of folding intermediates in the homeodomain superfamily. *Proceedings of the National Academy of Sciences USA* **108**(14), 5596–5601. <https://doi.org/10.1073/pnas.1101752108>.
- Bartlett AI and Radford SE (2010) Desolvation and development of specific hydrophobic core packing during Im7 folding. *Journal of Molecular Biology* **396**(5), 1329–1345. <https://doi.org/10.1016/j.jmb.2009.12.048>.
- Batey S and Clarke J (2008) The folding pathway of a single domain in a multidomain protein is not affected by its neighbouring domain. *Journal of Molecular Biology* **378**(2), 297–301. <https://doi.org/10.1016/j.jmb.2008.02.032>.
- Baxa MC and Sosnick TR (2022) Engineered metal-binding sites to probe protein folding transition states: Psi analysis. *Methods in Molecular Biology* **2376**, 31–63. https://doi.org/10.1007/978-1-0716-1716-8_2.
- Best RB and Clarke J (2002) What can atomic force microscopy tell us about protein folding? *Chemical Communications (Cambridge)* **3**, 183–192. <https://doi.org/10.1039/b108159b>.
- Best RB, Fowler SB, Toca-Herrera JL and Clarke J (2002) A simple method for probing the mechanical unfolding pathway of proteins in detail. *Proceedings of the National Academy of Sciences USA* **99**(19), 12143–12148. <https://doi.org/10.1073/pnas.192351899>.
- Best RB and Hummer G (2016) Microscopic interpretation of folding phi-values using the transition path ensemble. *Proceedings of the National Academy of Sciences of the United States of America* **113**(12), 3263–3268. <https://doi.org/10.1073/pnas.1520864113>.
- Bodenreider C and Kiefhaber T (2005) Interpretation of protein folding psi values. *Journal of Molecular Biology* **351**(2), 393–401. <https://doi.org/10.1016/j.jmb.2005.05.062>.
- Bond CJ, Wong KB, Clarke J, Fersht AR and Daggett V (1997) Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: Description of the folding pathway. *Proceedings of the National Academy of Sciences USA* **94**(25), 13409–13413. <https://doi.org/10.1073/pnas.94.25.13409>.
- Booth PJ (2012) A successful change of circumstance: A transition state for membrane protein folding. *Current Opinion in Structural Biology* **22**(4), 469–475. <https://doi.org/10.1016/j.sbi.2012.03.008>.

- Brønsted JN and Pedersen K** (1924) The catalytic disintegration of nitramide and its physical-chemical relevance. *Zeitschrift Fur Physikalische Chemie--Stoichiometrie Und Verwandtschaftslehre* **108**(3/4), 185–235.
- Brunori M, Gianni S, Giri R, Morrone A and Travaglini-Allocatelli C** (2012) Morphogenesis of a protein: Folding pathways and the energy landscape. *Biochemical Society Transactions* **40**(2), 429–432. <https://doi.org/10.1042/BST20110683>.
- Bryan PN** (2000) Protein engineering of subtilisin. *Biochimica et Biophysica Acta* **1543**(2), 203–222. [https://doi.org/10.1016/s0167-4838\(00\)00235-1](https://doi.org/10.1016/s0167-4838(00)00235-1).
- Bryngelson JD, Onuchic JN, Succi ND and Wolynes PG** (1995) Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* **21**(3), 167–195. <https://doi.org/10.1002/prot.340210302>.
- Buchner J and Kiefhaber T** (1990) Folding pathway enigma. *Nature* **343**(6259), 601–602. <https://doi.org/10.1038/343601b0>.
- Bueno M, Ayuso-Tejedor S and Sancho J** (2006) Do proteins with similar folds have similar transition state structures? A diffuse transition state of the 169 residue apoflavodoxin. *Journal of Molecular Biology* **359**(3), 813–824. <https://doi.org/10.1016/j.jmb.2006.03.067>.
- Bulaj G and Goldenberg DP** (2001) Phi-values for BPTI folding intermediates and implications for transition state analysis. *Nature Structural Biology* **8**(4), 326–330. <https://doi.org/10.1038/86200>.
- Burton RE, Myers JK and Oas TG** (1998) Protein folding dynamics: Quantitative comparison between theory and experiment. *Biochemistry* **37**(16), 5337–5343. <https://doi.org/10.1021/bi980245c>.
- Cadugan DJ and Auerbach A** (2007) Conformational dynamics of the alpha M3 transmembrane helix during acetylcholine receptor channel gating. *Biophysical Journal* **93**(3), 859–865. <https://doi.org/10.1529/biophysj.107.105171>.
- Cafilisch A and Karplus M** (1995) Acid and thermal denaturation of barnase investigated by molecular dynamics simulations. *Journal of Molecular Biology* **252**(5), 672–708. <https://doi.org/10.1006/jmbi.1995.0528>.
- Calosci N, Chi CN, Richter B, Camilloni C, Engstrom A, Eklund L, Travaglini-Allocatelli C, Gianni S, Vendruscolo M and Jemth P** (2008) Comparison of successive transition states for folding reveals alternative early folding pathways of two homologous proteins. *Proceedings of the National Academy of Sciences of the United States of America* **105**(49), 19241–19246. <https://doi.org/10.1073/pnas.0804774105>.
- Campbell-Valois FX and Michnick SW** (2007) The transition state of the ras binding domain of Raf is structurally polarized based on phi-values but is energetically diffuse. *Journal of Molecular Biology* **365**(5), 1559–1577. <https://doi.org/10.1016/j.jmb.2006.10.079>.
- Campos LA, Bueno M, Lopez-Llano J, Jimenez MA and Sancho J** (2004) Structure of stable protein folding intermediates by equilibrium phi-analysis: The apoflavodoxin thermal intermediate. *Journal of Molecular Biology* **344**(1), 239–255. <https://doi.org/10.1016/j.jmb.2004.08.081>.
- Carter JW, Baker CM, Best RB and De Sancho D** (2013) Engineering folding dynamics from two-state to downhill: Application to lambda-repressor. *Journal of Physical Chemistry B* **117**(43), 13435–13443. <https://doi.org/10.1021/jp405904g>.
- Carter PJ, Winter G, Wilkinson AJ and Fersht AR** (1984) The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (*Bacillus stearothermophilus*). *Cell* **38**(3), 835–840. [https://doi.org/10.1016/0092-8674\(84\)90278-2](https://doi.org/10.1016/0092-8674(84)90278-2).
- Chedad A, Van Dael H, Vanhooren A and Hanssens I** (2005) Influence of Trp mutation on native, intermediate, and transition states of goat alpha-lactalbumin: An equilibrium and kinetic study. *Biochemistry* **44**(46), 15129–15138. <https://doi.org/10.1021/bi0512400>.
- Chiti F, Taddei N, White PM, Bucciantini M, Magherini F, Stefani M and Dobson CM** (1999) Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nature Structural Biology* **6**(11), 1005–1009. <https://doi.org/10.1038/14890>.
- Cho JH, O'Connell N, Raleigh DP and Palmer AG** (2010) Phi-value analysis for ultrafast folding proteins by NMR relaxation dispersion. *Journal of the American Chemical Society* **132**(2), 450. <https://doi.org/10.1021/ja909052h>.
- Cho JH and Raleigh DP** (2006) Denatured state effects and the origin of nonclassical phi values in protein folding. *Journal of the American Chemical Society* **128**(51), 16492–16493. <https://doi.org/10.1021/ja0669878>.
- Clarke J and Fersht AR** (1993) Engineered disulfide bonds as probes of the folding pathway of barnase: Increasing the stability of proteins against the rate of denaturation. *Biochemistry* **32**(16), 4322–4329. <https://doi.org/10.1021/bi00067a022>.
- Clementi C, Nymeyer H and Onuchic JN** (2000) Topological and energetic factors: What determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. *Journal of Molecular Biology* **298**(5), 937–953. <https://doi.org/10.1006/jmbi.2000.3693>.
- Cymes GD, Grosman C and Auerbach A** (2002) Structure of the transition state of gating in the acetylcholine receptor channel pore: A phi-value analysis. *Biochemistry* **41**(17), 5548–5555. <https://doi.org/10.1021/bi011864f>.
- Daggett V and Fersht AR** (2003) Is there a unifying mechanism for protein folding? *Trends in Biochemical Sciences* **28**(1), 18–25. [https://doi.org/10.1016/s0968-0004\(02\)00012-9](https://doi.org/10.1016/s0968-0004(02)00012-9).
- Daggett V and Levitt M** (1992) A model of the molten globule state from molecular dynamics simulations. *Proceedings of the National Academy of Sciences USA* **89**(11), 5142–5146. <https://doi.org/10.1073/pnas.89.11.5142>.
