THE UNIVERSITY OF CHICAGO

EARLY EVENTS AND BARRIERS IN FAST PROTEIN FOLDING

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

BY

WILLIAM KEVIN MEISNER

CHICAGO, ILLINOIS

AUGUST 2004

TABLE OF CONTENTS

Page

LIST OF FIGURES..... v
LIST OF ABBREVIATIONS.....vi
ACKNOWLEDGEMENTS.....vii
ABSTRACT..... viii

Chapter

1.0 INTRODUCTION TO PROTEIN FOLDING.....1

1.1 Beginnings 1
1.2 Two-state folding 5
1.3 Early events in protein folding 10
1.4 Kinetic burst phases 14
1.5 Reaction rate theories 19
1.6 Diffusion-collision 21
1.7 Diffusion-collision and experiments 23
1.8 Landscape Theory 25
1.9 Downhill folding 28
1.10 Downhill folding experiments 30
1.11 Native state hydrogen exchange 32

2.0 FAST FOLDING OF A HELICAL PROTEIN INITIATED BY THE COLLISION OF UNSTRUCTURED CHAINS.....36

2.1 Abstract 37
2.2 Introduction 37
2.3 Materials and methods 41
2.4 Results and discussion 42
2.4.1 Design of a stable coiled coil with negligible intrinsic helicity 42
2.4.2 Two-state folding kinetics 45
2.4.3 Folding is faster than predicted by D-C Model 49
2.4.4 GCN4-E_9G_4 Folding pathway 51
2.4.5 The role of helicity in folding 54

3.0 INVESTIGATING BARRIER-FREE PROTEIN FOLDING USING NATIVE-STATE HYDROGEN EXCHANGE METHODS.....56

3.1 Abstract 57
3.2 Introduction 58
3.3 Materials and methods 62
3.4 Results and discussion 65

3.4.1 Standard equilibrium and kinetic measurements 65
3.4.2 HX measurements 68
3.4.3 Folding remains barrier-limited 74
3.4.4 Denaturant does not alter folding behavior 75
3.4.5 Cooperativity and downhill folding 75
3.4.6 Early folding steps are uphill 76
3.5 Conclusions 78

4.0 CONCLUSIONS AND FUTURE DIRECTIONS.....79

4.1 Summary of conclusions 79
4.1.1 Collision between unstructured chains can induce folding 79
4.1.2 The μsec folding of a stable protein is barrier-limited under native conditions 81
4.2 Future directions 82
4.2.1 Success frequency of collisions 82
4.2.2 Size vs. orientation in the folding transition state 83
4.2.3 Design of a downhill folder 84

REFERENCES.....86

Appendix

CHEVRON ANALYSIS.....96

LIST OF FIGURES

Figure		Page
1.1	Two-state and three-state folding	6
1.2	Kinetic "burst phases"	15
1.3	Folding free energy surfaces	26
1 4	Native state hydrogen exchange	33
2.1	Acceleration of folding rates through helix stabilization does not identify whether helix formation comes before or after collision	39
2.2	Design properties of stable coiled coil with negligible intrinsic helicity	43
2.3	Folding behavior of the designed coiled coil GCN4-E_9G_4	46
2.4	Folding pathway of a designed coiled coil	52
3.1	Barrier-limited versus downhill folding	59
3.2	Equilibrium and kinetic measurements	66
3.3	Native state hydrogen exchange	70
3.4	μsec folding is barrier-limited	72
A.1	Chevron analysis	99

LIST OF ABBREVIATIONS

BPTI	bovine pancreatic trypsin inhibitor
CD	far-UV circular dichroism
CI2	chymotrypsin inhibitor II
ctAcP	common type acyl phosphatase
Cyt c	cytochrome c
D-C model	diffusion-collision model
FRET	fluorescence resonance energy transfer
gdmHCl	guanidinium hydrochloride
HEWL	hen egg white lysozyme
IM7	colicin immunity protein 7
PPII	polyproline II
ProtG	protein G
ProtL	protein L
RNase	ribonuclease
SAXS	small angle X-ray scattering
TFE	trifluoroethanol
Ub	ubiquitin

