Hammond Behavior versus Ground State Effects in Protein Folding: Evidence for Narrow Free Energy Barriers and Residual Structure in Unfolded States

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · MAY 2003

Impact Factor: 4.33 · DOI: 10.1016/S0022-2836(03)00171-2 · Source: PubMed

2 AUTHORS, INCLUDING:

Ignacio Enrique Sánchez
University of Buenos Aires
51 PUBLICATIONS **984** CITATIONS

SEE PROFILE

doi:10.1016/S0022-2836(03)00171-2 J. Mol. Biol. (2003) 327, 867–884

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Hammond Behavior *versus* Ground State Effects in Protein Folding: Evidence for Narrow Free Energy Barriers and Residual Structure in Unfolded States

Ignacio E. Sánchez and Thomas Kiefhaber*

Department of Biophysical Chemistry, Biozentrum der Universität Basel Klingelbergstrasse 70 CH-4056 Basel, Switzerland

Apparent transition state movement upon mutation or changes in solvent conditions is frequently observed in protein folding and is often interpreted in terms of Hammond behavior. This led to the conclusion that barrier regions in protein folding are broad maxima on the free energy landscape. Here, we use the concept of self-interaction and cross-interaction parameters to test experimental data of 21 well-characterized proteins for Hammond behavior. This allows us to characterize the origin of transition state movements along different reaction coordinates. Only one of the 21 proteins shows a small but coherent transition state movement in agreement with the Hammond postulate. In most proteins the structure of the transition state is insensitive to changes in protein stability. The apparent change in the position of the transition state upon mutation, which is frequently observed in ϕ-value analysis, is in most cases due to ground-state effects caused by structural changes in the unfolded state. This argues for significant residual structure in unfolded polypeptide chains of many proteins. Disruption of these residual interactions by mutation often leads to decreased folding rates, which implies that these interactions are still present in the transition state. The failure to detect Hammond behavior shows that the free energy barriers encountered by a folding polypeptide chain are generally rather narrow and robust maxima for all experimentally explorable reaction coordinates.

Keywords: protein folding; Hammond postulate; transition state; unfolded state; cross-interaction parameters

*Corresponding author

Introduction

The nature of the rate-limiting steps in the folding of polypeptide chains is still not well understood. For the characterization of free energy barriers of protein folding reactions several concepts from physical organic chemistry have been applied. A frequently used approach is the analysis of linear free energy relationships. For many reactions in organic chemistry the changes in activation free energy ($\Delta G^{0\ddagger}$) induced by changes in solvent conditions or by modifications in structure are linearly related to the corresponding changes in equilibrium free energy (ΔG^0) between reactants and products[1] (linear rate-equilibrium free energy relationships; REFERs). A proportionality constant, α, can be defined to quantify the energetic sensitivity of the transition state relative to the ground states in respect to a perturbation, ∂x:[1]

$$\alpha_x = \frac{\partial \Delta G^{0\ddagger}/\partial x}{\partial \Delta G^0/\partial x} \quad (1)$$

α_x is a measure for the position of the transition state along the reaction coordinate probed by ∂x and it is commonly used to gain information on the structural properties of the transition state. For many reactions the α_x-values are constant over a broad range of $\delta\Delta G^0$, indicating little effect of the perturbations on the structure of the transition state.[2] For other reactions, α_x is sensitive to changes in ΔG^0, which leads to non-linear REFERs, which indicate structural changes in the transition state relative to the ground states. It was shown that the characterization of non-linear rate-equilibrium free energy relationships can give valuable information on the mechanism of a reaction and on the

Abbreviations used: REFER, rate-equilibrium free energy relationship; GdmCl, guanidinium chloride.

E-mail address of the corresponding author: t.kiefhaber@unibas.ch

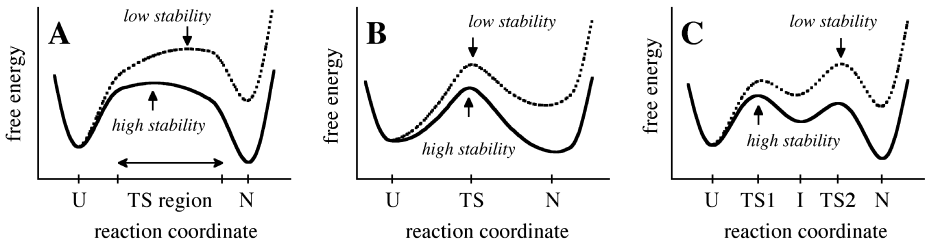

Figure 1. Schematic representation of the response of different types of free energy barriers to the same perturbation. The position of the transition state along the reaction coordinate is more sensitive to the perturbation if the free energy shows a broad maximum (A) than if the maximum is narrow (B). An apparent movement of the position of the transition state can also be due to a switch between consecutive transition states on a linear pathway (C).

properties of the transition barriers, since the non-linearities can have different origins: (i) a change in the rate-limiting step on a sequential pathway,[2] (ii) a change in the mechanism of the reaction due to a switch between parallel pathways;[2] (iii) a movement of the position of the transition state along a broad barrier region following Hammond behavior;[3,4] or (iv) structural changes in the ground state(s).[5] Detailed analyses of non-linear free energy relationships have therefore been widely used to elucidate reaction mechanisms in organic chemistry and the mechanism of complex biological reactions like the catalytic mechanisms of enzymes.[6]

In protein folding studies various perturbations like changes in pressure, in the concentration of a chemical denaturant or changes in the amino acid sequence have been shown to induce both linear and non-linear REFERs.[7] In particular, mutational studies and denaturant-induced folding/unfolding experiments revealed that the position of the transition state may significantly change with protein stability.[8–13] These results were commonly interpreted in terms of the Hammond postulate. According to the Hammond postulate the position of a transition state is significantly shifted towards the ground state that is destabilized by the perturbation. Accordingly, α_x increases if the native state is destabilized relative to the unfolded state, which will lead to curvatures in REFERs. In principle, Hammond behavior should be observed for any transition state. If the free energy landscape in the vicinity of the transition state is a broad and smoothly curved maximum, α_x will be strongly influenced by changes in protein stability[4] (Figure 1A). If the transition region represents a rather narrow free energy maximum on the reaction coordinate, the changes in α_x will be too small to be detected experimentally[4] (Figure 1B). The sensitivity of α_x to changes in protein stability can thus be used to obtain information on the broadness of the free energy barriers. From the observation of non-linear free energy relationships it was concluded that folding reactions of many proteins proceed over a single broad barrier region with a continuum of states with similar free energy[13,14] (Figure 1A). However, also other origins have been demonstrated to cause apparent transition state movements in protein folding reactions. The frequently observed denaturant-induced changes in the position of the transition state can be explained by a switch between a few consecutive narrow barriers on a sequential folding pathway[15–18] (Figure 1C). The apparent transition state movement in destabilized variants of protein L was shown to be due to structural changes in the unfolded state with an unchanged structure of the transition state.[7] This study showed that mutations in proteins can lead to significant ground state effects, which can easily be mistaken for transition state movements.

To understand the origins of experimentally observed transition state movements in protein folding reactions and to characterize the properties of the free energy barriers we analyzed published data for 21 different small single domain proteins. The data set includes REFERs induced by changes in denaturant concentration, pressure, temperature, pH and structure. About half of the proteins for which extensive mutational studies have been performed were postulated to show Hammond behavior. Our analysis shows that in most cases these transition state movements are due to ground state effects. In particular, residual structure in the unfolded state seems to be frequently disrupted by mutations, whereas the structure of the transition state usually remains unchanged relative to the native state. Only one protein in our data set shows consistent transition state movement detected by several different perturbations, compatible with Hammond behavior. These results suggest that protein folding transition states are for most proteins narrow and robust maxima in the free energy landscape.

Theory

Linear rate equilibrium free energy relationships in protein folding

Protein folding reactions can be perturbed in different ways to gain information on different properties of the transition state.[7] This allows us to probe different reaction coordinates for the folding reaction. α_p and α_T-values can be obtained

by changes in pressure at constant temperature or changes in temperature at constant pressure, respectively:[7]

$$\alpha_p = \frac{\partial \Delta G^{0\ddagger}/\partial p}{\partial \Delta G^0/\partial p} = \frac{\Delta V^{0\ddagger}}{\Delta V^0} \quad (2)$$

$$\alpha_T = \frac{\partial \Delta G^{0\ddagger}/\partial T}{\partial \Delta G^0/\partial T} = \frac{\Delta S^{0\ddagger}}{\Delta S^0} \quad (3)$$

α_p and α_T give information on the volume and on the entropy of the transition state, respectively.[19,20] From equation (3) we can easily calculate the changes in enthalpy along the reaction coordinate using the Gibbs-Helmholtz equation. The value of α_T changes with temperature due to the change in molar heat capacity (ΔC_p^0) associated with protein folding reactions.[21] This allows the definition of another reaction coordinate:[22]

$$\alpha_C = \frac{\Delta C_p^{0\ddagger}}{\Delta C_p^0} \quad (4)$$

α_C allows the characterization of a transition state in terms of its relative solvent exposure, since ΔC_p^0 mainly arises from differences in interactions of the protein with the solvent.[23]

The addition of chemical denaturants like urea or guanidinium chloride (GdmCl) is the most commonly used perturbation in the study of protein folding reactions. The concentration of these chaotropic compounds generally has a linear effect on both the equilibrium free energy (ΔG^0)[24,25] and on the activation free energies[26] for folding $\Delta G_f^{0\ddagger}$ and unfolding $\Delta G_u^{0\ddagger}$. The proportionality constants, m, are defined as:

$$m_{eq} = \frac{\partial \Delta G^0}{\partial [\text{Denaturant}]} \quad (5)$$

$$m_{f,u} = \frac{\partial \Delta G_{f,u}^{0\ddagger}}{\partial [\text{Denaturant}]} \quad (6)$$

and a denaturant-induced free energy relationship can be obtained:[26]

$$\alpha_D = \frac{\partial \Delta G_f^{0\ddagger}/\partial [\text{Denaturant}]}{\partial \Delta G^0/\partial [\text{Denaturant}]} = \frac{m_f}{m_{eq}} \quad (7)$$

α_D reflects the relative sensitivity of the transition state to changes in denaturant concentration. Since the m_{eq}-value was shown to be proportional to the changes in accessible surface area upon unfolding of the protein,[27] α_D is interpreted as the relative amount of accessible surface area buried in the transition state and it should correlate with α_C.