- Daggett V, Li AJ and Fersht AR** (1998) Combined molecular dynamics and Φ -value analysis of structure-reactivity relationships in the transition state and unfolding pathway of barnase: Structural basis of Hammond and anti-Hammond effects. *Journal of the American Chemical Society* **120**(49), 12740–12754. <https://doi.org/10.1021/ja981558y>.
- Daggett V, Li A, Itzhaki LS, Otzen DE and Fersht AR** (1996) Structure of the transition state for folding of a protein derived from experiment and simulation. *Journal of Molecular Biology* **257**(2), 430–440. <https://doi.org/10.1006/jmbi.1996.0173>.
- Dalby PA, Clarke J, Johnson CM and Fersht AR** (1998a) Folding intermediates of wild-type and mutants of barnase. II. Correlation of changes in equilibrium amide exchange kinetics with the population of the folding intermediate. *Journal of Molecular Biology* **276**(3), 647–656. <https://doi.org/10.1006/jmbi.1997.1547>.
- Dalby PA, Oliveberg M and Fersht AR** (1998b) Folding intermediates of wild-type and mutants of barnase. I. Use of phi-value analysis and m-values to probe the cooperative nature of the folding pre-equilibrium. *Journal of Molecular Biology* **276**(3), 625–646. <https://doi.org/10.1006/jmbi.1997.1546>.
- Dalby PA, Oliveberg M and Fersht AR** (1998c) Movement of the intermediate and rate determining transition state of barnase on the energy landscape with changing temperature. *Biochemistry* **37**(13), 4674–4679. <https://doi.org/10.1021/bi972798d>.
- Dave K, Jager M, Nguyen H, Kelly JW and Gruebele M** (2016) High-resolution mapping of the folding transition state of a WW domain. *Journal of Molecular Biology* **428**(8), 1617–1636. <https://doi.org/10.1016/j.jmb.2016.02.008>.
- Day R, Bennion BJ, Ham S and Daggett V** (2002) Increasing temperature accelerates protein unfolding without changing the pathway of unfolding. *Journal of Molecular Biology* **322**(1), 189–203. [https://doi.org/10.1016/s0022-2836\(02\)00672-1](https://doi.org/10.1016/s0022-2836(02)00672-1).
- Day R and Daggett V** (2005) Ensemble versus single-molecule protein unfolding. *Proceedings of the National Academy of Sciences USA* **102**(38), 13445–13450. <https://doi.org/10.1073/pnas.0501773102>.
- Day R and Daggett V** (2007) Direct observation of microscopic reversibility in single-molecule protein folding. *Journal of Molecular Biology* **366**(2), 677–686. <https://doi.org/10.1016/j.jmb.2006.11.043>.
- Debye P and Huckel E** (1923) The theory of electrolytes I. The lowering of the freezing point and related occurrences. *Physikalische Zeitschrift* **24**, 185–206.
- DeMarco ML, Alonso DO and Daggett V** (2004) Diffusing and colliding: The atomic level folding/unfolding pathway of a small helical protein. *Journal of Molecular Biology* **341**(4), 1109–1124. <https://doi.org/10.1016/j.jmb.2004.06.074>.
- Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD and Chan HS** (1995) Principles of protein folding – A perspective from simple exact models. *Protein Science* **4**(4), 561–602. <https://doi.org/10.1002/pro.5560040401>.
- Dill KA, Ozkan SB, Shell MS and Weikl TR** (2008) The protein folding problem. *Annual Review of Biophysics* **37**, 289–316. <https://doi.org/10.1146/annurev.biophys.37.092707.153558>.
- Dill KA and Shortle D** (1991) Denatured states of proteins. *Annual Review of Biochemistry* **60**, 795–825. <https://doi.org/10.1146/annurev.bi.60.070191.004051>.
- Dodson CA and Arbely E** (2015) Protein folding of the SAP domain, a naturally occurring two-helix bundle. *FEBS Letters* **589**(15), 1740–1747. <https://doi.org/10.1016/j.febslet.2015.06.002>.

- Dogan J, Gianni S and Jemth P** (2014) The binding mechanisms of intrinsically disordered proteins. *Physical Chemistry Chemical Physics* **16**(14), 6323–6331. <https://doi.org/10.1039/c3cp54226b>.
- Dogan J, Mu X, Engstrom A and Jemth P** (2013) The transition state structure for coupled binding and folding of disordered protein domains. *Scientific Reports* **3**, 2076. <https://doi.org/10.1038/srep02076>.
- Eaton WA, Thompson PA, Chan CK, Hage SJ and Hofrichter J** (1996) Fast events in protein folding. *Structure* **4**(10), 1133–1139. [https://doi.org/10.1016/s0969-2126\(96\)00121-9](https://doi.org/10.1016/s0969-2126(96)00121-9).
- Edelstein SJ and Changeux JP** (2010) Relationships between structural dynamics and functional kinetics in Oligomeric membrane receptors. *Biophysical Journal* **98**(10), 2045–2052. <https://doi.org/10.1016/j.bpj.2010.01.050>.
- Englander SW** (2023) HX and me: Understanding Allostery, folding, and protein machines. *Annual Review of Biophysics* **52**, 1–18. <https://doi.org/10.1146/annurev-biophys-062122-093517>.
- Englander SW, Mayne L, Bai Y and Sosnick TR** (1997) Hydrogen exchange: The modern legacy of Linderstrom-Lang. *Protein Science* **6**(5), 1101–1109. <https://doi.org/10.1002/pro.5560060517>.
- Epand RM and Scheraga HA** (1968) The influence of long-range interactions on the structure of myoglobin. *Biochemistry* **7**(8), 2864–2872. <https://doi.org/10.1021/bi00848a024>.
- Evans MG and Polanyi M** (1935) Some applications of the transition state method to the calculation of reaction velocities, especially in solution. *Transactions of the Faraday Society* **31**(1), 0875–0893. <https://doi.org/10.1039/tf9353100875>.
- Eyring H** (1935) The activated complex and the absolute rate of chemical reactions. *Chemical Reviews* **17**(1), 65–77. <https://doi.org/10.1021/cr60056a006>.
- Ferguson N, Day R, Johnson CM, Allen MD, Daggett V and Fersht AR** (2005) Simulation and experiment at high temperatures: Ultrafast folding of a thermophilic protein by nucleation-condensation. *Journal of Molecular Biology* **347**(4), 855–870. <https://doi.org/10.1016/j.jmb.2004.12.061>.
- Ferguson N, Sharpe TD, Johnson CM and Fersht AR** (2006) The transition state for folding of a peripheral subunit-binding domain contains robust and ionic-strength dependent characteristics. *Journal of Molecular Biology* **356**(5), 1237–1247. <https://doi.org/10.1016/j.jmb.2005.12.016>.
- Fersht AR** (1974) Catalysis, binding and enzyme-substrate complementarity. *Proceedings of the Royal Society of London. Series B Biological Sciences* **187**(1089), 397–407. <https://doi.org/10.1098/rspb.1974.0084>.
- Fersht A** (1977) *Enzyme Structure and Mechanism*. Reading Eng. San Francisco: W. H. Freeman.
- Fersht AR** (1979) Fidelity of replication of phage phi X174 DNA by DNA polymerase III holoenzyme: Spontaneous mutation by misincorporation. *Proceedings of the National Academy of Sciences USA* **76**(10), 4946–4950. <https://doi.org/10.1073/pnas.76.10.4946>.
- Fersht AR** (1981) Enzymic editing mechanisms and the genetic code. *Proceedings of the Royal Society of London. Series B Biological Sciences* **212**(1189), 351–379. <https://doi.org/10.1098/rspb.1981.0044>.
- Fersht A** (1985) *Enzyme Structure and Mechanism*, 2nd edn. New York: W.H. Freeman.
- Fersht AR** (1987) The hydrogen-bond in molecular recognition. *Trends in Biochemical Sciences* **12**(8), 301–304. [https://doi.org/10.1016/0968-0004\(87\)90146-0](https://doi.org/10.1016/0968-0004(87)90146-0).
- Fersht AR** (1988) Relationships between apparent binding energies measured in site-directed mutagenesis experiments and energetics of binding and catalysis. *Biochemistry* **27**(5), 1577–1580. <https://doi.org/10.1021/bi00405a027>.
- Fersht AR** (1995) Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications. *Proceedings of the National Academy of Sciences USA* **92**(24), 10869–10873. <https://doi.org/10.1073/pnas.92.24.10869>.
- Fersht A** (1999) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. New York: W.H. Freeman.