ACKNOWLEDGEMENTS

Whatever successes I've enjoyed while pursuing my doctorate degree at the University of Chicago are due in large part to the contributions of many others. First and foremost, my thesis advisor Tobin Sosnick, whose enthusiasm and expertise fosters a stimulating work environment certain to produce and attract outstanding scientists. Notable among them is Bryan Krantz, a graduate student one year in advance of myself, whose practical knowledge of the lab kept the daily experimental machinery running as smoothly as possible. I am equally indebted to Josh Kurutz and Shohei Koide, for their continuing efforts to shape and improve the NMR facility, and to Ron Crouch of Varian, Inc. for additional NMR instruction and advice. I've also benefited from numerous thought provoking discussions and criticisms provided by my thesis committee: Wouter Hoff (chair), Marvin Makinen, and Keith Moffat. I thank the chairman of the Department of Biochemistry and Molecular Biology, Tony Kossiakoff, and the office staff for keeping the "behind the scenes" workings of the department behind the scenes. Lastly, I thank my family, and friends too numerous to mention for their encouragement and support during my time in Chicago.

ABSTRACT

The current thesis investigates the ultra-fast folding behavior of two coiled coil variants. Chapter two addresses whether helix formation or chain collision is the first step of folding. The relative positions of these events leading up to the free energy barrier was determined using a dimeric version of a coiled coil with such low intrinsic helicity that far fewer than 1% of collisions occur between helical, rather than unstructured chains. Folding of this protein approaches the diffusion-limited rate and is much faster than possible if folding is contingent upon pre-collision helix formation. Thus, the collision of two unstructured chains is the initial step of the dominant kinetic pathway, while helicity exerts its influence only at a later step. Folding from an unstructured encounter complex may be efficient and robust, which has implications for any biological process that couples folding to binding.

In chapter three, a second coiled coil is investigated with results pertinent to the identification of the elusive downhill folding behavior predicted by theory. Because this protein is among the fastest and most stable yet described, it is a prime candidate for downhill folding. A combination of traditional stop-flow experiments and native state hydrogen exchange methods demonstrate that the kinetic and thermodynamic characteristics of folding in the absence of denaturant are consistent with those at the folding transition, where behavior is known to be barrier limited. Therefore, folding of this molecule remains cooperative and barrier limited even under conditions where downhill folding is predicted to occur.

CHAPTER 1

INTRODUCTION TO PROTEIN FOLDING

1.1 Beginnings

As a unique field of biological investigation, protein folding was birthed from two observations in the 1960's. First, Anfinsen demonstrated that fully denatured proteins require nothing more than dilution of denaturant to return to their active form (Anfinsen 1973). This observation identified each protein's unique amino acid sequence as sufficient for the determination of its structure. Before long, Levinthal and others noted that a simple random search through the absurdly large number of unfolded configurations could not locate the folded structure in a biologically relevant time (Levinthal 1968).

So how do proteins circumvent what has come to be called Levinthal's paradox? The obvious answer is that the search is not simply random, but directed.

Subsequent folding studies attempted to identify the events, factors, and mechanisms responsible for guidance. One proposed solution that was popular through most of the first two decades of study is the formation of a step-wise series of obligate intermediates along a defined pathway. Formation of each intermediate pays the great majority of the remaining entropic penalty, enabling subsequent structural changes to fine tune the native structure through smaller interactions at little additional cost. A consequence of this understanding was a vision of the folding reaction as a progression through a series of well defined steps to the native structure, in a manner quite similar to a metabolic pathway.

Creighton's investigation of disulfide bond formation during folding of bovine pancreatic trypsin inhibitor (BPTI) is among the most detailed of this genre (Creighton 1978; Creighton 1985). His approach utilized thiol blocking reagents to trap free reactive thiol groups, allowing him to identify four early intermediates containing just one of the protein's four possible disulfide bonds, and four two-disulfide intermediates. Approximately 30% of the early intermediates were mispaired – that is, the disulfide bond formed between residues unpaired in the native conformation – and a small amount of the two-disulfide intermediates contained one native and one non-native pair. This led to the perhaps surprising hypothesis that those intermediates containing two native pairs were non-productive, and that the remaining two-disulfide species were obligate non-native conformations encountered during a successful folding event.

Creighton's pioneering work has fallen out of favor for a number of reasons. Weisman and Kim showed that the integrity of the blocking reaction was suspect (Weissman and Kim 1991), because the charged or highly polar blocking reagents Creighton employed were unlikely to penetrate the hydrophobic core and therefore preferentially protected surface residues, leading to non-native disulphide formation. Furthermore, disulphide bonds are absent from a great number of proteins, and even when present, protonation and potential burial of free thiol groups at physiological pH renders them largely non-reactive. Thus, while an understanding of disulphide bond formation may be important in special cases, its utility for protein folding in general is limited.