Other solvent additives or ligands such as alcohols (2,2,2-trifluoroethanol), polyols (glycerol, sugars), salts (Na_2SO_4, NaCl), ^2H_2O as solvent or proton concentration (change in pH) can also change the stability of a protein and can therefore be used to define the corresponding α-values.[7]

Equations (2)–(4) and (7) can be considered as medium-induced rate-equilibrium free energy relationships. The effect of structural changes introduced by protein engineering are commonly used to obtain a structure-induced free energy relationship:[28–30]

$$\alpha_S = \phi_f = \frac{\partial \Delta G^{0\ddagger}/\partial \text{Structure}}{\partial \Delta G^0/\partial \text{Structure}} \quad (8)$$

α_S, which is commonly called ϕ_f, reports on the energetics of the interactions formed by a side-chain with the rest of the protein in the transition state relative to the native state. The free energy of the unfolded state serves as a reference. It can be defined for a single residue (ϕ_f^i), for the whole protein[31] (ϕ_f^{prot}) or for a substructure.[32]

It should be noted that the type of free energy relationships described here are commonly termed "Brønsted plots" in protein folding literature, which is misleading. Brønsted plots were originally used to relate the effect of a change in the rate constant of an acid or base-catalyzed reaction to the dissociation constant of the catalyst. The rate-equilibrium relationships considered here are rather Leffler-type relationships (equation (1)), since they directly relate the rate constants of a reaction to the equilibrium constants of the same reaction.

Self-interaction and cross-interaction parameters

A practical way to detect and to analyze transition state movements was proposed by Jencks and co-workers[4] by defining self-interaction and cross-interaction parameters. A self-interaction parameter (p_x) measures the shift in the position of the transition state along the reaction coordinate with changing ΔG^0 upon a perturbation ∂x:

$$p_x = \frac{\partial \alpha_x}{\partial \Delta G_x^0} = \frac{\partial^2 \Delta G^{0\ddagger}}{(\partial \Delta G_x^0)^2} \quad (9)$$

By definition, a shift in the position of the transition state towards the destabilized state, e.g. as a result of Hammond behavior or due to sequential barriers, will give a positive p_x-value.

For a folding reaction perturbed by addition of denaturant or destabilized by mutations the corresponding self-interaction parameters p_D and p_S are:[7]

$$p_D = \frac{\partial \alpha_D}{\partial \Delta G_D^0} = \frac{\partial^2 \Delta G^{0\ddagger}}{(\partial \Delta G_D^0)^2} \quad (10)$$

$$p_S = \frac{\partial \phi_f}{\partial \Delta G_S^0} = \frac{\partial^2 \Delta G^{0\ddagger}}{(\partial \Delta G_S^0)^2} \quad (11)$$

The use of self-interaction parameters to detect transition state movements is often not sensitive enough, because the energy range of the measurements is too narrow or the curvatures are too small.[4] A more sensitive test for transition state

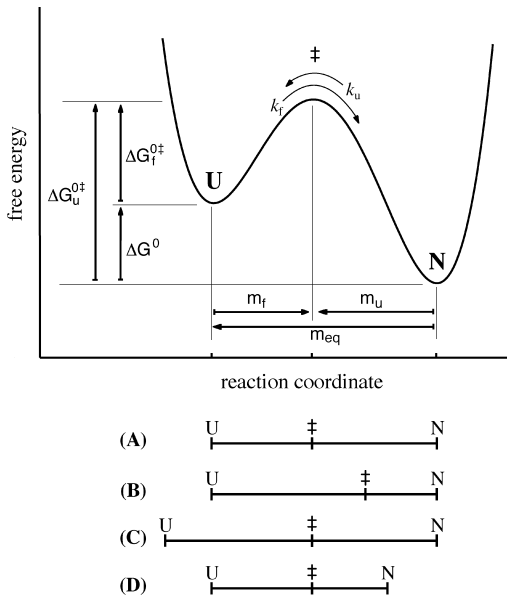

Figure 2. Schematic representation of the possible effect of a perturbation on the position of the ground states and the transition state along a hypothetical reaction coordinate. The reference conditions (A) are compared to real Hammond behavior (B) and to apparent transition state movements caused by ground state effects due to changes in the structure of the unfolded (C) and native state (D). In all three cases (A)–(C) the position of the transition state relative to the ground states (α_x) will change by the same amount.

movements is provided by cross-interaction parameters,[4] which measure changes in the position α_x of the transition state of a reaction (measured using the perturbation ∂x) caused by a second perturbation ∂y:

$$p_{xy} = \frac{\partial \alpha_x}{\partial \Delta G_y^0} = \frac{\partial^2 \Delta G^{0‡}}{(\partial \Delta G_x^0)(\partial \Delta G_y^0)} = \frac{\partial \alpha_y}{\partial \Delta G_x^0} = p_{yx} \quad (12)$$

For non-linear REFERs the value of α_x will change with the amount of a second perturbation ∂y, resulting in a non-zero p_{xy}-value. A shift in the position of the transition state towards the destabilized state will yield positive p_{xy}-values.

In mutational studies the changes in kinetic and equilibrium free energies are usually determined as a function of the denaturant concentration for each mutant. This directly allows the calculation of the denaturant-structure cross-interaction parameter (p_{DS}):[7]

$$p_{DS} = \frac{\partial \alpha_D}{\partial \Delta G_S^0} = \frac{\partial^2 \Delta G^{0‡}}{(\partial \Delta G_S^0)(\partial \Delta G_D^0)} = \frac{\partial \phi_f}{\partial \Delta G_D^0} = p_{SD} \quad (13)$$

p_{DS} tests whether changes in the stability of the native state caused by mutations ($\partial \Delta G_S^0$) have an effect on the relative solvent exposure of the transition state for folding (α_D).

If the effect of a change in structure ∂S is tested in the wild-type and mutant ($\partial S'$) backgrounds, we can calculate a structure–structure cross-interaction parameter:

$$p_{SS'} = \frac{\partial \phi_f^S}{\partial \Delta G_{S'}^0} \quad (14)$$

$p_{SS'}$ indicates whether ϕ_f ($= \alpha_S$) changes when the protein stability is altered by an additional mutation.[33]

Ground state effects

Changes in the length of the reaction coordinate due to ground state effects can easily be mistaken for genuine transition state movement.[5,7] Both phenomena have very similar effects on the experimentally observed data but they are based on completely different free energy barriers. Figure 2 shows possible effects of a perturbation ∂x on the location of the transition state for a protein folding reaction with the reference conditions given in Figure 2(A). In Figure 2(B) a perturbation leads to a transition state movement along the reaction coordinate relative to both ground states. The length of the reaction coordinate is unchanged, i.e. the structure of both ground states is not affected by the perturbation or they are affected to the same extent. This scenario is in accordance with Hammond behavior if we assume that the native state is destabilized by the perturbation. Figure 2(C) and (D) shows apparent transition state movements caused by ground state effects. In both cases the absolute position and the structure of the transition state remain unchanged but the structure of either the unfolded state (Figure 2(C)) or of the native state (Figure 2(D)) changes, which leads to a change in length of the reaction coordinate. In this case an apparent transition state movement will be observed if the position of the transition state is normalized against the length of the reaction coordinate. This shows that the characterization of transition state movements requires the determination of the effects of a perturbation on the rate constants for the forward and backward reaction, on the equilibrium constant and on the length of the reaction coordinate.

Sequential *versus* parallel transition states

Non-linear REFERs provide a unique tool to characterize complex reaction mechanisms and to distinguish between sequential and parallel pathways.[2] If destabilization of the native state leads to a shift in the rate-limiting step to a more native-like transition state (larger α_x) on a sequential pathway a positive p_x value will be observed. If, in contrast, folding occurs along parallel pathways, the reaction with the most native-like transition state will be most sensitive to a reduced stability of the native-state and an alternative parallel pathway with a more unfolded-like transition state can become rate-limiting. This leads to a decrease in α_x with decreasing protein stability, which will result in a negative p_x-value. This

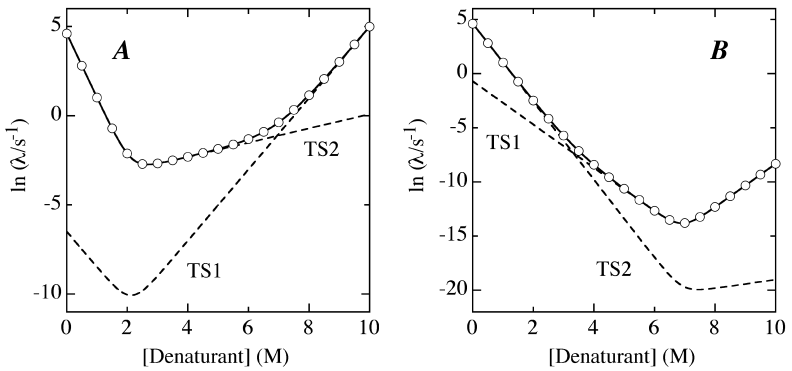

Figure 3. Simulated effect of denaturant concentration on the apparent rate constant (○ and continuous line) for the folding of a protein through two parallel pathways with the transition states TS1 and TS2. The individual chevron plots corresponding to folding over TS1 and TS2 are shown as broken lines. For two parallel pathways with the folding and unfolding rate constants of k_{f1}, k_{u1} (for TS1) and k_{f2}, k_{u2} (for TS2) the single observable rate constant (λ) is given by the sum of the two apparent rate constants ($\lambda_1 = k_{f1} + k_{u1}$ and $\lambda_2 = k_{f2} + k_{u2}$) for the parallel pathways: $\lambda = \lambda_1 + \lambda_2$. Data were simulated with the rate constants for folding and unfolding of (A) $k_{f1} = 0.0015\ s^{-1}$, $k_{u1} = 3 \times 10^{-7}\ s^{-1}$; $k_{f2} = 100\ s^{-1}$; $k_{u2} = 0.02\ s^{-1}$ and (B) $k_{f1} = 5\ s^{-1}$, $k_{u1} = 3 \times 10^{-7}\ s^{-1}$; $k_{f2} = 100\ s^{-1}$; $k_{u2} = 0.02\ s^{-1}$. The m_i-values ($m_i = RT(\partial \ln k_i/\partial[\text{Denaturant}])$) were $m_{f1} = -4.95$ (kJ/mol)/M, $m_{u1} = 4.95$ (kJ/mol)/M; $m_{f2} = -8.92$ (kJ/mol)/M; $m_{u2} = 0.99$ (kJ/mol)/M in both cases. This corresponds to α_D-values of 0.5 and 0.9 for TS1 and TS2, respectively. The plots show that significantly different α_D-values for the two parallel pathways are required to obtain a clear non-linearity.

phenomenon is sometimes referred to as anti-Hammond behavior.[11,32] Figure 3 shows the effect of increasing denaturant concentration on the apparent rate constant (λ) for folding for a mechanism with two parallel pathways. Obviously, if the transition states on the parallel pathways have different α_x-values an upward curvature (negative p_D-value) in the denaturant dependence of ln λ will be observed, since folding always proceeds over the lowest barrier.