- Fersht AR** (2000) Transition-state structure as a unifying basis in protein-folding mechanisms: Contact order, chain topology, stability, and the extended nucleus mechanism. *Proceedings of the National Academy of Sciences USA* **97**(4), 1525–1529. <https://doi.org/10.1073/pnas.97.4.1525>.
- Fersht AR** (2004a) Phi value versus psi analysis. *Proceedings of the National Academy of Sciences USA* **101**(50), 17327–17328. <https://doi.org/10.1073/pnas.0407863101>.
- Fersht AR** (2004b) Relationship of Leffler (Bronsted) alpha values and protein folding phi values to position of transition-state structures on reaction coordinates. *Proceedings of the National Academy of Sciences USA* **101**(40), 14338–14342. <https://doi.org/10.1073/pnas.0406091101>.
- Fersht AR** (2008) From the first protein structures to our current knowledge of protein folding: Delights and scepticisms. *Nature Reviews. Molecular Cell Biology* **9**(8), 650–654. <https://doi.org/10.1038/nrm2446>.
- Fersht A** (2017) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. Series in Structural Biology. New Jersey: World Scientific.
- Fersht A** (2018) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. Amazon Kindle. New York: Kaissa Publications.
- Fersht AR** (2021) AlphaFold - A personal perspective on the impact of machine learning. *Journal of Molecular Biology* **433**(20), 167088. <https://doi.org/10.1016/j.jmb.2021.167088>.
- Fersht AR and Daggett V** (2002) Protein folding and unfolding at atomic resolution. *Cell* **108**(4), 573–582. [https://doi.org/10.1016/s0092-8674\(02\)00620-7](https://doi.org/10.1016/s0092-8674(02)00620-7).
- Fersht AR, Itzhaki LS, elMasry NF, Matthews JM and Otzen DE** (1994) Single versus parallel pathways of protein folding and fractional formation of structure in the transition state. *Proceedings of the National Academy of Sciences USA* **91**(22), 10426–10429. <https://doi.org/10.1073/pnas.91.22.10426>.
- Fersht AR and Kirby AJ** (1967) Structure and mechanism in intramolecular catalysis. The hydrolysis of substituted aspirins. *Journal of the American Chemical Society* **89**(19), 4853–4856. <https://doi.org/10.1021/ja00995a006>.
- Fersht AR, Leatherbarrow RJ and Wells TNC** (1986) Quantitative-analysis of structure-activity-relationships in engineered proteins by linear free-energy relationships. *Nature* **322**(6076), 284–286. <https://doi.org/10.1038/322284a0>.
- Fersht AR, Leatherbarrow RJ and Wells TN** (1987) Structure-activity relationships in engineered proteins: Analysis of use of binding energy by linear free energy relationships. *Biochemistry* **26**(19), 6030–6038. <https://doi.org/10.1021/bi00393a013>.
- Fersht AR, Matouschek A and Serrano L** (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *Journal of Molecular Biology* **224**(3), 771–782. [https://doi.org/10.1016/0022-2836\(92\)90561-w](https://doi.org/10.1016/0022-2836(92)90561-w).
- Fersht AR and Petrovich M** (2013) Reply to Campos and Munoz: Why phosphate is a bad buffer for guanidinium chloride titrations. *Proceedings of the National Academy of Sciences USA* **110**(14), E1244–E1245. <https://doi.org/10.1073/pnas.1303286110>.
- Fersht AR and Sato S** (2004) Phi-value analysis and the nature of protein-folding transition states. *Proceedings of the National Academy of Sciences USA* **101**(21), 7976–7981. <https://doi.org/10.1073/pnas.0402684101>.
- Fersht AR, Shi JP, Knill-Jones J, Lowe DM, Wilkinson AJ, Blow DM, Brick P, Carter P, Waye MM and Winter G** (1985) Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* **314**(6008), 235–238. <https://doi.org/10.1038/314235a0>.
- Finkelstein AV, Bogatyreva NS, Ivankov DN and Garbuzynskiy SO** (2022) Protein folding problem: Enigma, paradox, solution. *Biophysical Reviews* **14**(6), 1255–1272. <https://doi.org/10.1007/s12551-022-01000-1>.
- Font J, Benito A, Lange R, Ribo M and Vilanova M** (2006) The contribution of the residues from the main hydrophobic core of ribonuclease a to its pressure-folding transition state. *Protein Science* **15**(5), 1000–1009. <https://doi.org/10.1110/ps.052050306>.
- Fowler SB, Best RB, Toca Herrera JL, Rutherford TJ, Steward A, Paci E, Karplus M and Clarke J** (2002) Mechanical unfolding of a titin Ig domain: Structure of unfolding intermediate revealed by combining AFM, molecular dynamics simulations. *NMR and Protein Engineering. Journal of Molecular Biology* **322**(4), 841–849. [https://doi.org/10.1016/s0022-2836\(02\)00805-7](https://doi.org/10.1016/s0022-2836(02)00805-7).
- Fowler SB and Clarke J** (2001) Mapping the folding pathway of an immunoglobulin domain: Structural detail from phi value analysis and movement of the transition state. *Structure* **9**(5), 355–366. [https://doi.org/10.1016/s0969-2126\(01\)00596-2](https://doi.org/10.1016/s0969-2126(01)00596-2).
- Friel CT, Capaldi AP and Radford SE** (2003) Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: Similarities and

- differences in the folding of homologous proteins. *Journal of Molecular Biology* **326**(1), 293–305. [https://doi.org/10.1016/s0022-2836\(02\)01249-4](https://doi.org/10.1016/s0022-2836(02)01249-4).
- Galano-Frutos JJ and Sancho J** (2019) Accurate calculation of Barnase and SNase folding energetics using short molecular dynamics simulations and an atomistic model of the unfolded ensemble: Evaluation of force fields and water models. *Journal of Chemical Information and Modeling* **59**(10), 4350–4360. <https://doi.org/10.1021/acs.jcim.9b00430>.
- Galano-Frutos JJ, Torreblanca R, Garcia-Cebollada H and Sancho J** (2022) A look at the face of the molten globule: Structural model of the helicobacter pylori apoflavodoxin ensemble at acidic pH. *Protein Science* **31**(11), e4445. <https://doi.org/10.1002/pro.4445>.
- Garcia-Mira MM, Boehringer D and Schmid FX** (2004) The folding transition state of the cold shock protein is strongly polarized. *Journal of Molecular Biology* **339**(3), 555–569. <https://doi.org/10.1016/j.jmb.2004.04.011>.
- Geierhaas CD, Paci E, Vendruscolo M and Clarke J** (2004) Comparison of the transition states for folding of two Ig-like proteins from different superfamilies. *Journal of Molecular Biology* **343**(4), 1111–1123. <https://doi.org/10.1016/j.jmb.2004.08.100>.
- Geierhaas CD, Salvatella X, Clarke J and Vendruscolo M** (2008) Characterisation of transition state structures for protein folding using ‘high’, ‘medium’ and ‘low’ phi-values. *Protein Engineering Design & Selection* **21**(3), 215–222. <https://doi.org/10.1093/protein/gzm092>.
- Gelman H and Gruebele M** (2014) Fast protein folding kinetics. *Quarterly Reviews of Biophysics* **47**(2), 95–142. <https://doi.org/10.1017/s003358351400002x>.
- Gianni S, Geierhaas CD, Calosci N, Jemth P, Vuister GW, Travaglini-Allocatelli C, Vendruscolo M and Brunori M** (2007) A PDZ domain recapitulates a unifying mechanism for protein folding. *Proceedings of the National Academy of Sciences USA* **104**(1), 128–133. <https://doi.org/10.1073/pnas.0602770104>.
- Gianni S, Guydosh NR, Khan F, Caldas TD, Mayor U, White GW, DeMarco ML, Daggett V and Fersht AR** (2003) Unifying features in protein-folding mechanisms. *Proceedings of the National Academy of Sciences USA* **100**(23), 13286–13291. <https://doi.org/10.1073/pnas.1835776100>.
- Gianni S and Jemth P** (2014) Conserved nucleation sites reinforce the significance of phi value analysis in protein-folding studies. *IUBMB Life* **66**(7), 449–452. <https://doi.org/10.1002/iub.1287>.
- Goldenberg DP, Frieden RW, Haack JA and Morrison TB** (1989) Mutational analysis of a protein-folding pathway. *Nature* **338**(6211), 127–132. <https://doi.org/10.1038/338127a0>.
- Grantcharova VP, Riddle DS, Santiago JV and Baker D** (1998) Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nature Structural Biology* **5**(8), 714–720. <https://doi.org/10.1038/1412>.