Ensuing years witnessed the discovery of many proteins that fold in a single step, with no accumulation of intermediate structures. In addition, much of the evidence for intermediates in non-disulfide containing proteins was found to be a consequence of non-native isomers of proline in the unfolded state, or misfolding. Ribonuclease A (RNase A) is an extreme case of the former (Brandts, Halvorson and Brennan 1975), while cytochrome *c* (Cyt c) usefully illustrates the latter (Sosnick et al. 1994).

Unfolded RNase A refolds with multi-phasic kinetics, an observation that could be considered indicative of a folding intermediate. Baldwin and co-workers proposed that fast and slow populations of unfolded RNase A slowly interconvert (Garel and Baldwin 1975), a behavior attributed to proline isomerization by Brandts *et al.* Prolines at two sites in native RNase A adopt the cis-conformation, an

orientation sufficiently unfavorable that it exists in only 20% of the unfolded molecules. This fraction folds quickly and directly to the native state, but the remaining 80% fold more slowly because they must undergo a slow isomerization at these two positions before folding. The slow fraction also folds through a stable intermediate, further illustrating that observable intermediates are not necessarily beneficial to folding (Udgaonkar and Baldwin 1995).

Sosnick, Englander, and coworkers attributed the biphasic kinetics of Cyt c folding at neutral to basic pH to the misligation of its heme cofactor (Sosnick et al. 1994). The native ligation occurs with Met80 but during folding, a population of molecules with the heme misligated to His26 or His33 transiently accumulates. However, reducing solvent pH to 4.8 protonates the heme and largely eliminates the misligation which slowed folding. For the major fraction, folding under these conditions follows a single exponential, and is consistent kinetically with a two-state system. This work demonstrated that while intermediates most certainly exist, they do not have to accumulate. They suggested that intermediates generally accumulate due to misfolding. Furthermore, it was proposed that folding is limited by a global search for a native-like nucleus (Sosnick et al. 1995; Sosnick, Mayne and Englander 1996).

The growing evidence that most small single-domain proteins can fold without accumulating intermediates led to a new model: that protein folding can be well described as a two-state, *unfolded* $\leftrightarrow$ *folded* transition (also notated as *denatured* $\leftrightarrow$ *native*, or combinations thereof) where only the denatured and native conformations are present in observable amounts. Perhaps surprisingly, this idea

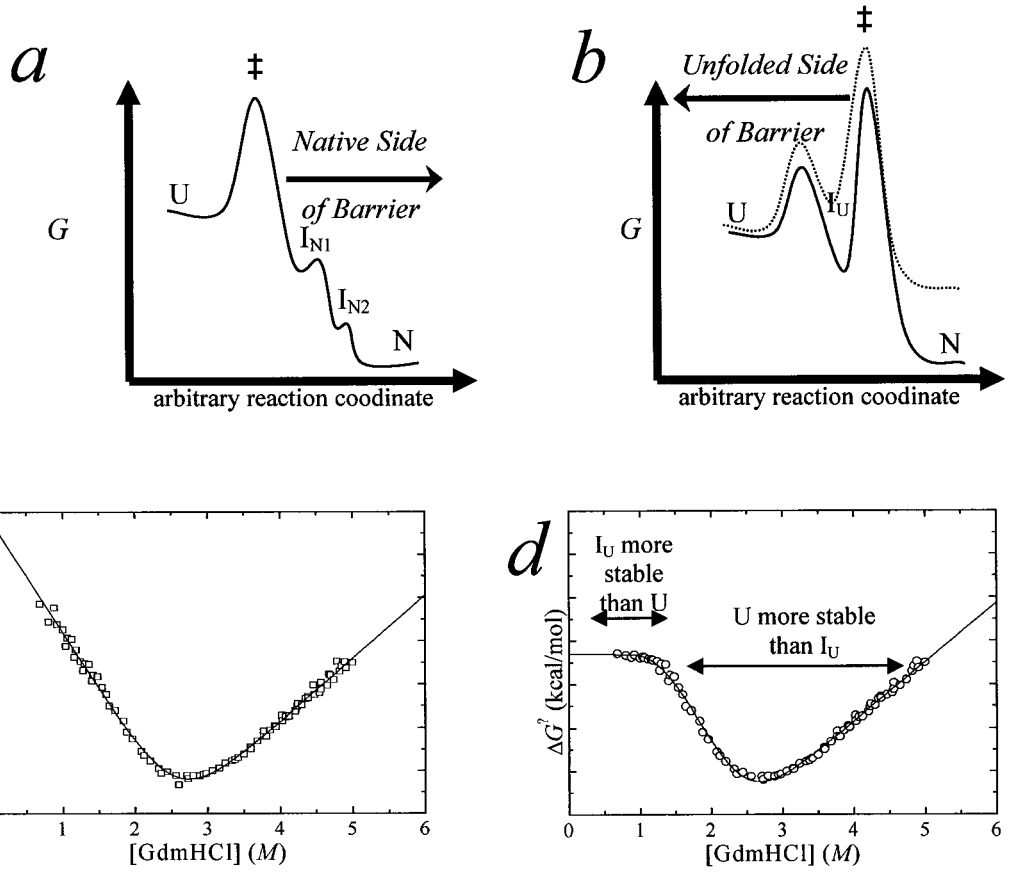
predates Levinthal's insight, having been proposed by Lumry, Biltonen and Brandts in 1966 (Lumry, Biltonen and Brandts 1966), but was largely ignored for nearly three decades while intermediates were characterized in BPTI and RNase A.