Results

To analyze the properties of the transition barriers in protein folding we calculated self-interaction and cross-interaction parameters for published data on folding of 21 different proteins. The perturbations include structural changes (mutations) as well as variations in temperature, pressure, pH and denaturant concentration. In addition, we tested the data set for ground state effects in order to be able to characterize effects of the perturbations on the complete reaction coordinate. We considered data on small monomeric single domain proteins that belong to a variety of structural classes: all-helical proteins (ACBP, cyt *b*562, Im7, λ repressor), αβ-proteins (ctAcP, mAcP, ADA2h, barnase, CheY, CI2, CTL9, FKBP12, NTL9, protein G, protein L) and all-β proteins (CspB, drkN SH3, fyn SH3, spectrin SH3, Sso7d SH3, tendamistat). In the analysis of the folding of barnase the core region and helix 1 were analyzed separately, since they show different behavior in the analysis of transition state movement.[8,11,32,34] The behavior of a fifth SH3 domain (c-src SH3) is more complex and will be discussed elsewhere together with additional results from our laboratory (I.E.S. & T.K., unpublished results). The chain length of the proteins varies from 58 to 130 residues and the folding rates differ by more than four orders of magnitude. The denaturants used were either urea (for 11 of the proteins) or GdmCl (for ten of the proteins). The diversity of our database in terms of structure, size, denaturant and folding rates makes it suitable to look for common properties of protein folding transition states.

Denaturant self-interaction parameter

The most commonly applied perturbation in protein folding is a change in denaturant concentration. Linear denaturant-dependencies of $\Delta G_f^{0\ddagger}$ and $\Delta G_u^{0\ddagger}$ have been observed for many two-state folders over the entire experimentally accessible range of denaturant concentrations.[35] However, for about 50% of the two-state folders the α_D-value increases with increasing denaturant concentration, indicating a shift to a more native-like transition state at high denaturant concentrations.[18] This phenomenon was commonly ascribed to Hammond behavior.[9] However, for most proteins only some variants show positive p_D-values, which cannot be explained by the existence of a single broad free energy barrier. In a previous work we showed that a detailed comparison of data for several mutants of the same protein and of experiments carried out under different solvent conditions suggests that positive p_D-values can be explained by a switch between two or three distinct transition states on a sequential pathway.[18] This model was supported by the observation that the α_D-values change in a narrow range of denaturant concentrations for many proteins with constant α_D values at higher and lower denaturant concentrations.[17,18] These results argued against Hammond behavior and suggested protein folding landscapes with high energy intermediates and consecutive narrow barriers.

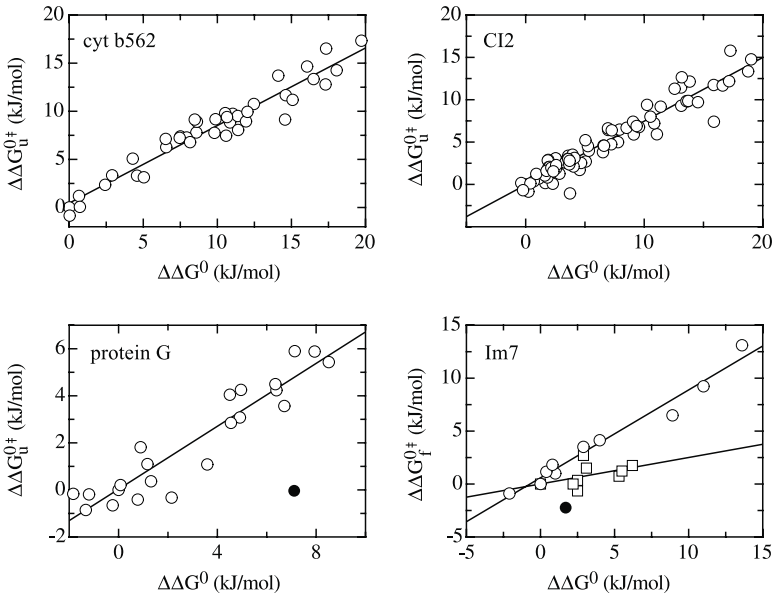

Figure 4. Effect of changes in protein stability caused by mutations of protein side-chains on the change in free energy of activation for unfolding ($\Delta\Delta G_u^{0\ddagger}$) for cyt *b*562, CI2 and protein G and the free energy of activation for folding ($\Delta\Delta G_f^{0\ddagger}$) for Im7. The lines represent linear fits of the data and their slopes correspond to $1 - \langle\phi_f\rangle$[32] for cyt *b*562, CI2 and protein G and $\langle\phi_f\rangle$ for Im7 (see Table 1). For protein G and Im7 a second transition state seems to become rate-limiting in some variants (●), indicated by the drastic changes in all parameters.

To test for evidence for parallel folding pathways (negative p_D-values; Figure 3) we analyzed the published data on the denaturant-dependence of protein folding reactions. Several proteins show slight upward curvatures in the denaturant dependence of ln λ, indicating a $p_D < 0$ (data not shown). However, the effect is too small to unequivocally attribute it to parallel folding pathways. The clearest example for a negative p_D-value was published for the c-src SH3 domain, which shows a strong upward curvature with increasing denaturant concentration in the unfolding region of the ln λ *versus* [denaturant] plot.[36] However, these data could not be reproduced in our laboratory using the same c-src SH3 domain (I.E.S. & T.K., unpublished results).

The results from the p_D-value analysis show that the analysis of the denaturant dependence of protein folding rates does not yield any clear evidence for parallel folding pathways or Hammond behavior. We therefore analyzed other self and cross-interaction parameters to get a more detailed picture on the shape of the free energy barriers in protein folding.

Structure self-interaction parameters

In ϕ_f-value analysis usually single mutations are introduced at many different positions in the protein. This gives a single data point at each position for the determination of the ϕ_f-value, which does not allow us to test for linear free energy relationships. In two recent studies by Davidson and co-workers, several mutations at three different positions in two SH3 domains were analyzed.[37,38] The results indicated a p_S-value of zero for all three positions. An alternative way to obtain ϕ_f-values is a global analysis of all mutations that have been introduced at different positions of the same protein.[31] This gives average ϕ_f and p_S-values for the complete structure. In order to draw conclusions on the properties of the transition state from a global analysis it is necessary that all mutagenized positions are evenly distributed in the structure of the protein and have similar ϕ_f-values. This scenario is sometimes called "diffuse" or "delocalized" transition state, which is the case for eight of the 14 proteins for which enough data on p_s are available (ADA2h, mAcP, CI2, cyt *b*562, FKBP12, fyn SH3, protein G and Sso7d SH3). Representative examples for p_S plots of cyt *b*562, CI2 and protein G are shown in Figure 4. Linear relationships between protein stability and folding rate are observed for all proteins in this group, indicating a zero p_S-value and an overall transition state structure that is insensitive to mutations. Figure 4 reveals a large amount of scattering in the plots, which might be due to different behavior of different structural regions of the proteins or to each position being probed with only one mutation. For another six proteins in our study (ACBP, barnase, CheY, Im7, protein L and spectrin SH3) different substructures of the protein clearly show different average ϕ_f-values and therefore have been analyzed independently.[32,39] Figure 4 shows the p_S analysis of Im7 as a representative example of this class of proteins with a "polarized" transition state structure. The data indicate two structural regions with $\langle\phi_f\rangle$-values of 0.83(±0.05) and 0.25(±0.18). However, for each region the p_S-value is zero, indicating that the structure of the transition state is insensitive to changes in protein stability. A non-zero p_S-value was only observed for the first helix of barnase.[32] In this case an abrupt increase in the slope of $\langle\phi_f\rangle$ occurs in a narrow range of ΔG_S^0 ($p_S < 0$). At higher and lower ΔG_S^0-values $\langle\phi_f\rangle$ is constant, indicating the existence of two parallel pathways for the folding of barnase. Analysis of the data showed $\langle\phi_f\rangle$-values of 0.33(±0.04) and 0.78(±0.03) for the first helix of

Table 1. Structural and thermodynamic information on the proteins used in the calculation of denaturant-structure cross-interaction parameters (p_{DS})

Protein	Structure	Number of variants	Distribution of mutations	Maximum $\partial\Delta G^0_S$ (kJ/mol)	Denaturant
ACBP[63]	α	29	Core	21	GdmCl
Cyt *b*562[64]	α	43	~Uniform	27	GdmCl
Im7[65]	α	28	~Uniform	27	Urea
λ Repressor[66]	α	8	A single position	15	Urea
ctAcP[67]	α/β	7	Helices 1 and 2	8	Urea
mAcP[68]	α/β	27	~Uniform	23	Urea
ADA2h[69]	α/β	20	~Uniform	15	Urea
Barnase core[8,70]	α/β	9	Main core	15	Urea
Barnase helix 1[11]	α/β	23	Helix 1	20	Urea
CheY[39]	α/β	43	~Uniform	26	Urea
CI2[31]	α/β	83	~Uniform	19	GdmCl
FKBP12[71]	α/β	32	~Uniform	17	Urea
Protein G[72]	α/β	24	~Uniform	10	GdmCl
Protein L[73]	α/β	62	~Uniform	24	GdmCl
CspB[74]	β	14	Chain termini	9	GdmCl
drkNSH3[37]	β	9	Diverging turn	12	GdmCl
fyn SH3[38,75]	β	69	~Uniform	19	GdmCl
Spectrin SH3[76]	β	13	~Uniform	11	Urea
Sso7d SH3[77]	β	21	~Uniform	14	GdmCl

barnase in the two parallel transition states.[32] All other substructures of proteins in our data show p_S-values of zero, indicating that also the structures of polarized transition states are generally insensitive to mutation (data not shown).

Denaturant-structure cross-interaction parameters

To increase the sensitivity of the analysis of REFERs we determined cross-interaction parameters.

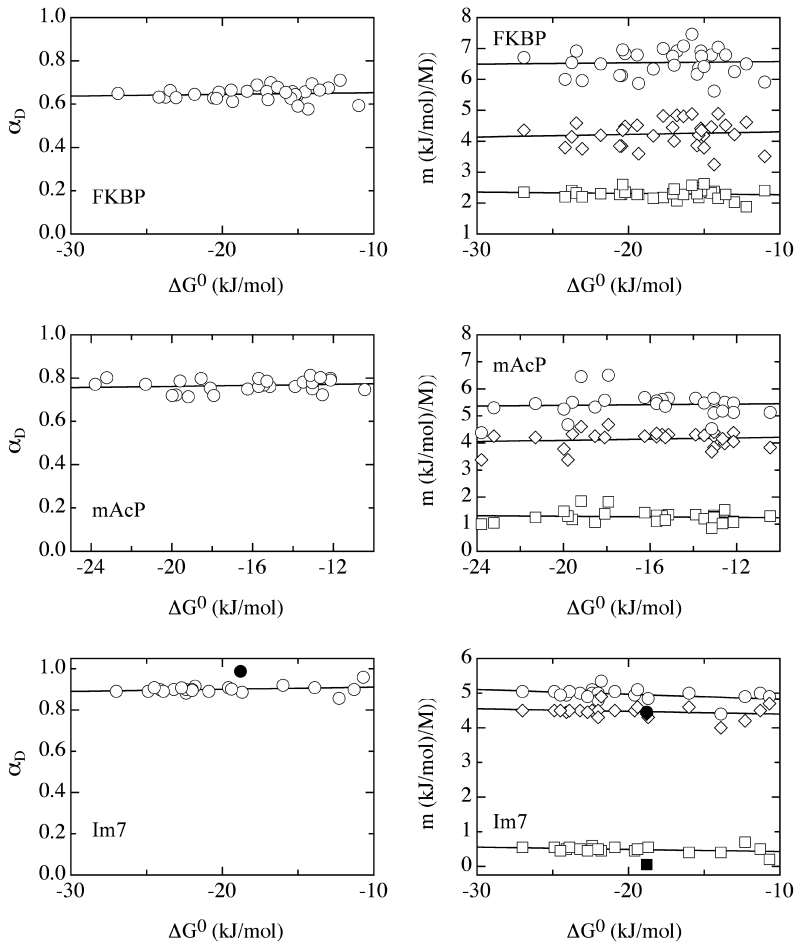

Figure 5. Effect of changes in ΔG^0_S on α_D (left panels) and on m_{eq} (○), m_f (◇) and m_u (□; right panels) for FKBP12, mAcP and Im7. Linear fits to the data show that there is no significant effect of ΔG^0_S on α_D or on the *m*-values (see Table 2). For one variant of Im7 a second transition state seems to become rate-limiting (●), indicated by the drastic changes in all parameters.