- Guerois R and Serrano L** (2000) The SH3-fold family: Experimental evidence and prediction of variations in the folding pathways. *Journal of Molecular Biology* **304**(5), 967–982. <https://doi.org/10.1006/jmbi.2000.4234>.
- Haldane JBS** (1930) *Enzymes*. London: Longman Green and Co.
- Hammett LP** (1937) The effect of structure upon the reactions of organic compounds benzene derivatives. *Journal of the American Chemical Society* **59**, 96–103. <https://doi.org/10.1021/ja01280a022>.
- Hammett LP** (1940) *Physical Organic Chemistry*. New York: McGraw-Hill Book Company.
- Hammond GS** (1955) A correlation of reaction rates. *Journal of the American Chemical Society* **77**(2), 334–338. <https://doi.org/10.1021/ja01607a027>.
- Homouz D, Stagg L, Wittung-Stafshede P and Cheung MS** (2009) Macromolecular crowding modulates folding mechanism of alpha/beta protein apoflavodoxin. *Biophysical Journal* **96**(2), 671–680. <https://doi.org/10.1016/j.bpj.2008.10.014>.
- Horowitz A and Fersht AR** (1990) Strategy for analysing the co-operativity of intramolecular interactions in peptides and proteins. *Journal of Molecular Biology* **214**(3), 613–617. [https://doi.org/10.1016/0022-2836\(90\)90275-Q](https://doi.org/10.1016/0022-2836(90)90275-Q).
- Horowitz A and Fersht AR** (1992) Co-operative interactions during protein folding. *Journal of Molecular Biology* **224**(3), 733–740. [https://doi.org/10.1016/0022-2836\(92\)90557-z](https://doi.org/10.1016/0022-2836(92)90557-z).
- Horowitz A, Serrano L and Fersht AR** (1991) COSMIC analysis of the major alpha-helix of barnase during folding. *Journal of Molecular Biology* **219**(1), 5–9. [https://doi.org/10.1016/0022-2836\(91\)90852-w](https://doi.org/10.1016/0022-2836(91)90852-w).
- Huang F, Settanni G and Fersht AR** (2008) Fluorescence resonance energy transfer analysis of the folding pathway of engrailed homeodomain. *Protein Engineering, Design, and Selection* **21**(3), 131–146. <https://doi.org/10.1093/protein/gzm069>.
- Hutchison CA 3rd, Phillips S, Edgell MH, Gillam S, Jahnke P and Smith M** (1978) Mutagenesis at a specific position in a DNA sequence. *Journal of Biological Chemistry* **253**(18), 6551–6560.
- Itzhaki LS, Otzen DE and Fersht AR** (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. *Journal of Molecular Biology* **254**(2), 260–288. <https://doi.org/10.1006/jmbi.1995.0616>.
- Ivarsson Y, Travaglini-Allocatelli C, Brunori M and Gianni S** (2009) Engineered symmetric connectivity of secondary structure elements highlights malleability of protein folding pathways. *Journal of the American Chemical Society* **131**(33), 11727–11733. <https://doi.org/10.1021/ja900438b>.
- Jackson SE** (1998) How do small single-domain proteins fold? *Folding & Design* **3**(4), R81–R91. [https://doi.org/10.1016/S1359-0278\(98\)00033-9](https://doi.org/10.1016/S1359-0278(98)00033-9).
- Jackson SE and Fersht AR** (1991a) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **30**(43), 10428–10435. <https://doi.org/10.1021/bi00107a010>.
- Jackson SE and Fersht AR** (1991b) Folding of chymotrypsin inhibitor 2. 2. Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding. *Biochemistry* **30**(43), 10436–10443. <https://doi.org/10.1021/bi00107a011>.
- Jackson SE, Suma A and Micheletti C** (2017) How to fold intricately: Using theory and experiments to unravel the properties of knotted proteins. *Current Opinion in Structural Biology* **42**, 6–14. <https://doi.org/10.1016/j.sbi.2016.10.002>.
- Jager M, Nguyen H, Crane JC, Kelly JW and Gruebele M** (2001) The folding mechanism of a beta-sheet: The WW domain. *Journal of Molecular Biology* **311**(2), 373–393. <https://doi.org/10.1006/jmbi.2001.4873>.
- Jemth P, Day R, Gianni S, Khan F, Allen M, Daggett V and Fersht AR** (2005) The structure of the major transition state for folding of an FF domain from experiment and simulation. *Journal of Molecular Biology* **350**(2), 363–378. <https://doi.org/10.1016/j.jmb.2005.04.067>.
- Jencks WP** (1975) Binding energy, specificity, and enzymic catalysis: The circe effect. *Advances in Enzymology and Related Areas of Molecular Biology* **43**, 219–410. <https://doi.org/10.1002/9780470122884.ch4>.
- Jencks WP** (1985) A primer for the BEMA HAPOTHELE – An empirical approach to the characterization of changing transition-state structures. *Chemical Reviews* **85**(6), 511–527. <https://doi.org/10.1021/cr00070a001>.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P and Hassabis D** (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
- Karlsson E, Andersson E, Dogan J, Gianni S, Jemth P and Camilloni C** (2019) A structurally heterogeneous transition state underlies coupled binding and folding of disordered proteins. *Journal of Biological Chemistry* **294**(4), 1230–1239. <https://doi.org/10.1074/jbc.RA118.005854>.
- Karlsson OA, Chi CN, Engstrom A and Jemth P** (2012) The transition state of coupled folding and binding for a flexible beta-finger. *Journal of Molecular Biology* **417**(3), 253–261. <https://doi.org/10.1016/j.jmb.2012.01.042>.
- Karlsson E and Jemth P** (2021) Kinetic methods of deducing binding mechanisms involving intrinsically disordered proteins. *Methods in Molecular Biology* **2263**, 105–133. https://doi.org/10.1007/978-1-0716-1197-5_4.
- Karlsson E, Pisoni C, Erkelens AM, Tehranizadeh ZA, Sorgenfrei FA, Andersson E, Ye W, Camilloni C and Jemth P** (2020) Mapping the transition state for a binding reaction between ancient intrinsically disordered proteins. *Journal of Biological Chemistry* **295**(51), 17698–17712. <https://doi.org/10.1074/jbc.RA120.015645>.
- Karplus M** (2011) Behind the folding funnel diagram. *Nature Chemical Biology* **7**(7), 401–404. <https://doi.org/10.1038/nchembio.565>.

- Karplus M and Weaver DL** (1976) Protein-folding dynamics. *Nature* **260** (5550), 404–406. <https://doi.org/10.1038/260404a0>.
- Kazmirski SL, Wong KB, Freund SM, Tan YJ, Fersht AR and Daggett V** (2001) Protein folding from a highly disordered denatured state: The folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proceedings of the National Academy of Sciences USA* **98**(8), 4349–4354. <https://doi.org/10.1073/pnas.071054398>.
- Kellis JT, Nyberg K and Fersht AR** (1989) Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry* **28**(11), 4914–4922. <https://doi.org/10.1021/bi00437a058>.
- Kellis JT, Nyberg K, Sali D and Fersht AR** (1988) Contribution of hydrophobic interactions to protein stability. *Nature* **333**(6175), 784–786. <https://doi.org/10.1038/333784a0>.
- Kelly SE, Meisl G, Rowling PJ, McLaughlin SH, Knowles T and Itzhaki LS** (2014) Diffuse transition state structure for the unfolding of a leucine-rich repeat protein. *Physical Chemistry Chemical Physics* **16**(14), 6448–6459. <https://doi.org/10.1039/c3cp54818j>.
- Khan F, Chuang JI, Gianni S and Fersht AR** (2003) The kinetic pathway of folding of barnase. *Journal of Molecular Biology* **333**(1), 169–186. <https://doi.org/10.1016/j.jmb.2003.08.024>.
- Kim PS and Baldwin RL** (1990) Intermediates in the folding reactions of small proteins. *Annual Review of Biochemistry* **59**, 631–660. <https://doi.org/10.1146/annurev.bi.59.070190.003215>.
- Kim DE, Fisher C and Baker D** (2000) A breakdown of symmetry in the folding transition state of protein L. *Journal of Molecular Biology* **298**(5), 971–984. <https://doi.org/10.1006/jmbi.2000.3701>.
- Kim J and Shin JS** (2010) Probing the transition state for nucleic acid hybridization using phi-value analysis. *Biochemistry* **49**(16), 3420–3426. <https://doi.org/10.1021/bi902047x>.