Adoption of this paradigm shifted focus from the search for a linear sequence of obligate intermediates to characterization of the ensemble of structures present at the rate-limiting step (though reference to this ensemble as a "transition state" may recall the earlier view's analogy to reactions in metabolic chemistry). It was believed by many that an adequate understanding of the interactions present in the transition state would identify key parts of the protein folding problem, and from a purely practical standpoint, it was the only intermediary position amenable to characterization. These interactions can be probed by changes in reaction kinetics due to mutation through a formalism referred to as ϕ-analysis, mentioned briefly here for its historical significance.

1.2 Two-state folding

Conceptually, two-state behavior (Fig 1.1, ***a*** and ***b***) and single exponential kinetics require that all conformational changes within the unfolded and native populations equilibrate rapidly relative to the folding and unfolding processes. From an experimental standpoint, four criteria are considered requisite for two-state folding of a protein. As discussed earlier, the relaxation kinetics must be simple exponential. Fits by a sum of exponentials can substitute, provided it can be shown due to heterogeneity within the unfolded state caused by, for example, proline

Figure 1.1 *Two-state and three-state folding.* Idealized free energy diagrams and kinetic behavior of two-state *(left)* and three-state *(right)* folding. ***(a)*** Two-state folding requires that all intermediates be either of higher free energy than the unfolded state, or exist only on the native side of the major free energy barrier to folding (I_{N1}, I_{N2}). ***(b)*** Denaturants exert their influence by interacting with the additional exposed surface area of unfolded, or partially unfolded, proteins. Higher denaturant concentrations preferentially stabilize the unfolded state relative to an intermediate because the unfolded protein has the most exposed surface area. Thus, in three-state folding, an intermediate prior to the energy barrier, I_U, may accumulate in the absence of denaturant (solid line), but be unobservable at higher denaturant concentrations (dotted line). ***(c)*** For two-state folding, the denaturant dependence of the free energies are linear to the origin, whereas three-state systems ***(d)*** display "roll-over" at low denaturant because I_U becomes more stable than U.


a
‡
U
Native Side
of Barrier
G
I_{N1}
I_{N2}
N
arbitrary reaction coodinate
b
‡
Unfolded Side
of Barrier
U
I_U
G
N
arbitrary reaction coodinate
c
$\Delta G^?$ (kcal/mol)
0
1
2
3
4
5
6
[GdmHCl] (M)
d
I_U more stable than U
U more stable than I_U
$\Delta G^?$ (kcal/mol)
0
1
2
3
4
5
6
[GdmHCl] (M)

isomerization. In this case, individual unfolded molecules group into two or more populations, each with a characteristic folding rate.

A consequence of the inability of intermediates to accumulate is that all regions of the protein fold in the same kinetic step. This property is referred to as cooperativity, and is best confirmed by monitoring multiple probes of protein structure. Coincidence between the equilibrium signals of a global parameter such as secondary structure acquisition – monitored by far-UV circular dichroism (CD) – with a local environmental probe – frequently tryptophan fluorescence – is used as strong evidence of a molecule's cooperative behavior, though changes in heat capacity, fluorescence resonance energy transfer (FRET) and other properties are also sometimes used. Kinetic evidence of two-state behavior dominated by a single energy barrier can be found in the denaturant dependence of folding kinetics, termed "chevron analysis" (Fig 1.1, ***c*** and ***d***) (Matthews 1987). This method was first developed by Matthews, then later popularized by Jackson and Fersht's seminal investigation of chymotrypsin inhibitor II (CI2) folding, the first clear demonstration of kinetically two-state folding (Jackson and Fersht 1991).