Table 2. Effect of changes in protein stability induced by mutations ($\partial\Delta G^0_S$) on the relative solvent accssibility of the transition state (α_D-value) for proteins that have been extensively mutagenized

Protein	$p_{DS} = \partial\alpha_D/\partial\Delta G^0_S$ (mol/J)	$\partial m_{eq}/\partial\Delta G^0_S$ ($M^{-1} \times 10^2$)	$\partial m_f/\partial\Delta G^0_S$ ($M^{-1} \times 10^2$)	$\partial m_u/\partial\Delta G^0_S$ ($M^{-1} \times 10^2$)	$(\partial m_f/\partial\Delta G^0_S)/(\partial m_{eq}/\partial\Delta G^0_S)$	$(\partial m_u/\partial\Delta G^0_S)/(\partial m_{eq}/(\partial\Delta G^0_S)$	$\partial m_f/\partial m_{eq}$
ACBP[63]	6.5 ± 2.3	≅ 0	≅ 0	≅ 0			
ctAcP[67]	≅ 0	12.4 ± 3.7	≅ 0	≅ 0	0.80 ± 0.47	0.20 ± 0.15	0.95 ± 0.14
mAcP[68]	≅ 0	≅ 0	≅ 0	≅ 0			
ADA2h[69]	6.6 ± 1.3	4.3 ± 1.2	5.8 ± 1.1	−1.5 ± 0.6	1.35 ± 0.44	−0.35 ± 0.17	1.03 ± 0.11
Barnase (core)[8,70], [a,b]	2.5 ± 1.2	≅ 0	≅ 0	−2.7 ± 0.7			
Barnase (helix 1)[11], [a,b]	≅ 0	≅ 0	≅ 0	≅ 0			
CheY[39], [a]	≅ 0	−4.4 ± 1.0	−3.7 ± 1.0	≅ 0	0.84 ± 0.30	0.16 ± 0.12	0.86 ± 0.06
CI2[31]	5.2 ± 0.7	6.2 ± 1.4	8.6 ± 1.4	−2.5 ± 0.4	1.40 ± 0.39	−0.40 ± 0.11	1.04 ± 0.04
CspB[74]	≅ 0	≅ 0	≅ 0	≅ 0			
Cyt *b*562[64]	4.3 ± 1.3	−4.5 ± 1.6	≅ 0	−5.1 ± 1.0	−0.15 ± 0.29	1.15 ± 0.48	0.48 ± 0.08
drkNSH3[37]	≅ 0	≅ 0	≅ 0	≅ 0			
FKBP12[71]	≅ 0	≅ 0	≅ 0	≅ 0			
Fyn SH3[38,75]	3.1 ± 1.2	−4.7 ± 1.3	≅ 0	−3.8 ± 0.8	0.20 ± 0.33	0.80 ± 0.28	0.85 ± 0.07
Im7[65], [a]	≅ 0	≅ 0	≅ 0	≅ 0			
λ Repressor[66]	≅ 0	≅ 0	≅ 0	≅ 0			
Protein G[72]	4.5 ± 2.0	7.6 ± 3.3	9.6 ± 3.2	≅ 0	1.27 ± 0.69	−0.27 ± 0.25	0.94 ± 0.10
Protein L[73]	3.8 ± 0.8	19.0 ± 2.2	17.6 ± 2.0	≅ 0	0.92 ± 0.15	0.08 ± 0.04	0.90 ± 0.03
Spectrin SH3[76], [b]	≅ 0	4.2 ± 1.8	4.8 ± 2.4	≅ 0	1.15 ± 0.76	−0.15 ± 0.45	0.92 ± 0.25
Sso7d SH3[77]	6.8 ± 1.0	13.4 ± 2.6	13.8 ± 2.1	≅ 0	1.03 ± 0.26	−0.03 ± 0.07	0.92 ± 0.05

α_D, ΔG^0_S and m_{eq}-value were obtained from kinetic data except for FKBP12, CheY and barnase, for which ΔG^0_S and m_{eq} were determined in equilibrium measurements. The numbers and standard deviations represent the result of linear fits to the data.

[a] These proteins populate an intermediate during refolding. Only the transition state of highest energy is considered in our analysis.

[b] For these proteins the m_u-value varies with denaturant concentration and α_D in water was used in our analysis.

The popularity of the ϕ_f-value analysis in combination with denaturant-induced folding/unfolding measurements provides a wealth of data on the effect of a change in ΔG^0 on the α_D-value for several proteins, i.e. on the relative change in solvent-accessible surface area in the transition state. To test these data sets for transition state movements we calculated the denaturant-structure cross-interaction parameter (p_{DS}) for 18 proteins that have been extensively mutagenized (Table 1). Of the 18 proteins, nine have α_D-values that are independent of ΔG^0 ($p_{DS} = 0$), which argues against transition state movement (Figure 5, Table 2). For eight proteins, α_D changes significantly with protein stability, as indicated by positive p_{DS}-values in Table 2 and Figures 6 and 8. This demonstrates a significant change in the relative position of the transition state for these proteins.

To discriminate between ground state effects and Hammond behavior we calculated the effect of $\partial\Delta G^0_S$ on the equilibrium *m*-value (m_{eq}) and on both kinetic *m*-values (m_f and m_u) for all proteins in our data set (Table 2). If the structures of the ground states are not affected by the mutations the m_{eq}-value should not change. The analysis shows that the proteins can be divided into different groups (Table 2). For six proteins (mACP, CspB, FKBP, Im7, λ repressor, drkNSH3) and for the first helix of barnase both m_{eq} and the position of the transition state ($\alpha_D = m_f/m_{eq}$) remain unchanged upon mutation (Figure 5). Obviously, in these proteins the structures of the unfolded state, of the native state and of the transition state are insensitive to mutations. For ten proteins (ctAcP, ADA2h, CheY, CI2, cyt *b*562, protein G, protein L, fyn SH3, spectrin SH3, Sso7d SH3) the m_{eq}-value changes significantly upon mutation, indicating structural changes in at least one of the ground states. In CheY, fyn SH3 and cyt *b*562 the m_{eq}-value decreases when the protein is destabilized and for the other proteins it increases (Table 2). The effect of protein stability on the kinetic and equilibrium *m*-values for Sso7d SH3, protein G and protein L, which show an increased m_{eq}-value with decreasing protein stability, is shown in Figure 6. In these proteins m_f changes by the same amount as m_{eq} whereas m_u is essentially not affected by the change in m_{eq} (Figure 7). This behavior is found for all proteins with an increased m_{eq}-value, indicating that in these proteins the structure of the transition state does not move relative to the structure of the native state (Table 2). Obviously, the apparent shift of the position of the transition state along the reaction coordinate seen for these proteins is due to a ground state effect rather than Hammond behavior. The observed increase in the m_{eq}-value is most likely caused by the disruption of interactions in the unfolded state upon mutation which leads to a less compact structure. Another possible origin of an increased m_{eq} value is the destabilization of equilibrium intermediates in destabilized variants. This will lead to an apparent increase in the m_{eq}-values, when the transitions are fit to a two-state model.[26,40] However, there is no evidence for significantly populated equilibrium intermediates in any of the proteins from our data set that show increased m_{eq}-values.

The decreased m_{eq}-value in destabilized mutants of cyt *b*562, fyn SH3 and CheY might be due to a

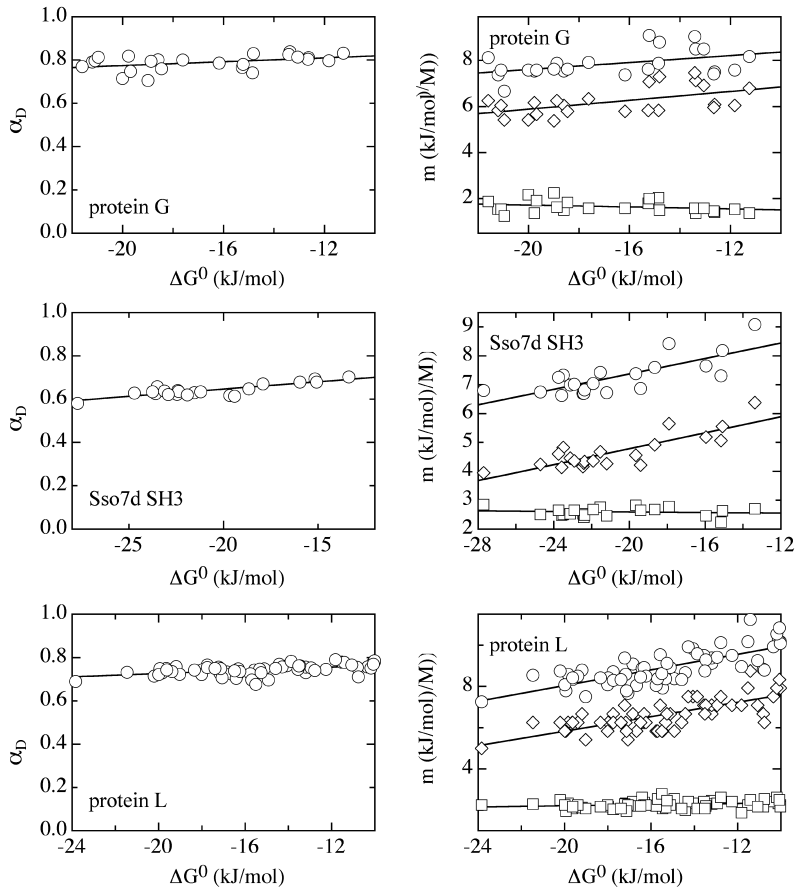

Figure 6. Effect of changes in ΔG^0_S on α_D (left panels) and on m_{eq} (○), m_f (◇) and m_u (□) (right panels) for protein G, Sso7d SH3 and protein L. Linear fits of the data show that α_D, m_{eq} and m_f are sensitive to changes in ΔG^0_S, whereas m_u is constant (see Table 2).

ground state effect on the native state. It is very unlikely that destabilizing mutations, which are commonly obtained by truncation of side-chains, lead to a more compact unfolded state. An alternative source for this effect is an increased population of equilibrium intermediates, which will decrease the apparent m_{eq}-value in fits using a two-state model.[26,40] In fyn SH3 and cyt *b*562 there is no evidence for significantly populated equilibrium intermediates, which makes a ground state effect on the native state the most likely source for the decreased m_{eq}-value. In these proteins m_f is not sensitive to mutation and the change in m_u matches the change in m_{eq} (Table 2). This indicates that in these two proteins the native state properties altered by the mutations are not present in the transition state, which argues against Hammond behavior. In CheY the m_u-value does not change significantly. Here the change in m_{eq} matches the change in m_f (Table 2). For CheY an equilibrium intermediate has been described,[41] which makes deviations from a two-state mechanism the most likely source for the decreased m_{eq}-value in destabilized variants.