- Konuma T, Kimura T, Matsumoto S, Goto Y, Fujisawa T, Fersht AR and Takahashi S** (2011) Time-resolved small-angle X-ray scattering study of the folding dynamics of barnase. *Journal of Molecular Biology* **405**(5), 1284–1294. <https://doi.org/10.1016/j.jmb.2010.11.052>.
- Kragelund BB, Poulsen K, Andersen KV, Baldursson T, Kroll JB, Neergård TB, Jepsen J, Roepstorff P, Kristiansen K, Poulsen FM and Knudsen J** (1999) Conserved residues and their role in the structure, function, and stability of acyl-coenzyme A binding protein. *Biochemistry* **38**(8), 2386–2394. <https://doi.org/10.1021/bi982427c>.
- Krantz BA and Sosnick TR** (2001) Engineered metal binding sites map the heterogeneous folding landscape of a coiled coil. *Nature Structural Biology* **8** (12), 1042–1047. <https://doi.org/10.1038/nsb723>.
- Kubelka J, Hofrichter J and Eaton WA** (2004) The protein folding ‘speed limit’. *Current Opinion in Structural Biology* **14**(1), 76–88. <https://doi.org/10.1016/j.sbi.2004.01.013>.
- Kukic P, Pustovalova Y, Camilloni C, Gianni S, Korzhnev DM and Vendruscolo M** (2017) Structural characterization of the early events in the nucleation-condensation mechanism in a protein folding process. *Journal of the American Chemical Society* **139**(20), 6899–6910. <https://doi.org/10.1021/jacs.7b01540>.
- Ladurner AG, Itzhaki LS, Daggett V and Fersht AR** (1998) Synergy between simulation and experiment in describing the energy landscape of protein folding. *Proceedings of the National Academy of Sciences USA* **95**(15), 8473–8478. <https://doi.org/10.1073/pnas.95.15.8473>.
- Lappalainen I, Hurley MG and Clarke J** (2008) Plasticity within the obligatory folding nucleus of an immunoglobulin-like domain. *Journal of Molecular Biology* **375**(2), 547–559. <https://doi.org/10.1016/j.jmb.2007.09.088>.
- Lawrence C, Vallee-Belisle A, Pfeil SH, de Mornay D, Lipman EA and Plaxco KW** (2014) A comparison of the folding kinetics of a small, artificially selected DNA aptamer with those of equivalently simple naturally occurring proteins. *Protein Science* **23**(1), 56–66. <https://doi.org/10.1002/pro.2390>.
- Lazaridis T and Karplus M** (1997) “New view” of protein folding reconciled with the old through multiple unfolding simulations. *Science* **278**(5345), 1928–1931. <https://doi.org/10.1126/science.278.5345.1928>.
- Leatherbarrow RJ and Fersht AR** (1987) Investigation of transition-state stabilization by residues histidine-45 and threonine-40 in the tyrosyl-tRNA synthetase. *Biochemistry* **26**(26), 8524–8528. <https://doi.org/10.1021/bi00400a005>.
- Leatherbarrow RJ, Fersht AR and Winter G** (1985) Transition-state stabilization in the mechanism of tyrosyl-tRNA synthetase revealed by protein engineering. *Proceedings of the National Academy of Sciences USA* **82**(23), 7840–7844. <https://doi.org/10.1073/pnas.82.23.7840>.
- Leffler JE** (1953) Parameters for the description of transition states. *Science* **117** (3039), 340–341. <https://doi.org/10.1126/science.117.3039.340>.
- Levinthal C** (1968) Are there pathways for protein folding. *Journal de Chimie Physique et de Physico-Chimie Biologique* **65**(1), 44. <https://doi.org/10.1051/jcp/1968650044>.
- Li A and Daggett V** (1994) Characterization of the transition state of protein unfolding by use of molecular dynamics: Chymotrypsin inhibitor 2. *Proceedings of the National Academy of Sciences USA* **91**(22), 10430–10434. <https://doi.org/10.1073/pnas.91.22.10430>.
- Li A and Daggett V** (1996) Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 by molecular dynamics simulations. *Journal of Molecular Biology* **257**(2), 412–429. <https://doi.org/10.1006/jmbi.1996.0172>.
- Li A and Daggett V** (1998) Molecular dynamics simulation of the unfolding of barnase: Characterization of the major intermediate. *Journal of Molecular Biology* **275**(4), 677–694. <https://doi.org/10.1006/jmbi.1997.1484>.
- Lindberg MO, Haglund E, Hubner IA, Shakhnovich EI and Oliveberg M** (2006) Identification of the minimal protein-folding nucleus through loop-entropy perturbations. *Proceedings of the National Academy of Sciences USA* **103**(11), 4083–4088. <https://doi.org/10.1073/pnas.0508863103>.
- Lopez-Hernandez E and Serrano L** (1996) Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, CI-2. *Folding & Design* **1**(1), 43–55. [https://doi.org/10.1016/s1359-0278\(96\)00011-9](https://doi.org/10.1016/s1359-0278(96)00011-9).
- Lopez-Llano J, Campos LA, Bueno M and Sancho J** (2006) Equilibrium phi-analysis of a molten globule: The 1-149 apoflavodoxin fragment. *Journal of Molecular Biology* **356**(2), 354–366. <https://doi.org/10.1016/j.jmb.2005.10.086>.
- Main ER, Fulton KF, Daggett V and Jackson SE** (2001) A comparison of experimental and computational methods for mapping the interactions present in the transition state for folding of FKBP12. *Journal of Biological Physics* **27**(2–3), 99–117. <https://doi.org/10.1023/A:1013137924581>.
- Malagrino F, Visconti L, Pagano L, Toto A, Troilo F and Gianni S** (2020) Understanding the binding induced folding of intrinsically disordered proteins by protein engineering: Caveats and pitfalls. *Current Opinion in Structural Biology* **21**(10), 3484. <https://doi.org/10.3390/ijms21103484>.
- Mallam AL, Morris ER and Jackson SE** (2008) Exploring knotting mechanisms in protein folding. *Proceedings of the National Academy of Sciences USA* **105** (48), 18740–18745. <https://doi.org/10.1073/pnas.0806697105>.
- Manavalan B, Kuwajima K and Lee J** (2019) PFDB: A standardized protein folding database with temperature correction. *Scientific Reports* **9**(1), 1588. <https://doi.org/10.1038/s41598-018-36992-y>.
- Marcus RA** (1968) Theoretical relations among rate constants barriers and Brønsted slopes of chemical reactions. *Journal of Physical Chemistry* **72**(3), 891. <https://doi.org/10.1021/j100849a019>.
- Martinez JC and Serrano L** (1999) The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nature Structural Biology* **6**(11), 1010–1016. <https://doi.org/10.1038/14896>.
- Matouschek A and Fersht AR** (1993) Application of physical organic chemistry to engineered mutants of proteins: Hammond postulate behavior in the transition state of protein folding. *Proceedings of the National Academy of Sciences USA* **90**(16), 7814–7818. <https://doi.org/10.1073/pnas.90.16.7814>.
- Matouschek A, Kellis JT, Serrano L, Bycroft M and Fersht AR** (1990) Transient folding intermediates characterized by protein engineering. *Nature* **346** (6283), 440–445. <https://doi.org/10.1038/346440a0>.
- Matouschek A, Kellis JT, Serrano L and Fersht AR** (1989) Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340**(6229), 122–126. <https://doi.org/10.1038/340122a0>.
- Matouschek A, Otzen DE, Itzhaki LS, Jackson SE and Fersht AR** (1995) Movement of the position of the transition state in protein folding. *Biochemistry* **34**(41), 13656–13662. <https://doi.org/10.1021/bi00041a047>.
- Matouschek A, Serrano L and Fersht AR** (1992) The folding of an enzyme. IV. Structure of an intermediate in the refolding of barnase analysed by a

- protein engineering procedure. *Journal of Molecular Biology* **224**(3), 819–835. [https://doi.org/10.1016/0022-2836\(92\)90564-z](https://doi.org/10.1016/0022-2836(92)90564-z).
- Matthews CR** (1987) Effect of point mutations on the folding of globular proteins. *Methods in Enzymology* **154**, 498–511. [https://doi.org/10.1016/0076-6879\(87\)54092-7](https://doi.org/10.1016/0076-6879(87)54092-7).
- Matthews JM and Fersht AR** (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**(20), 6805–6814. <https://doi.org/10.1021/bi00020a027>.
- Mayor U, Grossmann JG, Foster NW, Freund SM and Fersht AR** (2003a) The denatured state of engrailed homeodomain under denaturing and native conditions. *Journal of Molecular Biology* **333**(5), 977–991. <https://doi.org/10.1016/j.jmb.2003.08.062>.