Chevron analysis uses a linear dependence of the equilibrium and activation free energies for folding (*f*) and unfolding (*u*) on denaturant concentration.

$$\Delta G^{o}([\text{denaturant}]) = \Delta G^{o}_{\text{H2O}} - m^{o}\,[\text{denaturant}] \qquad (1.2.1)$$

$$\Delta G^{\ddagger}_{f}([\text{denaturant}]) = -\,\text{RT ln}\, k_f^{\,H2O} - m_f\,[\text{denaturant}] + \text{Constant} \qquad (1.2.2)$$

$$\Delta G^{\ddagger}_{u}([\text{denaturant}]) = -\,\text{RT ln}\, k_u^{\,H2O} - m_u\,[\text{denaturant}] + \text{Constant} \qquad (1.2.3)$$

The Constant is proportional to the logarithm of the pre-exponential barrier-crossing attempt frequency (i.e the pre-factor in reaction rate theory), discussed in *section 1.5*.

These relationships allow kinetic free energies in the absence of denaturant to be extrapolated from measurements at moderate denaturant concentrations. The free energies of folding and unfolding are determined by the heights of the barrier relative to the unfolded and native states, respectively. Thus, in a two-state, barrier-limited system, the ratio of folding and unfolding rates is equal to the equilibrium constant, $K_{eq}=k_f/k_u$. Recovery of the equilibrium free energy from kinetic folding and unfolding is highly suggestive that both processes encounter the same rate-limiting step, rather than two different steps with coincidental free energies.

This unlikely caveat can be discarded by assessments of surface area burial during different portions of the reaction. The slopes of the free energy dependencies, their *m*-values, correlate to surface burial in amounts of 0.22 kcal mol^{-1} M^{-1} $Å^{-2}$ when denatured by guanidinium hydrochloride (GdmHCl), and 0.11 kcal mol^{-1} M^{-1} $Å^{-2}$ by urea (ref). The denaturant dependence of folding, m_f, unfolding, m_u, and stability, m^0, reflect surface area burial between the unfolded and transition states, the transition and native states, and the unfolded and native states, respectively. The ratio m_f/m^0 calculates the fractional amount of the change in exposed surface area that is buried in the transition state. When $m_u - m_f = m^0$, the amount of surface area monitored in kinetic reactions fully accounts for that observed in equilibrium. Agreement between chevron- and equilibrium-derived free energy and surface burial parameters denies the significant accumulation of an intermediate (unless that intermediate possesses the

same free energy and surface burial as the unfolded state), and identifies the reaction as two-state.

1.3 Early events in protein folding

The inability to observe intermediate stages in two-state folding has not discouraged speculation regarding the nature and importance of very fast events in protein folding. In particular, the elementary steps or processes that initiate folding have been the subject of much conjecture, in part because their rates could determine an overall speed limit for protein folding. While any number of elementary processes may be imagined, only a few are amenable to experimental analysis. Chief among these are collapse of the unfolded chain, formation of secondary structure, and "hidden" or unstable intermediates that appear but disappear faster so that their fractional population is below the resolution of the experimental apparatus. An important caveat to this discussion is the understanding that these steps need not occur sequentially (e.g. collapse followed by formation of secondary structure units which then assemble into an intermediate). All the processes could, and potentially do, occur in the same kinetic step. Furthermore, the relative contributions of these events may very well depend on the identity of the protein and experimental conditions.

Investigations of chain collapse have yielded surprising and contradictory results. Hagen and Eaton monitored Cyt c fluorescence with nanosecond resolution as the molecule unfolded in response to a laser-induced temperature jump (Hagen and

Eaton 2000). After accounting for the intrinsic temperature dependence of the fluorescence, they observed an exponential decay that they attributed to a re-equilibration of the molecular population from a compact to a more expanded conformation. However, the measured timescale of 45 μsec – 1.30 msec, between 21 and 43 °C is unexpectedly slow and the exponential behavior implies that collapse is two-state, cooperative, and barrier-limited. For these reasons re-equilibration of partially unfolded intermediates on the native side of the barrier, as seen in native state hydrogen exchange experiments on the same molecule (Bai et al. 1995), seems a more likely explanation.