Three proteins (ACBP, ADA2h and CI2) show a significant non-zero p_{DS}-value that cannot be explained solely by ground state effects (Figures 8 and 9; Table 2). In ACBP the mutations have only little effect on the m_{eq}-value, indicating a true transition state movement relative to both the native and the unfolded state, in agreement with Hammond behavior. For CI2 and ADA2h the m_{eq}-value strongly increases when the protein is destabilized, compatible with disruption of residual structure in the unfolded state upon mutation. This change in m_{eq} is, however, overcompensated by an even larger increase in m_f (Table 2; Figures 8 and 9), indicating Hammond behavior in addition to the ground-state effect. Figure 9 shows that for CI2 and ADA2h the major source for the observed transition state movement is the ground state effect since only minor changes in m_u are observed compared to the changes in m_{eq} and m_f.

We further analyzed the structure-denaturant cross-interaction parameters for the first two groups of proteins to test for local transition state movements of substructures. In agreement with the previous analysis from Fersht and co-workers[8] a non-zero p_{DS}-value was seen for mutations in the core of barnase, which is highly structured in the transition state. In this case the mutations did not affect m_{eq}, indicating Hammond behavior without contributions from ground state effects. For all other proteins in our data set we were not able to detect Hammond behavior in any substructure.

The results from the analysis of denaturant-structure cross-interaction parameters show that only three out of 18 proteins and the core of barnase show a small but significant shift in the

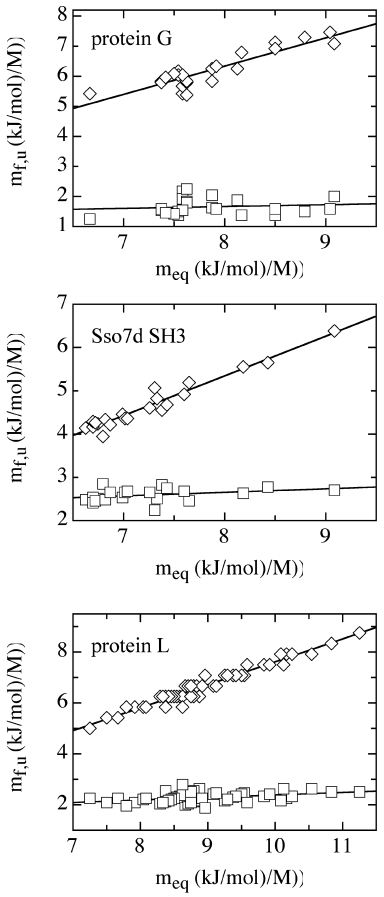

Figure 7. Correlation of the changes in m_{eq} with changes in m_u (◇) and m_f (□) for protein G, Sso7d SH3 and protein L. Linear fits of the data indicate that the variation in m_{eq} is correlated with the changes in m_f, whereas m_u is virtually not affected by changes in m_{eq}. Correlation coefficients between m_f and m_{eq} are 0.90, 0.97 and 0.98 for protein G, Sso7d SH3 and protein L, respectively. The respective correlation coefficients between m_u and m_{eq} are 0.13, 0.34 and 0.44.

position of the transition state relative to the native state when the protein is destabilized by a structural change. Only in ACBP and in the core of barnase are the transition state movements not accompanied by ground state effects. An interesting observation in the p_{DS}-analysis is a discrete jump in the position of the transition state for some mutants in several proteins (filled symbols in Figures 5 and 8). Obviously, in these proteins the mutations lead to a complete change in the structure of the transition state compatible with a change in the rate-limiting step of the folding reaction. The same effect was observed for mutants in tendamistat, which were shown to induce a change in the rate-limiting step on a sequential pathway.[17]

Other self-interaction and cross-interaction parameters

For several proteins other parameters like pH, pressure and temperature were varied to characterize the transition barriers for folding. Table 3 shows the self-interaction parameters p_D and p_S and various cross-interaction parameters for ten thoroughly characterized proteins. Only for CI2 are most p_x and p_{xy}-values non-zero and of the same sign. The behavior of barnase is intricate. It shows Hammond behavior in the p_{DS} and p_{DT} cross-interaction parameters and anti Hammond behavior in the p_S self-interaction parameter and the $p_{SS'}$ cross-interaction parameter for helix 1. This may be due to the more complex folding mechanism with transient intermediates,[18] which complicates data analysis and interpretation. For all other proteins there is either no evidence for Hammond behavior or only one of the p_x or p_{xy}-values is different from zero. In addition, none of the p_x or p_{xy} values are negative, demonstrating that there is no evidence for parallel folding pathways.

Discussion

Transition state movements and ground state effects

The analysis of experimental data on 21 proteins shows that apparent transition state movements in protein folding are observed in about half of the proteins in our data set. The results reveal that the observed transition state movements have different origins. CI2 is the only well-characterized protein that shows a small but consistent transition state movement detected by several independent self and cross-interaction parameters (Table 3). ACBP and ADA2h show genuine transition state movements in destabilized mutants ($p_{DS} > 0$), but none of the other self or cross-interaction parameters detect significant changes in the position of the transition state for these proteins. The large majority of apparent transition state movements in proteins destabilized by mutations can be ascribed to ground state effects, i.e. to changes in the structure of the unfolded state or native state. More than half of the proteins in our database show ground state effects upon mutation. In most cases the structure of the unfolded state seems to be altered by small changes in sequence, as indicated by an increase in m_{eq} with decreasing protein stability. This supports the idea of a substantial amount of residual structure in unfolded proteins (Table 2). Mutations introduced for studies of protein folding transition states usually represent a truncation of side-chains, which generally destabilizes intramolecular side-chain interactions. Our findings indicate that this often leads to a more extended unfolded state. NMR studies revealed specific side-chain interactions in denaturant-unfolded states of several proteins,[42–47] including protein G,[48] protein L[49] and spectrin SH3,[50] which show increased m_{eq} values for destabilized mutants in our data analysis. Interactions in unfolded polypeptide chains are often of short or medium range

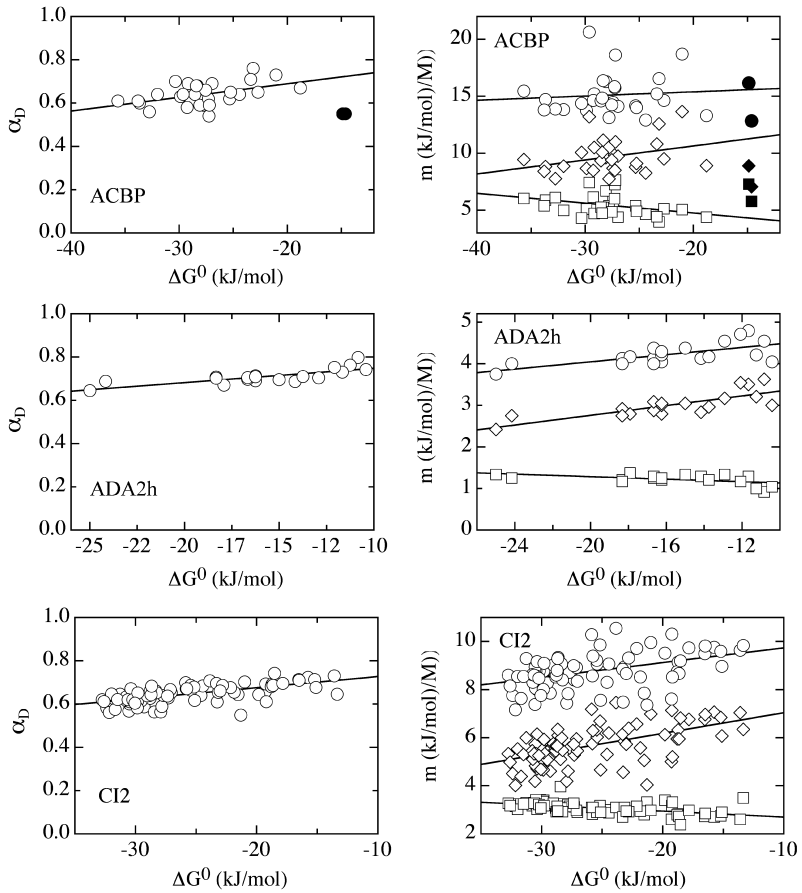

Figure 8. Effect of changes in ΔG^0_S on α_D (left panels) and on m_{eq} (○), m_f (◇) and m_u (□) (right panels) for ACBP, ADA2h and CI2. Linear fits to the data indicate a transition state movement for all three proteins and a ground state effect on the unfolded state in ADA2h and CI2. For two variants of ACBP a second transition state seems to become rate-limiting (●), indicated by the drastic changes in all parameters.

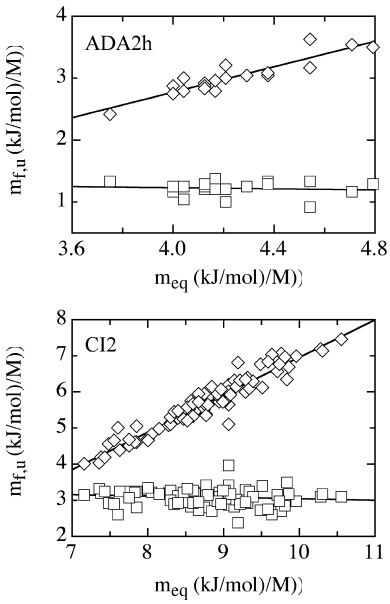

Figure 9. Correlation of the changes in m_{eq} with changes in m_u (◇) and m_f (□) for ADA2h and CI2. Linear fits of the data indicate that nearly all of the variation in m_{eq} is due to changes in m_f. Correlation coefficients between m_f and m_{eq} are 0.91 and 0.95 for ADAh2 and CI2, respectively. The respective correlation coefficients between m_u and m_{eq} are −0.04 and −0.12.

and are formed mainly between a few hydrophobic residues, but also long-range interactions have been described. The drastic disruption of such interactions in the unfolded state of lysozyme by a single point mutation has recently been characterized by NMR.[51] Our results indicate that the destabilization and disruption of these interactions in unfolded polypeptide chains gives rise to significant ground state effects, which are easily mistaken for Hammond behavior if only the effect of a mutation on the α_D-value is considered. The analysis of the p_{DS} cross-interaction parameters shows, however, that for all proteins that exhibit a ground state effect on the unfolded state only the m_f-value is altered while m_u is virtually unchanged (Figures 6–9 and Table 2). This indicates that the structures of the native state and of the transition state are usually rather insensitive to mutations. These results demonstrate the need to analyze all kinetic and equilibrium m-values in order to discriminate between ground state effects and real transition state movements.