- Mayor U, Guydosh NR, Johnson CM, Grossmann JG, Sato S, Jas GS, Freund SM, Alonso DO, Daggett V and Fersht AR** (2003b) The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* **421**(6925), 863–867. <https://doi.org/10.1038/nature01428>.
- Mayor U, Johnson CM, Daggett V and Fersht AR** (2000) Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proceedings of the National Academy of Sciences USA* **97**(25), 13518–13522. <https://doi.org/10.1073/pnas.250473497>.
- McCully ME, Beck DA and Daggett V** (2008) Microscopic reversibility of protein folding in molecular dynamics simulations of the engrailed homeodomain. *Biochemistry* **47**(27), 7079–7089. <https://doi.org/10.1021/bi800118b>.
- Meiering EM, Serrano L and Fersht AR** (1992) Effect of active site residues in barnase on activity and stability. *Journal of Molecular Biology* **225**(3), 585–589. [https://doi.org/10.1016/0022-2836\(92\)90387-y](https://doi.org/10.1016/0022-2836(92)90387-y).
- Mitra A, Bailey TD and Auerbach AL** (2004) Structural dynamics of the M4 transmembrane segment during acetylcholine receptor gating. *Structure* **12**(10), 1909–1918. <https://doi.org/10.1016/j.str.2004.08.004>.
- Muralidhara BK, Chen M, Ma J and Wittung-Stafshede P** (2005) Effect of inorganic phosphate on FMN binding and loop flexibility in Desulfovibrio desulfuricans apo-flavodoxin. *Journal of Molecular Biology* **349**(1), 87–97. <https://doi.org/10.1016/j.jmb.2005.03.054>.
- Naganathan AN and Orozco M** (2011) The protein folding transition-state ensemble from a G(o)over-bar-like model. *Physical Chemistry Chemical Physics* **13**(33), 15166–15174. <https://doi.org/10.1039/c1cp20964g>.
- Nasedkin A, Marcellini M, Religa TL, Freund SM, Menzel A, Fersht AR, Jemth P, van der Spoel D and Davidsson J** (2015) Deconvoluting protein (un)folding structural ensembles using X-ray scattering, nuclear magnetic resonance spectroscopy and molecular dynamics simulation. *PLoS One* **10**(5), e0125662. <https://doi.org/10.1371/journal.pone.0125662>.
- Neuweiler H, Banachewicz W and Fersht AR** (2010) Kinetics of chain motions within a protein-folding intermediate. *Proceedings of the National Academy of Sciences USA* **107**(51), 22106–22110. <https://doi.org/10.1073/pnas.1011666107>.
- Neuweiler H, Sharpe TD, Rutherford TJ, Johnson CM, Allen MD, Ferguson N and Fersht AR** (2009) The folding mechanism of BBL: Plasticity of transition-state structure observed within an ultrafast folding protein family. *Journal of Molecular Biology* **390**(5), 1060–1073. <https://doi.org/10.1016/j.jmb.2009.05.011>.
- Nickson AA, Stoll KE and Clarke J** (2008) Folding of a LysM domain: Entropy-enthalpy compensation in the transition state of an ideal two-state folder. *Journal of Molecular Biology* **380**(3), 557–569. <https://doi.org/10.1016/j.jmb.2008.05.020>.
- Nolting B and Agard DA** (2008) How general is the nucleation-condensation mechanism? *Proteins* **73**(3), 754–764. <https://doi.org/10.1002/prot.22099>.
- Nolting B, Golbik R, Neira JL, Soler-Gonzalez AS, Schreiber G and Fersht AR** (1997) The folding pathway of a protein at high resolution from microseconds to seconds. *Proceedings of the National Academy of Sciences USA* **94**(3), 826–830. <https://doi.org/10.1073/pnas.94.3.826>.
- Oliveberg M and Wolynes PG** (2005) The experimental survey of protein-folding energy landscapes. *Quarterly Reviews of Biophysics* **38**(3), 245–288. <https://doi.org/10.1017/S0033583506004185>.
- Onuchic JN, Wolynes PG, Luthey-Schulten Z and Socci ND** (1995) Toward an outline of the topography of a realistic protein-folding funnel. *Proceedings of the National Academy of Sciences USA* **92**(8), 3626–3630. <https://doi.org/10.1073/pnas.92.8.3626>.
- Ooka K and Arai M** (2023) Accurate prediction of protein folding mechanisms by simple structure-based statistical mechanical models. *Nature Communications* **14**(1), 6338. <https://doi.org/10.1038/s41467-023-41664-1>.
- Otzen DE** (2011) Mapping the folding pathway of the transmembrane protein DsbB by protein engineering. *Protein Engineering Design & Selection* **24**(1–2), 139–149. <https://doi.org/10.1093/protein/gzq079>.
- Otzen DE and Oliveberg M** (2002) Conformational plasticity in folding of the split beta-alpha-beta protein S6: Evidence for burst-phase disruption of the native state. *Journal of Molecular Biology* **317**(4), 613–627. <https://doi.org/10.1006/jmbi.2002.5423>.
- Paci E, Friel CT, Lindorff-Larsen K, Radford SE, Karplus M and Vendruscolo M** (2004) Comparison of the transition state ensembles for folding of Im7 and Im9 determined using all-atom molecular dynamics simulations with phi value restraints. *Proteins-Structure Function and Bioinformatics* **54**(3), 513–525. <https://doi.org/10.1002/prot.10595>.
- Padmanabhan S, Marqusee S, Ridgeway T, Laue TM and Baldwin RL** (1990) Relative helix-forming tendencies of nonpolar amino acids. *Nature* **344**(6263), 268–270. <https://doi.org/10.1038/344268a0>.
- Pagano L, Toto A, Malagrino F, Visconti L, Jemth P and Gianni S** (2021) Double mutant cycles as a tool to address folding, binding, and allostery. *International Journal of Molecular Sciences* **22**(2), 828. <https://doi.org/10.3390/ijms22020828>.
- Pande VS, Grosberg A, Tanaka T and Rokhsar DS** (1998) Pathways for protein folding: Is a new view needed? *Current Opinion in Structural Biology* **8**(1), 68–79. [https://doi.org/10.1016/S0959-440X\(98\)80012-2](https://doi.org/10.1016/S0959-440X(98)80012-2).
- Paslowski W, Lillelund OK, Kristensen JV, Schafer NP, Baker RP, Urban S and Otzen DE** (2015) Cooperative folding of a polytopic alpha-helical membrane protein involves a compact N-terminal nucleus and nonnative loops. *Proceedings of the National Academy of Sciences of the United States of America* **112**(26), 7978–7983. <https://doi.org/10.1073/pnas.1424751112>.
- Pauling L** (1948) Chemical achievement and hope for the future. *American Scientist* **36**(1), 51–58.
- Pelzer H and Wigner E** (1932) The speed constants of the exchange reactions. *Zeitschrift Fur Physikalische Chemie-Abteilung B-Chemie Der Elementarprozesse Aufbau Der Materie* **15**(6), 445–471.
- Pereyaslavets LB and Galzitskaya OV** (2015) Theoretical search for RNA folding nuclei. *Entropy* **17**(11), 7827–7847. <https://doi.org/10.3390/e17117827>.
- Petrovich M, Jonsson AL, Ferguson N, Daggett V and Fersht AR** (2006) Phi-analysis at the experimental limits: Mechanism of beta-hairpin formation. *Journal of Molecular Biology* **360**(4), 865–881. <https://doi.org/10.1016/j.jmb.2006.05.050>.
- Prigozhin MB and Gruebele M** (2013) Microsecond folding experiments and simulations: A match is made. *Physical Chemistry Chemical Physics* **15**(10), 3372–3388. <https://doi.org/10.1039/c3cp43992e>.
- Ptitsyn OB** (1973) Stages in the mechanism of self-organization of protein molecules. *Doklady Akademii Nauk SSSR* **210**(5), 1213–1215.
- Ptitsyn OB** (1987) Protein folding – hypotheses and experiments. *Journal of Protein Chemistry* **6**(4), 273–293.
- Ptitsyn OB** (1991) How does protein synthesis give rise to the 3D-structure? *FEBS Letters* **285**(2), 176–181. [https://doi.org/10.1016/0014-5793\(91\)80799-9](https://doi.org/10.1016/0014-5793(91)80799-9).
- Religa TL, Johnson CM, Vu DM, Brewer SH, Dyer RB and Fersht AR** (2007) The helix-turn-helix motif as an ultrafast independently folding domain: The pathway of folding of engrailed homeodomain. *Proceedings of the National Academy of Sciences USA* **104**(22), 9272–9277. <https://doi.org/10.1073/pnas.0703434104>.