Kiefhaber and coworkers examined the distance dependence of FRET rates between extrinsic dyes located at the termini of unstructured peptides, noting an inverse relationship between the number of peptide bonds and the observed rates (Bieri et al. 1999). As with Cyt c, the rate is well described by a single exponential suggestive of specific collapse, and here the observed timescale is in the nanosecond range, with a minimum at 22 °C of 20 ns when the dyes are separated by four residues.

These studies use pairs of molecular probes that report size in only a single dimension. A more relevant measure of collapse would give a sense of volume to the molecule of study. Small angle x-ray scattering (SAXS) is of limited utility for structure determination. However, because it monitors the radius of gyration (R_g) of molecules in solution, it is an ideal tool for the study of unfolded molecules (Sosnick and Trewhella 1992; Millet, Doniach and Plaxco 2002; Svergun and Koch 2003).

Recent advances have sufficiently improved the resolution of time-resolved SAXS methods at synchrotron sources to enable data collection within milliseconds of the initiation of folding (Eliezer et al. 1993; Chen, Hodgson and Doniach 1996; Noppert et al. 1998; Akiyama et al. 2002; Jacob et al. 2004).

Both equilibrium and time-resolved SAXS studies have found that some denatured proteins do not collapse upon transfer to more native solutions. An analog of unfolded hen egg-white lysozyme (HEWL), created by reduction of its four disulphide bonds, was subjected to equilibrium measurements. The addition of trifluoroethanol (TFE), known to induce helix formation, to unfolded HEWL increased both the helical content and molecular dimensions (Hoshino et al. 1997). In a time-resolved study of Protein L (ProtL), the molecule was found to display dimensions indistinguishable from those observed at high denaturant immediately after transfer to lower denaturant concentration (Plaxco et al. 1999). Furthermore, this initial state collapses with the same rate as that of the overall folding reaction. Lastly, once a protein has unfolded, further increase in the denaturant concentration does not result in increased molecular dimensions (Millet, Doniach and Plaxco 2002). The dimensions of reduced and unfolded HEWL and RNase A did not change upon the addition of 4 and 6 M GdmHCl, respectively (J. Jacob, T. Sosnick, unpublished results). These results demonstrate that early, rapid chain collapse need not occur prior to protein folding. Rather, collapse progresses concurrently with folding to the native state. In retrospect, since early collapse would bury surface area, this is an

unsurprising assertion for systems confirmed to be two-state through chevron analysis.

Kiefhaber's results are perhaps more pertinent to discussions of early secondary structure formation. Contact between dye labels separated by 3-4 peptide bonds is analogous to helix and hairpin formation, which forms hydrogen bonds between equivalently spaced residues. Though a second, independent study calculated similar rates (Lapidus, Eaton and Hofrichter 2000), both studies were performed on glycine-rich peptides, which may be expected to have greater flexibility than realistic protein sequences, so the reported rates may be artificially high. There is also some evidence suggesting that helix formation is multiphasic (Thompson, Eaton and Hofrichter 1997; Huang et al. 2002), with the slower rate attributed to helix initiation and the faster rate to propagation.