The results from the analysis of other self and cross-interaction parameters give further support to the finding that Hammond behavior is not common for protein folding reactions (Table 3). Obviously, structures of protein folding transition states are robust against changes in solvent conditions. The few transition state movements that are observed upon variation of external parameters

Table 3. Proteins for which more than one self or cross-interaction coefficient can be determined

Protein	$p_D = \partial\alpha_D/\partial\Delta G^0_D$	$p_S = \partial\phi_f^{prot}/\partial\Delta G^0_S$	$p_{DS} = \partial\alpha_D/\partial\Delta G^0_S$	$p_{DT} = \partial\alpha_D/\partial\Delta G^0_T$	$p_{DpH} = \partial\alpha_D/\partial\Delta G^0_{pH}$	$p_{Dp} = \partial\alpha_D/\partial\Delta G^0_p$	$p_{DGl} = \partial\alpha_D/\partial\Delta G^0_{Gl}$	$p_{ST} = \partial\phi_f^i/\partial\Delta G^0_T$	$p_{SpH} = \partial\phi_f^i/\partial\Delta G^0_{pH}$	$p_{SS'} = \partial\phi_f^i/\partial\Delta G^0_{S'}$
ACBP[63]	$\cong 0$	$\cong 0$	>0							
ADA2h[69]	$\cong 0$	$\cong 0$	>0							
Barnase[8–12,32]	>0	<0 for helix 1	>0 for core	>0				>0 for helix 1		<0 for helix 1
CI2[10,13,31]	>0 for some mutants	$\cong 0$	>0	>0				>0		
CspB[74]	>0 for some mutants		$\cong 0$							
CTL9[78]	$\cong 0$				$\cong 0$					
FKBP12[71,79]	>0 for some mutants	$\cong 0$	$\cong 0$	$\cong 0$						
NTL9[80,81]	$\cong 0$			$\cong 0$	$\cong 0$					
Protein L[73,82,83]	$\cong 0$	$\cong 0$	$\cong 0$	$\cong 0$			$\cong 0$			
Spectrin SH3[33,76]	>0	$\cong 0$	$\cong 0$		$\cong 0$				$\cong 0$	$\cong 0$
Tendamistat[17,19,59]	>0 for some mutants			$\cong 0$	$\cong 0$	>0				

A p_x or p_{xy}-value is considered to be larger (>0) or smaller (<0) than zero if it differs from zero in more than two standard deviations.

can in most cases be ascribed to changes in the rate-limiting step of the reaction, as shown for the downward curvature in the denaturant-dependence of the folding rates (positive p_D-values).[18]

Parallel folding pathways

Theoretical studies on protein-like model heteropolymers suggested a highly heterogeneous transition state ensemble with a manifold of parallel routes leading to the native state.[52,53] In this case a perturbation should affect the substates of the transition state ensemble to different degrees. The free energy of more native-like transition states would be more strongly increased when the native state is destabilized, which would lead to an apparent movement to more unfolded-like transition states (see Figure 3). This would result in negative self and cross-interaction parameters in our data analysis. The same reasoning applies to a scenario with defined parallel pathways of similar free energy leading to the native state (Figure 3). At least three parallel pathways were described for lysozyme folding. In this case two of the pathways transiently populate an intermediate, which facilitated the detection and characterization of the parallel routes.[54] The existence of two parallel pathways was deduced for folding of a variant of GCN4, a dimeric coiled-coil.[55] Negative p_S and $p_{SS'}$-values indicate parallel pathways for the formation of the first helix of barnase[11,32] and negative $p_{SS'}$-values were also found for protein G.[56] For an immunoglobulin-like domain of titin a negative p_{DS}-value was reported,[57] which also suggests the existence of parallel folding pathways. Our data analysis did not detect any additional negative p_x and p_{xy}-values in any other protein in our data set. The lack of evidence for negative p_x and p_{xy}-values in the majority of proteins could indicate that parallel folding pathways are rare or that they have transition states with similar α_x-values. This would lead to only minor shifts in the apparent position of the transition state that would escape detection in our data analysis.

Shape of free energy barriers in protein folding

In organic reactions the transition state movements probed by different perturbations can be decoupled, i.e. a certain p_x or p_{xy} can have a positive value while others are zero. For these reactions the experimental reaction coordinates x and y describe the formation or breakage of different covalent bonds in a very direct way and thus x and y can change from totally synchronous to totally asynchronous.[4] In protein folding many non-covalent protein–protein, protein–solvent and solvent–solvent interactions are formed and broken during the reaction and many "microscopic" reaction coordinates are averaged into the various "macroscopic" α_x-values, which therefore are likely to be highly correlated. Thus, the different p_x and p_{xy}-values should generally be coupled in protein folding reactions if they are due to global structural changes in the transition state. The results from the p_{DS} analysis (Figures 5 and 6; Table 2) and from the analysis of other self-interaction and cross-interaction parameters (Table 3) indicate the absence of genuine transition state movements for folding of most of the proteins in our data set. Out of the 21 investigated proteins only CI2 shows a small but coherent transition state movement. These results suggest that Hammond behavior is not common in protein folding reactions. Most protein folding transition states seem to be robust, conformationally restricted maxima in the free energy landscape and, similar to native states, they usually do not undergo extensive structural rearrangements upon mutation or changes in solvent conditions.

The discrete jumps in the position of the transition state observed in the p_{DS} analysis for mutants of several proteins (filled symbols in Figures 5 and 8) indicate changes in the rate-limiting step of the folding process. This might be due to parallel folding pathways or to consecutive barriers on a sequential pathway. Sequential pathways for apparent two-state folding are in agreement with our earlier findings from an analysis of proteins with non-zero p_D-values. These results indicated the presence of two or a few consecutive barriers on a sequential pathway with high energy intermediates.[18] The existence of multiple consecutive barriers may further explain transition state movements observed only with some p_x or p_{xy}-values. If two consecutive transition states have a very similar α_x-value in that particular reaction coordinate, the range of experimentally accessible perturbations may lie in a region where the switch between the two transition states occurs. Therefore, conditions where only one of the transition states is limiting may be missed. This will show up as an apparent movement of the transition state.[17,19] Our results indicate that the individual transition states on the sequential pathways are rather robust maxima in most proteins.

Effect of residual structure in the unfolded state on the folding kinetics

Our analysis has revealed that about half of the proteins in our data set have significantly structured unfolded states, which are sensitive to mutation. This raises the question, in which way residual interactions in unfolded proteins influence the rate of protein folding. Figure 10 and Table 4 show that six of the seven proteins that exhibit a ground state effect on the unfolded state (ADAh2, ctACP, CI2, protein G, protein L and Sso7d SH3) show significantly decreased folding rates with increasing m_{eq} and m_f-values (significance of correlation larger than 90%). Interestingly, none of the proteins shows increased folding rates with increasing m_f and m_{eq} values. Obviously, the disruption of residual structure in the unfolded state on average slows down folding, although the

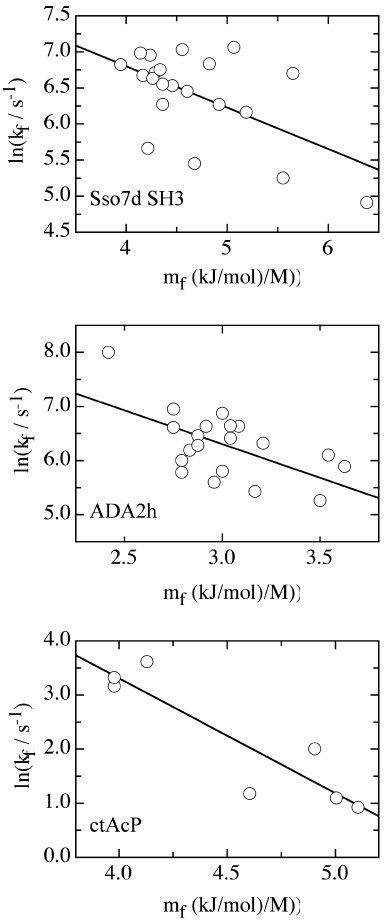

Figure 10. Effect of changes in m_f on ln(k_f) for Sso7d SH3, ADA2h and ctAcP, which exhibit ground state effects on the unfolded state. The correlation coefficients are −0.57, −0.59 and −0.90, respectively (see Table 4).

structure of the transition state is unchanged. This has three important consequences. First, it shows that the majority of the residual interactions in the unfolded state of these proteins are also present in the transition state. The presence of interactions in the unfolded state that are not present in the transition state would result in an energetic cost for the folding reaction, since the unfolded state would be stabilized relative to the transition state. This would lead to increased folding rates when these interactions are destabilized, which is not observed. Our data do not allow us, however, to discriminate between native and non-native interactions. Since ϕ-value analysis showed that transition states commonly contain a subset of the native interactions at reduced strength, our results indicate that the residual interactions in the unfolded state of these six proteins are predominantly native-like. The second consequence of the results is that residual interactions in unfolded proteins can accelerate folding. It has been proposed that the accumulation of native-like structure in the unfolded state should stabilize the unfolded state relative to the transition state and therefore slow down folding.[58] This reasoning assumes, however, that the free energy of the transition state is independent of the strength of a specific interaction in the unfolded state, which can only be assumed for interactions that are not present in the transition state. Changes in stability of an interaction will affect the free energies of the unfolded state and of the transition state, if the interaction is present in both states. The destabilization and disruption of residual interactions obviously has on average a stronger effect on the free energy of the transition state compared to the unfolded state. This indicates that the interactions are stronger in the transition state than in the unfolded state and thus more sensitive to mutations. The third conclusion from these results is that a subset of the interactions present in the transition state for folding can be directly pointed out from structural studies on the unfolded state. NMR studies in combination with *m*-value analysis for each mutation would be required to identify and characterize these interactions in more detail. This can only be done if the changes in the *m*-values for a mutation are large enough, as normally seen for the introduction or disruption of disulfide bonds.[59]

Table 4. Effect of a change in m_f on the refolding kinetics (ln(k_f)) for proteins which show a ground state effect on the unfolded state

Protein	$\partial RT \ln(k_f/s^{-1})/\partial m_f$ (M)	Correlation coefficient	Significance of correlation (%)[a]
ctAcP[67]	−5.3 ± 1.1	−0.90	9.91
ADA2h[69]	−3.1 ± 1.0	−0.59	2.20
CI2[31]	−0.9 ± 0.2	−0.42	0.03
Protein G[72]	−0.9 ± 0.4	−0.40	7.99
Protein L[73]	−1.0 ± 0.3	−0.40	0.27
Spectrin SH3[76]	−3.9 ± 2.9	−0.37	24.5
Sso7d SH3[77]	−1.4 ± 0.5	−0.57	2.23

[a] The significance of the correlation is the probability that the correlation arises by chance. Therefore small values indicate a statistically significant correlation.