- Religa TL, Markson JS, Mayor U, Freund SM and Fersht AR** (2005) Solution structure of a protein denatured state and folding intermediate. *Nature* **437**(7061), 1053–1056. <https://doi.org/10.1038/nature04054>.
- Rogers JM, Oleinikovas V, Shammass SL, Wong CT, De Sancho D, Baker CM and Clarke J** (2014) Interplay between partner and ligand facilitates the folding and binding of an intrinsically disordered protein. *Proceedings of the*

- National Academy of Sciences USA* **111**(43), 15420–15425. <https://doi.org/10.1073/pnas.1409122111>.
- Sali D, Bycroft M and Fersht AR** (1988) Stabilization of protein structure by interaction of alpha-helix dipole with a charged side chain. *Nature* **335**(6192), 740–743. <https://doi.org/10.1038/335740a0>.
- Sali A, Shakhnovich E and Karplus M** (1994a) How does a protein fold? *Nature* **369**(6477), 248–251. <https://doi.org/10.1038/369248a0>.
- Sali A, Shakhnovich E and Karplus M** (1994b) Kinetics of protein folding. A lattice model study of the requirements for folding to the native state. *Journal of Molecular Biology* **235**(5), 1614–1636. <https://doi.org/10.1006/jmbi.1994.1110>.
- Salvarella X, Dobson CM, Fersht AR and Vendruscolo M** (2005) Determination of the folding transition states of barnase by using PhiIJ-value-restrained simulations validated by double mutant PhiIJ-values. *Proceedings of the National Academy of Sciences USA* **102**(35), 12389–12394. <https://doi.org/10.1073/pnas.0408226102>.
- Sanchez IE and Kiefhaber T** (2003) Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *Journal of Molecular Biology* **325**(2), 367–376. [https://doi.org/10.1016/s0022-2836\(02\)01230-5](https://doi.org/10.1016/s0022-2836(02)01230-5).
- Sato S, Cho JH, Peran I, Soydaner-Azeloglu RG and Raleigh DP** (2017) The N-terminal domain of ribosomal protein L9 folds via a diffuse and delocalized transition state. *Biophysical Journal* **112**(9), 1797–1806. <https://doi.org/10.1016/j.bpj.2017.01.034>.
- Sato S, Religa TL and Fersht AR** (2006) Phi-analysis of the folding of the B domain of protein a using multiple optical probes. *Journal of Molecular Biology* **360**(4), 850–864. <https://doi.org/10.1016/j.jmb.2006.05.051>.
- Schramm VL** (1998) Enzymatic transition states and transition state analog design. *Annual Review of Biochemistry* **67**, 693–720. <https://doi.org/10.1146/annurev.biochem.67.1.693>.
- Schreiber G and Fersht AR** (1995) Energetics of protein-protein interactions: Analysis of the barnase-barstar interface by single mutations and double mutant cycles. *Journal of Molecular Biology* **248**(2), 478–486. [https://doi.org/10.1016/s0022-2836\(95\)80064-6](https://doi.org/10.1016/s0022-2836(95)80064-6).
- Schreiber G and Fersht AR** (1996) Rapid, electrostatically assisted association of proteins. *Nature Structural Biology* **3**(5), 427–431. <https://doi.org/10.1038/nsb0596-427>.
- Schymkowitz JWH, Rousseau F, Irvine LR and Itzhaki LS** (2000) The folding pathway of the cell-cycle regulatory protein p13suc1: Clues for the mechanism of domain swapping. *Structure* **8**(1), 89–100. [https://doi.org/10.1016/s0969-2126\(00\)00084-8](https://doi.org/10.1016/s0969-2126(00)00084-8).
- Scott KA, Alonso DO, Sato S, Fersht AR and Daggett V** (2007) Conformational entropy of alanine versus glycine in protein denatured states. *Proceedings of the National Academy of Sciences USA* **104**(8), 2661–2666. <https://doi.org/10.1073/pnas.0611182104>.
- Serrano L, Day AG and Fersht AR** (1993) Step-wise mutation of barnase to binase. A procedure for engineering increased stability of proteins and an experimental analysis of the evolution of protein stability. *Journal of Molecular Biology* **233**(2), 305–312. <https://doi.org/10.1006/jmbi.1993.1508>.
- Serrano L, Matouschek A and Fersht AR** (1992a) The folding of an enzyme. III. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. *Journal of Molecular Biology* **224**(3), 805–818. [https://doi.org/10.1016/0022-2836\(92\)90563-y](https://doi.org/10.1016/0022-2836(92)90563-y).
- Serrano L, Neira JL, Sancho J and Fersht AR** (1992b) Effect of alanine versus glycine in alpha-helices on protein stability. *Nature* **356**(6368), 453–455. <https://doi.org/10.1038/356453a0>.
- Serrano L, Sancho J, Hirshberg M and Fersht AR** (1992c) Alpha-helix stability in proteins. I. Empirical correlations concerning substitution of side-chains at the N and C-caps and the replacement of alanine by glycine or serine at solvent-exposed surfaces. *Journal of Molecular Biology* **227**(2), 544–559. [https://doi.org/10.1016/0022-2836\(92\)90906-z](https://doi.org/10.1016/0022-2836(92)90906-z).
- Shammas SL, Crabtree MD, Dahal L, Wicky BI and Clarke J** (2016) Insights into coupled folding and binding mechanisms from kinetic studies. *Journal of Biological Chemistry* **291**(13), 6689–6695. <https://doi.org/10.1074/jbc.R115.692715>.
- Sharpe TD, Ferguson N, Johnson CM and Fersht AR** (2008) Conservation of transition state structure in fast folding peripheral subunit-binding domains. *Journal of Molecular Biology* **383**(1), 224–237. <https://doi.org/10.1016/j.jmb.2008.06.081>.
- Silverman SK and Cech TR** (2001) An early transition state for folding of the P4-P6 RNA domain. *RNA* **7**(2), 161–166. <https://doi.org/10.1017/s1355838201001716>.
- Spector S, Kuhlman B, Fairman R, Wong E, Boice JA and Raleigh DP** (1998) Cooperative folding of a protein mini domain: The peripheral subunit-binding domain of the pyruvate dehydrogenase multienzyme complex. *Journal of Molecular Biology* **276**(2), 479–489. <https://doi.org/10.1006/jmbi.1997.1522>.
- Spector S and Raleigh DP** (1999) Submillisecond folding of the peripheral subunit-binding domain. *Journal of Molecular Biology* **293**(4), 763–768. <https://doi.org/10.1006/jmbi.1999.3189>.
- Spector S, Rosconi M and Raleigh DP** (1999a) Conformational analysis of peptide fragments derived from the peripheral subunit-binding domain from the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*: Evidence for nonrandom structure in the unfolded state. *Biopolymers* **49**(1), 29–40. [https://doi.org/10.1002/\(SICI\)1097-0282\(199901\)49:1<29::AID-BIP4>3.0.CO;2-7](https://doi.org/10.1002/(SICI)1097-0282(199901)49:1<29::AID-BIP4>3.0.CO;2-7).
- Spector S, Young P and Raleigh DP** (1999b) Nativelike structure and stability in a truncation mutant of a protein minidomain: The peripheral subunit-binding domain. *Biochemistry* **38**(13), 4128–4136. <https://doi.org/10.1021/bi982915k>.
- Srivastava AK and Sauer RT** (2000) Evidence for partial secondary structure formation in the transition state for arc repressor refolding and dimerization. *Biochemistry* **39**(28), 8308–8314. <https://doi.org/10.1021/bi000423d>.
- Stagg L, Samiotakis A, Homouz D, Cheung MS and Wittung-Stafshede P** (2010) Residue-specific analysis of frustration in the folding landscape of repeat beta/alpha protein apoflavodoxin. *Journal of Molecular Biology* **396**(1), 75–89. <https://doi.org/10.1016/j.jmb.2009.11.008>.
- Stollar EJ, Mayor U, Lovell SC, Federici L, Freund SM, Fersht AR and Luisi BF** (2003) Crystal structures of engrailed homeodomain mutants: Implications for stability and dynamics. *Journal of Biological Chemistry* **278**(44), 43699–43708. <https://doi.org/10.1074/jbc.M308029200>.
- Takada S** (2019) Go model revisited. *Biophysics and Physicobiology* **16**, 248–255. https://doi.org/10.2142/biophysico.16.0_248.
- Taketomi H, Ueda Y and Go N** (1975) Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *International Journal of Peptide and Protein Research* **7**(6), 445–459.