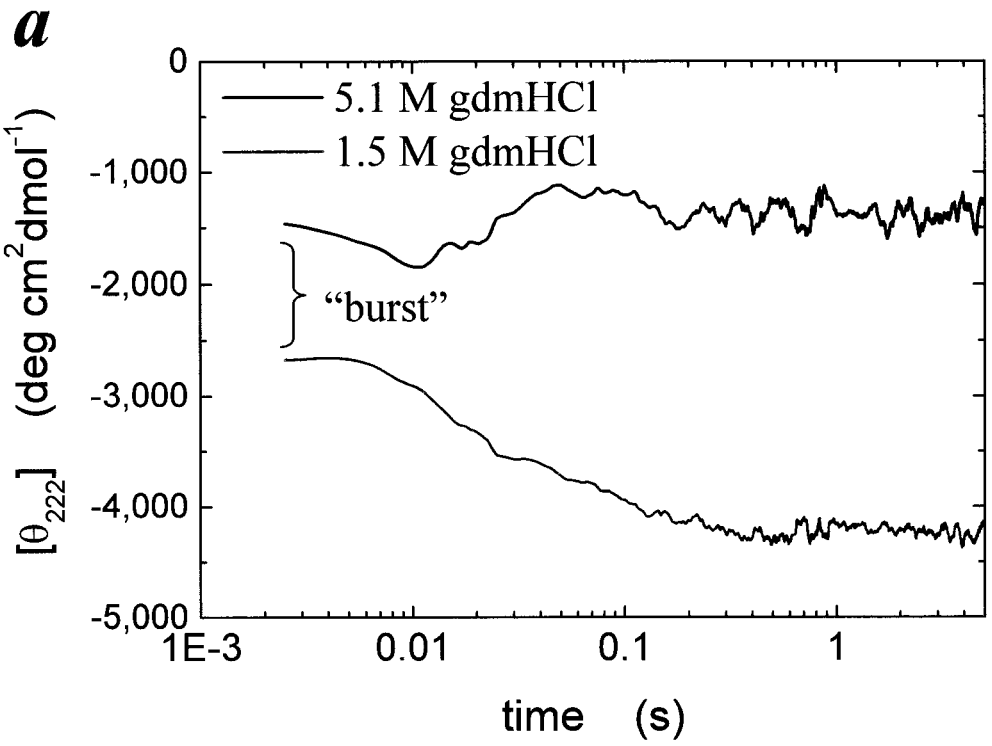
Rate constants for many of the fastest folding proteins have not been measured directly in the absence of denaturant, relying instead on the chevron extrapolation. The accumulation of early folding intermediates would be expected to cause a roll-over in the chevron plot at very low denaturant (Fig 1.1 ***d***) as has been previously seen in some systems. However, more rigorous experimental protocols found evidence for early intermediates in Cyt c (Bhuyan and Udgaonkar 2001), Barnase (Takei, Chu and Bai 2000), ubiquitin (Ub), protein G (ProtG), and CD2.D1 (Krantz and Sosnick 2000) to be either irreproducible, or consistent with a solvent-dependent response of the unfolded state (described more fully in *section 1.4*). Importantly, their revision does not deny the existence of intermediates entirely. It

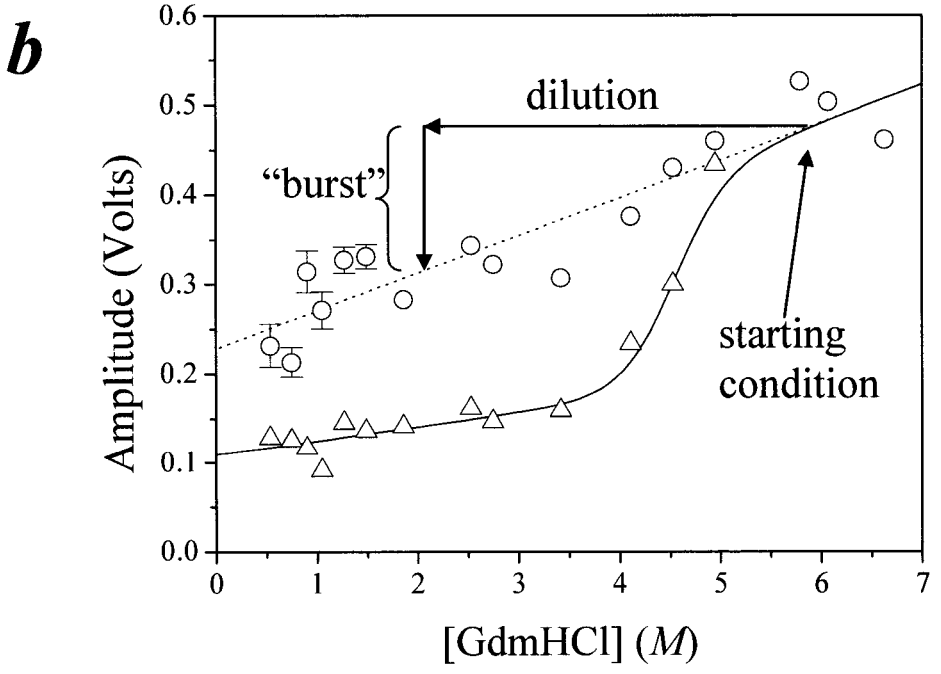
merely argues that stable intermediates generally accumulate due to misfolding, and many of the fast events can be accounted for by a solvent-dependent response of the unfolded state as discussed in the next section.