The interpretation of rate-equilibrium free energy relationships in protein folding

REFERs are empirical tools that cannot be deduced from the laws of thermodynamics and have thus been termed extra-thermodynamic free energy relationships.[2] A REFER is a macroscopic energetic reaction coordinate averaged over all the microscopic modes of the protein/solvent system. This raises two issues. The first one is whether a relationship can be found that links the different α_x to some microscopic reaction coordinate, such as dihedral angles or a solvation-related parameter. Such a relationship would enable a direct comparison between simulations on protein-like models and experimental results. However, the choice of a

microscopic reaction coordinate is not obvious in itself[60,61] and the many microscopic coordinates of a multidimensional system may give rise to complex behavior.[62] The second issue is the difficulty in the molecular interpretation of α_x-values in protein folding. α_x is a normalized energetic parameter and the molecular interpretation of the α_x-value is difficult due to the large variety of different effects on the protein and on the solvent. For α_D or α_C (equations (4) and (7)), empirical correlations between equilibrium free energy relationships and changes in solvent-accessible surface areas have been found,[27] which were also applied to the corresponding kinetic properties. These data thus give information on the solvent accessibility of the transition state, as long as the structure of the ground states is not changed upon variation of denaturant concentration (for α_D) or temperature (for α_C), which seems to be the case for many proteins. Our results show that structural changes are more difficult to interpret in the case of mutational studies, which were previously assumed to affect only interactions in the native state and in the transition state. However, mutations can significantly alter the structure of the unfolded state, as demonstrated by NMR and by our p_{DS}-value analysis (Figures 6–9 and Table 2). In this case a structural interpretation of the ϕ_f^i-value is not warranted any more and ϕ_f^i can only be interpreted in energetic terms.

Conclusions

The analysis of transition state movements in 21 well-characterized proteins using self-interaction and cross-interaction parameters shows that Hammond behavior is rare in protein folding reactions. This indicates that folding transition states are narrow regions on the free energy landscape that are robust against perturbations and supports our earlier finding that the commonly observed non-zero p_D-values are rather due to a change in the rate-limiting step on a sequential folding pathway and are not caused by Hammond behavior. Apparent transition state movements detected by p_{DS} cross-interaction parameters are in most proteins due to ground state effects, most commonly due to structural changes in the unfolded state. Our results show that residual structure is common in unfolded proteins and that it can accelerate protein folding. The use of cross-interaction parameters provides a sensitive tool to detect such structures.

Methods

The data were taken from the references indicated in the Tables. All rate and equilibrium constants and *m*-values were extrapolated to water. Only mutants for which $\Delta G^0 \leq -10$ kJ/mol are considered. The noise is higher in the data for protein variants with lower stability (not shown). The cause is probably a shorter refolding limb of the Chevron plot in the case of kinetic measurements and a shorter native state baseline in the case of equilibrium measurements. An exception was made for the drkN SH3 domain, which is unique in that nine variants at a single position of the protein were characterized and for which the quality of the data is very high. In this case $\Delta G^0 \leq -8$ kJ/mol was used as a cutoff. The software ProFit (QuantumSoft, Zürich, Switzerland) was used for data fitting and for the statistical analysis of the results.

Acknowledgements

We thank Annett Bachmann for discussion and comments on the manuscript. This work was supported by a grant from the Schweizerischen Nationalfonds.

References

1. Leffler, J. E. (1953). Parameters for the description of transition states. *Science*, **117**, 340–341.
2. Leffler, J. E. & Grunwald, E. (1963). *Rates and Equilibria of Organic Reactions*, Dover, New York.
3. Hammond, G. S. (1955). A correlation of reaction rates. *J. Am. Chem. Soc.* **77**, 334–338.
4. Jencks, W. P. (1985). A primer for the Bema Hapothle. An empirical approach to the characterization of changing transition-state structures. *Chem. Rev.* **85**, 511–527.
5. Farcasiu, D. (1975). The use and misuse of the Hammond postulate. *J. Chem. Ed.* **52**, 76–79.
6. Jencks, W. P. (1969). *Catalysis in Chemistry and Enzymology*, McGraw-Hill, New York.
7. Sánchez, I. E. & Kiefhaber, T. (2003). Non-linear rate-equilibrium free energy relationships and Hammond behavior in protein folding. *Biophys. Chem.* **100**, 397–407.
8. Matouschek, A. & Fersht, A. R. (1993). Application of physical organic chemistry to engineered mutants of proteins: Hammond postulate behavior in the transition state of protein folding. *Proc. Natl Acad. Sci. USA*, **90**, 7814–7818.
9. Matouschek, A., Matthews, J. M., Johnson, C. M. & Fersht, A. R. (1994). Extrapolation to water of kinetic and equilibrium data for the unfolding of barnase in urea solutions. *Protein Eng.* **7**, 1089–1095.
10. Matouschek, A., Otzen, D. E., Itzhaki, L., Jackson, S. E. & Fersht, A. R. (1995). Movement of the transition state in protein folding. *Biochemistry*, **34**, 13656–13662.
11. Matthews, J. M. & Fersht, A. R. (1995). Exploring the energy surface of protein folding by structure–reactivity relationship and engineered proteins: observation of Hammond behavior for the gross structure of the transitions state and anti-Hammond behavior for structural elements for unfolding/folding of barnase. *Biochemistry*, **34**, 6805–6814.
12. Dalby, P. A., Oliveberg, M. & Fersht, A. R. (1998). Movement of the intermediate and rate determining transition state of barnase on the energy landscape with changing temperature. *Biochemistry*, **37**, 4674–4679.

13. Oliveberg, M., Tan, Y.-J., Silow, M. & Fersht, A. R. (1998). The changing nature of the protein folding transition state: implications for the free-energy profile for folding. *J. Mol. Biol.* **277**, 933–943.
14. Silow, M. & Oliveberg, M. (1997). High-energy channeling in protein folding. *Biochemistry*, **36**, 7633–7637.
15. Walkenhorst, W. F., Green, S. & Roder, H. (1997). Kinetic evidence for folding and unfolding intermediates in staphylococcal nuclease. *Biochemistry*, **36**, 5795–5805.
16. Kiefhaber, T., Bachmann, A., Wildegger, G. & Wagner, C. (1997). Direct measurements of nucleation and growth rates in lysozyme folding. *Biochemistry*, **36**, 5108–5112.
17. Bachmann, A. & Kiefhaber, T. (2001). Apparent two-state tendamistat folding is a sequential process along a defined route. *J. Mol. Biol.* **306**, 375–386.
18. Sánchez, I. E. & Kiefhaber, T. (2002). Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* **325**, 367–376.
19. Pappenberger, G., Saudan, C., Becker, M., Merbach, A. E. & Kiefhaber, T. (2000). Denaturant-induced movement of the transition state of protein folding revealed by high pressure stopped-flow measurements. *Proc. Natl Acad. Sci. USA*, **97**, 17–22.
20. Jäger, M., Nguyen, H., Crane, J. C., Kelly, J. W. & Gruebele, M. (2001). The folding mechanism of a beta-sheet: the WW domain. *J. Mol. Biol.* **311**, 373–393.
21. Pohl, F. M. (1976). Temperature-dependence of the kinetics of folding of chymotrypsinogen A. *FEBS Letters*, **65**, 293–296.
22. Chen, B. L., Baase, W. A. & Schellman, J. A. (1989). Low-temperature unfolding of a mutant of phage T4 lysozyme. 2. Kinetic investigations. *Biochemistry*, **28**, 691–699.
23. Privalov, P. L. & Makhatadze, G. I. (1992). Contribution of hydration and non-covalent interactions to the heat capacity effect on protein unfolding. *J. Mol. Biol.* **224**, 715–723.
24. Greene, R. F. J. & Pace, C. N. (1974). Urea and guanidine-hydrochloride denaturation of ribonuclease, lysozyme, alpha-chyomtrypsinn and beta-lactoglobulin. *J. Biol. Chem.* **249**, 5388–5393.
25. Santoro, M. M. & Bolen, D. W. (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry*, **27**, 8063–8068.
26. Tanford, C. (1970). Protein denaturation. Part C. Theoretical models for the mechanism of denaturation. *Advan. Protein Chem.* **24**, 1–95.
27. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148.
28. Matthews, C. R. (1987). Effect of point mutations on the folding of globular proteins. *Methods Enzymol.* **154**, 498–511.
29. Goldenberg, D. P., Frieden, R. W., Haack, J. A. & Morrison, T. B. (1989). Mutational analysis of a protein-folding pathway. *Nature*, **338**, 127–132.
30. Fersht, A. R., Matouschek, A. & Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224**, 771–782.
31. Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995). The structure of the transition state for folding of chymotrypsin inhibitor 2 analyzed by protein engineering methods: evidence for a nucleation-condesation mechanism for protein folding. *J. Mol. Biol.* **254**, 260–288.
32. Fersht, A. R., Itzhaki, L. S., elMasry, N. F., Matthews, J. M. & Otzen, D. E. (1994). Single *versus* parallel pathways of protein folding and fractional formation of structure in the transition state. *Proc. Natl Acad. Sci. USA*, **91**, 10426–10429.
33. Martinez, J. C., Pisabarro, M. T. & Serrano, L. (1998). Obligatory steps in protein folding and the conformational diversity of the transition state. *Nature Struct. Biol.* **5**, 721–729.
34. Dalby, P. A., Oliveberg, M. & Fersht, A. R. (1998). Folding of wild-type and mutants of barnase. I. Use of ϕ analysis and m-values to probe the cooperative nature of the folding pre-equilibrium. *J. Mol. Biol.* **276**, 625–646.
35. Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold. Des.* **3**, R81–R91.
36. Grantcharova, V. P. & Baker, D. (1997). Folding dynamics of the src SH3 domain. *Biochemistry*, **36**, 15685–15692.
37. Mok, Y. K., Elisseeva, E. L., Davidson, A. R. & Forman-Kay, J. D. (2001). Dramatic stabilization of an SH3 domain by a single substitution: roles of the folded and unfolded states. *J. Mol. Biol.* **307**, 913–928.
38. Northey, J. G., Maxwell, K. L. & Davidson, A. R. (2002). Protein folding kinetics beyond the phi value: using multiple amino acid substitutions to investigate the structure of the SH3 domain folding transition state. *J. Mol. Biol.* **320**, 389–402.
39. Lopez-Hernandez, E. & Serrano, L. (1995). Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, CI-2. *Fold. Des.* **1**, 43–55.
40. Carra, J. H., Anderson, E. A. & Privalov, P. L. (1994). Three-state thermodynamic analysis of the denaturation of staphylococcal nuclease mutants. *Biochemistry*, **33**, 10842–10850.
41. Garcia, P., Serrano, L., Rico, M. & Bruix, M. (2002). An NMR view of the folding process of a CheY mutant at the residue level. *Structure (Camb.)*, **10**, 1173–1185.
42. Evans, P. A., Topping, K. D., Woolfson, D. N. & Dobson, C. M. (1991). Hydrophobic clustering in nonnative states of a protein: interpretation of chemical shifts in NMR spectra of denatured states of lysozyme. *Proteins: Struct. Funct. Genet.* **9**, 248–266.
43. Neri, D., Billeter, M., Wider, G. & Wüthrich, K. (1992). NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science*, **257**, 1559–1563.
44. Logan, T. M., Theriault, Y. & Fesik, S. W. (1994). Structural characterization of the FK506 binding protein unfolded in urea and guanidine hydrochloride. *J. Mol. Biol.* **236**, 637–648.
45. Wong, K. B., Clarke, J., Bond, C. J., Neira, J. L., Freund, S. M., Fersht, A. R. & Daggett, V. (2000). Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. *J. Mol. Biol.* **296**, 1257–1282.
46. Kazmirski, S. L., Wong, K. B., Freund, S. M., Tan, Y. J., Fersht, A. R. & Daggett, V. (2001). Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proc. Natl Acad. Sci. USA*, **98**, 4349–4354.