- Tanford C** (1968) Protein denaturation. *Advances in Protein Chemistry* **23**, 121–282. [https://doi.org/10.1016/s0065-3233\(08\)60401-5](https://doi.org/10.1016/s0065-3233(08)60401-5).
- Tanford C** (1970) Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Advances in Protein Chemistry* **24**, 1–95.
- Tanford C, Aune KC and Ikai A** (1973) Kinetics of unfolding and refolding of proteins. 3. Results for lysozyme. *Journal of Molecular Biology* **73**(2), 185–197. [https://doi.org/10.1016/0022-2836\(73\)90322-7](https://doi.org/10.1016/0022-2836(73)90322-7).
- Tang KS, Fersht AR and Itzhaki LS** (2003) Sequential unfolding of ankyrin repeats in tumor suppressor p16. *Structure* **11**(1), 67–73. [https://doi.org/10.1016/s0969-2126\(02\)00929-2](https://doi.org/10.1016/s0969-2126(02)00929-2).
- Tartakower SG** (1924) *Die Hypermoderne Schachpartie*, 1st edn. Vienna: Verlag der Wiener Schachzeitung.
- Teilum K, Thormann T, Caterer NR, Poulsen HI, Jensen PH, Knudsen J, Kragelund BB and Poulsen FM** (2005) Different secondary structure elements as scaffolds for protein folding transition states of two homologous four-helix bundles. *Proteins-Structure Function and Bioinformatics* **59**(1), 80–90. <https://doi.org/10.1002/prot.20340>.
- Ternstrom T, Mayor U, Akke M and Oliveberg M** (1999) From snapshot to movie: Phi analysis of protein folding transition states taken one step further. *Proceedings of the National Academy of Sciences of the United States of America* **96**(26), 14854–14859. <https://doi.org/10.1073/pnas.96.26.14854>.
- Toto A, Malagrino F, Nardella C, Pennacchietti V, Pagano L, Santorelli D, Diop A and Gianni S** (2022) Characterization of early and late transition states of the folding pathway of a SH2 domain. *Protein Science* **31**(6), e4332. <https://doi.org/10.1002/pro.4332>.
- Toto A, Malagrino F, Visconti L, Troilo F, Pagano L, Brunori M, Jemth P and Gianni S** (2020) Templated folding of intrinsically disordered proteins.

- Journal of Biological Chemistry* **295**(19), 6586–6593. <https://doi.org/10.1074/jbc.REV120.012413>.
- Toto A, Troilo F, Visconti L, Malagrino F, Bignon C, Longhi S and Gianni S** (2019) Binding induced folding: Lessons from the kinetics of interaction between N(TAIL) and XD. *Archives of Biochemistry and Biophysics* **671**, 255–261. <https://doi.org/10.1016/j.abb.2019.07.011>.
- Troilo F, Bonetti D, Camilloni C, Toto A, Longhi S, Brunori M and Gianni S** (2018) Folding mechanism of the SH3 domain from Grb2. *Journal of Physical Chemistry B* **122**(49), 11166–11173. <https://doi.org/10.1021/acs.jpcc.8b06320>.
- Troilo F, Bonetti D, Toto A, Visconti L, Brunori M, Longhi S and Gianni S** (2017) The folding pathway of the KIX domain. *ACS Chemical Biology* **12**(6), 1683–1690. <https://doi.org/10.1021/acscchembio.7b00289>.
- Varnai P, Dobson CM and Vendruscolo M** (2008) Determination of the transition state ensemble for the folding of ubiquitin from a combination of phi and psi analyses. *Journal of Molecular Biology* **377**(2), 575–588. <https://doi.org/10.1016/j.jmb.2008.01.012>.
- Villegas V, Martinez JC, Aviles FX and Serrano L** (1998) Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain. *Journal of Molecular Biology* **283**(5), 1027–1036. <https://doi.org/10.1006/jmbi.1998.2158>.
- Visconti L, Malagrino F, Gianni S and Toto A** (2019) Structural characterization of an on-pathway intermediate and transition state in the folding of the N-terminal SH2 domain from SHP2. *FEBS Journal* **286**(23), 4769–4777. <https://doi.org/10.1111/febs.14990>.
- Wang GZ and Fersht AR** (2015) Mechanism of initiation of aggregation of p53 revealed by phi-value analysis. *Proceedings of the National Academy of Sciences of the United States of America* **112**(8), 2437–2442. <https://doi.org/10.1073/pnas.1500243112>.
- Wells TNC and Fersht AR** (1985) Hydrogen-bonding in enzymatic catalysis analyzed by protein engineering. *Nature* **316**(6029), 656–657. <https://doi.org/10.1038/316656a0>.
- Wells TN and Fersht AR** (1986) Use of binding energy in catalysis analyzed by mutagenesis of the tyrosyl-tRNA synthetase. *Biochemistry* **25**(8), 1881–1886. <https://doi.org/10.1021/bi00356a007>.
- Wells TN and Fersht AR** (1989) Protection of an unstable reaction intermediate examined with linear free energy relationships in tyrosyl-tRNA synthetase. *Biochemistry* **28**(23), 9201–9209. <https://doi.org/10.1021/bi00449a036>.
- Wensley BG, Batey S, Bone FA, Chan ZM, Tumelty NR, Steward A, Kwa LG, Borgia A and Clarke J** (2010) Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. *Nature* **463**(7281), 685–688. <https://doi.org/10.1038/nature08743>.
- Wensley BG, Gartner M, Choo WX, Batey S and Clarke J** (2009) Different members of a simple three-helix bundle protein family have very different folding rate constants and fold by different mechanisms. *Journal of Molecular Biology* **390**(5), 1074–1085. <https://doi.org/10.1016/j.jmb.2009.05.010>.
- Went HM and Jackson SE** (2005) Ubiquitin folds through a highly polarized transition state. *Protein Engineering Design & Selection* **18**(5), 229–237. <https://doi.org/10.1093/protein/gzi025>.
- Wetlaufer DB** (1973) Nucleation, rapid folding, and globular intrachain regions in proteins. *Proceedings of the National Academy of Sciences USA* **70**(3), 697–701. <https://doi.org/10.1073/pnas.70.3.697>.
- Wilson CJ and Wittung-Stafshede P** (2005) Snapshots of a dynamic folding nucleus in zinc-substituted *Pseudomonas aeruginosa* azurin. *Biochemistry* **44**(30), 10054–10062. <https://doi.org/10.1021/bi050342n>.
- Winter G, Fersht AR, Wilkinson AJ, Zoller M and Smith M** (1982) Redesigning enzyme structure by site-directed mutagenesis: Tyrosyl tRNA synthetase and ATP binding. *Nature* **299**(5885), 756–758. <https://doi.org/10.1038/299756a0>.
- Wong KB, Clarke J, Bond CJ, Neira JL, Freund SM, Fersht AR and 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. *Journal of Molecular Biology* **296**(5), 1257–1282. <https://doi.org/10.1006/jmbi.2000.3523>.
- Wu L, Zhang J, Qin M, Liu F and Wang W** (2008) Folding of proteins with an all-atom Go-model. *Journal of Chemical Physics* **128**(23), 235103. <https://doi.org/10.1063/1.2943202>.
- Yang F, Wang HB, Logan DT, Mu X, Danielsson J and Oliveberg M** (2018) The cost of long catalytic loops in folding and stability of the ALS-associated protein SOD1. *Journal of the American Chemical Society* **140**(48), 16570–16579. <https://doi.org/10.1021/jacs.8b08141>.
- Young BT and Silverman SK** (2002) The GAAA tetraloop-receptor interaction contributes differentially to folding thermodynamics and kinetics for the P4-P6 RNA domain. *Biochemistry* **41**(41), 12271–12276. <https://doi.org/10.1021/bi0264869>.
- Zarrine-Afsar A, Larson SM and Davidson AR** (2005) The family feud: Do proteins with similar structures fold via the same pathway? *Current Opinion in Structural Biology* **15**(1), 42–49. <https://doi.org/10.1016/j.sbi.2005.01.011>.
- Zhou Z, Huang Y and Bai Y** (2005) An on-pathway hidden intermediate and the early rate-limiting transition state of Rd-apocytochrome b562 characterized by protein engineering. *Journal of Molecular Biology* **352**(4), 757–764. <https://doi.org/10.1016/j.jmb.2005.07.057>.
- Zong C, Wilson CJ, Shen T, Wolynes PG and Wittung-Stafshede P** (2006) Phi-value analysis of apo-azurin hvfolding: Comparison between experiment and theory. *Biochemistry* **45**(20), 6458–6466. <https://doi.org/10.1021/bi060025w>.