47. Garcia, P., Serrano, L., Durand, D., Rico, M. & Bruix, M. (2001). NMR and SAXS characterization of the denatured state of the chemotactic protein CheY: implications for protein folding initiation. *Protein Sci.* **10**, 1100–1112.
48. Sari, N., Alexander, P., Bryan, P. N. & Orban, J. (2000). Structure and dynamics of an acid-denatured protein G mutant. *Biochemistry*, **39**, 965–977.
49. Yi, Q., Scalley-Kim, M. L., Alm, E. J. & Baker, D. (2000). NMR characterization of residual structure in the denatured state of protein L. *J. Mol. Biol.* **299**, 1341–1351.
50. Kortemme, T., Kelly, M. J., Kay, L. E., Forman-Kay, J. D. & Serrano, L. (2000). Similarities between the spectrin SH3 domain denatured state and its folding transition state. *J. Mol. Biol.* **297**, 1217–1229.
51. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T. *et al.* (2002). Long-range interactions within a nonnative protein. *Science*, **295**, 1719–1722.
52. Guo, Z. & Thirumalai, D. (1995). Kinetics of protein folding: nucleation mechanism, time scales, and pathways. *Biopolymers*, **36**, 83–102.
53. Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. (1995). Navigating the folding routes. *Science*, **267**, 1619–1620.
54. Bieri, O., Wildegger, G., Bachmann, A., Wagner, C. & Kiefhaber, T. (1999). A salt-induced intermediate is on a new parallel pathway of lysozyme folding. *Biochemistry*, **38**, 12460–12470.
55. Krantz, B. A. & Sosnick, T. R. (2001). Engineered metal binding sites map the heterogeneous folding landscape of a coiled coil. *Nature Struct. Biol.* **8**, 1042–1047.
56. Nauli, S., Kuhlman, B. & Baker, D. (2001). Computer-based redesign of a protein folding pathway. *Nature Struct. Biol.* **8**, 602–605.
57. Fowler, S. B. & Clarke, J. (2001). Mapping the folding pathway of an immunoglobulin domain: structural detail from Phi value analysis and movement of the transition state. *Structure (Camb.)*, **9**, 355–366.
58. Fersht, A. R. (1995). Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc. Natl Acad. Sci. USA*, **92**, 10869–10873.
59. Schönbrunner, N., Pappenberger, G., Scharf, M., Engels, J. & Kiefhaber, T. (1997). Effect of pre-formed correct tertiary interactions on rapid two-state tendamistat folding: evidence for hairpins as initiation sites for β-sheet formation. *Biochemistry*, **36**, 9057–9065.
60. Du, R., Pande, V. S., Grosberg, A. Y., Tanaka, T. & Shaknovich, E. I. (1998). On the transition coordinate for protein folding. *J. Chem. Phys.* **108**, 334–350.
61. Bolhuis, P. G., Dellago, C. & Chandler, D. (2000). Reaction coordinates of biomolecular isomerization. *Proc. Natl Acad. Sci. USA*, **97**, 5877–5882.
62. Wales, D. J. (2001). A microscopic basis for the global appearance of energy landscapes. *Science*, **293**, 2067–2070.
63. Kragelund, B. B., Osmark, P., Neergaard, T. B., Schiodt, J., Kristiansen, K., Knudsen, J. & Poulsen, F. M. (1999). The formation of a native-like structure containing eight conserved hydrophobic residues is rate limiting in two-state protein folding of ACBP. *Nature Struct. Biol.* **6**, 594–601.
64. Chu, R., Pei, W., Takei, J. & Bai, Y. (2002). Relationship between the native-state hydrogen exchange and folding pathways of a four-helix bundle protein. *Biochemistry*, **41**, 7998–8003.
65. Capaldi, A. P., Kleanthous, C. & Radford, S. E. (2002). Im7 folding mechanism: misfolding on a path to the native state. *Nature Struct. Biol.* **9**, 209–216.
66. Myers, J. K. & Oas, T. G. (1999). Contribution of a buried hydrogen bond to lambda repressor folding kinetics. *Biochemistry*, **38**, 6761–6768.
67. Taddei, N., Chiti, F., Fiaschi, T., Bucciantini, M., Capanni, C., Stefani, M. *et al.* (2000). Stabilisation of alpha-helices by site-directed mutagenesis reveals the importance of secondary structure in the transition state for acylphosphatase folding. *J. Mol. Biol.* **300**, 633–647.
68. Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M. & Dobson, C. M. (1999). Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nature Struct. Biol.* **6**, 1005–1009.
69. Villegas, V., Martinez, J. C., Avilés, F. X. & Serrano, L. (1998). Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain. *J. Mol. Biol.* **283**, 1027–1036.
70. Serrano, L., Matouschek, A. & Fersht, A. R. (1992). The folding of an enzyme. III. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. *J. Mol. Biol.* **224**, 805–818.
71. Fulton, K. F., Main, E. R., Daggett, V. & Jackson, S. E. (1999). Mapping the interactions present in the transition state for unfolding/folding of FKBP12. *J. Mol. Biol.* **291**, 445–461.
72. McCallister, E. L., Alm, E. & Baker, D. (2000). Critical role of beta-hairpin formation in protein G folding. *Nature Struct. Biol.* **7**, 669–673.
73. Kim, D. E., Fisher, C. & Baker, D. (2000). A breakdown of symmetry in the folding transition state of protein L. *J. Mol. Biol.* **298**, 971–984.
74. Perl, D., Holtermann, G. & Schmid, F. X. (2001). Role of the chain termini for the folding transition state of the cold shock protein. *Biochemistry*, **40**, 15501–15511.
75. Northey, J. G., Di Nardo, A. A. & Davidson, A. R. (2002). Hydrophobic core packing in the SH3 domain folding transition state. *Nature Struct. Biol.* **9**, 126–130.
76. Martinez, J. C. & Serrano, L. (1999). The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nature Struct. Biol.* **6**, 1010–1016.
77. Guerois, R. & Serrano, L. (2000). The SH3-fold family: experimental evidence and prediction of variations in the folding pathways. *J. Mol. Biol.* **304**, 967–982.
78. Sato, S. & Raleigh, D. P. (2002). pH-dependent stability and folding kinetics of a protein with an unusual alpha-beta topology: the C-terminal domain of the ribosomal protein L9. *J. Mol. Biol.* **318**, 571–582.
79. Main, E. R. G., Fulton, K. F. & Jackson, S. E. (1999). Folding pathway of FKBP12 and characterization of the transition state. *J. Mol. Biol.* **291**, 429–444.
80. Kuhlman, B., Luisi, D. L., Evans, P. A. & Raleigh, D. P. (1998). Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the N-terminal domain of the protein L9. *J. Mol. Biol.* **284**, 1661–1670.
81. Luisi, D. L. & Raleigh, D. P. (2000). pH-dependent interactions and the stability and folding kinetics

of the N-terminal domain of L9. Electrostatic interactions are only weakly formed in the transition state for folding. *J. Mol. Biol.* **299**, 1091–1100.
82. Scalley, M. L. & Baker, D. (1997). Protein folding kinetics exhibit an Arrhenius temperature dependence when corrected for the temperature dependence of protein stability. *Proc. Natl Acad. Sci. USA*, **94**, 10636–10640.
83. Plaxco, K. W. & Baker, D. (1998). Limited internal friction in the rate-limiting step of a two-state protein folding reaction. *Proc. Natl Acad. Sci. USA*, **95**, 13591–13596.

Edited by R. Huber

(Received 14 November 2002; received in revised form 21 January 2003; accepted 23 January 2003)

Note added in proof: Recently an additional mutational study on Im9 protein was published (Friel, C. T., Capaldi, A. P. & Radford, S. E. (2003). *J. Mol. Biol.* **326**, 293–305) which supports the conclusions of our present paper. Our analysis of the data by Friel *et al.* reveals that also Im9 shows a ground state effect on the unfolded state without transition state movement. The corresponding parameters are (cf. Table 2): $p_{DS} \cong 0$, $\partial m_{eq}/\partial \Delta G_S^0 = 1.9 \pm 0.4$ $(M^{-1} \times 10^2)$, $\partial m_f/\partial \Delta G_S^0 = 2.0 \pm 0.2$ $(M^{-1} \times 10^2)$, $\partial m_u/\partial \Delta G_S^0 = -0.1 \pm 0.3$ $(M^{-1} \times 10^2)$, $(\partial m_f/\partial \Delta G_S^0)/(\partial m_{eq}/\partial \Delta G_S^0) = 0.97 \pm 0.21$, $(\partial m_u/\partial \Delta G_S^0)/(\partial m_{eq}/\partial \Delta G_S^0) = -0.03 \pm 0.13$, $\partial m_f/\partial m_{eq} = 0.77 \pm 0.09$. In addition there is a significant decrease in the folding rate constant (k_f) with increasing m_f-value with $(\partial RT\ln(k_f/s^{-1}))/(\partial m_f) = -13.9(\pm 2.3)$ M, a correlation coefficient of -0.77 and a significance of correlation of 0.08% (cf. Table 4).