1.4 Kinetic burst phases

A common observation in kinetic reactions is the presence of unaccounted-for signals in a folding reaction - termed "burst phase" - resulting from processes faster than the detection limit of the particular measurement, for example ~ 1 msec with a commercially available stopped-flow apparatus (Fig 1.2 ***a***). Such observations are often interpreted as evidence of an early intermediate, or an increase in the amount of transient structure that forms during the delay between initiation of folding and acquisition of the first data point. For example, burst CD may mean that a fraction of residues in an unfolded protein spend more time in a helical geometry under folding conditions (though not necessarily as a true helix, which requires a string of at least four consecutive residues). These invisible burst phase events provided much of the incentive to develop kinetic observation techniques with nanosecond time resolution, such as the laser temperature jumps described in *section 1.3*. However, non-folding analogues of Cyt c (fragments; (Sosnick et al. 1997) and RNAse A (reduced; (Qi, Sosnick and Englander 1998) display the same magnitude of burst observed in the whole molecule, despite having no ability to fold. This indicates that the observed burst is largely due to a non-specific change of the unfolded state in response to new solvent conditions.

Figure 1.2 *Kinetic "burst phases". **(a)*** Relaxation trajectories of many kinetic folding experiments appear to be missing a portion of the amplitude, even after extrapolation back to the moment at which folding begins. Here, the CD_{222nm} relaxation kinetics of ubiquitin as it folds in response to dilution from 5.1 M to 1.5 M gdmHCl (red line) is displayed with the observed kinetics from an undiluted sample (black line). The folding trajectory does not extrapolate back to the unfolded value, and the missing amplitude is referred to as the "burst phase" (*figure courtesy of Jacob et al. 2004*) ***(b)*** Recovered fluorescence amplitudes (circles) are plotted as a function of denaturant concentration over the corresponding denaturation profile (triangles). Recovered amplitudes are significantly smaller at low denaturant concentration, but in a linear manner consistent with the slope of the unfolded state at high denaturant. Thus, the observed burst phase reflects only the denaturant dependence of the unfolded state rather than structure formation (*figure courtesy of Krantz and Sosnick 2000*).


a
0
-1,000
-2,000
-3,000
-4,000
-5,000
5.1 M gdmHCl
1.5 M gdmHCl
"burst"
[θ222] (deg cm2dmol-1)
1E-3
0.01
0.1
1
time (s)


b
0.6
0.5
0.4
0.3
0.2
0.1
0.0
dilution
"burst"
starting condition
Amplitude (Volts)
0
1
2
3
4
5
6
7
[GdmHCl] (M)

Any discussion of burst phase CD changes in response to denaturant dilution must consider the contribution of polyproline II (PPII) structure. PPII is more extended than helical structure, with a more exposed backbone and increased solvent-protein interactions (Garcia 2004). It is important in the present context because, whereas the CD_{222nm} signal for helix is strongly negative and that of random coil close to zero, PPII generates a positive signal. Furthermore, PPII is probably the dominant conformation of the denatured state for most proteins (Pappu and Rose 2002; Shi, Woody and Kallenbach 2002; Kentsis et al. 2004; Mezei et al. 2004), and its presence is enhanced by denaturant (N. Kallenbach, personal communication). Thus, dilution of denaturant could increase helical signal by reducing the preference for backbone conformations with PPII geometry, rather than increasing authentic helical structure.

A time-resolved SAXS and CD study of Ub and common type acyl phosphatase (ctAcp) by Jacob *et al.* illustrate the effect of solvent conditions on the unfolded states of these molecules (Jacob et al. 2004). Both proteins exhibit a burst of CD during folding, but no contraction upon the shift to aqueous conditions. For most spectroscopic probes, including CD and especially fluorescence, the unfolded baseline of a denaturation profile is sloped, demonstrating a solvent-dependent characteristic of the unfolded state (Fig 1.2 ***b***). Extrapolation of the unfolded baseline to folding conditions fully accounts for the burst-phase CD observed in ctAcP. All the stopped-flow CD data on kinetically two-state proteins appear to share this property (Zitzewitz et al. 1995; Scalley et al. 1997; Sato, Luisi and Raleigh 2000) although

mild curvature is sometimes present (Sosnick et al. 1997; Qi, Sosnick and Englander 1998; Takei, Chu and Bai 2000). For Ub, the burst phase CD below 2.5 M gdmHCl is greater than can be explained by baseline effects. However, at 0.5 M the burst actually contains a larger CD signal than the native state, as does the thermally denatured protein. This may be due to the unusually high PPII content (24%) of the native structure, which artificially lowers the native CD signal (Sreerama and Woody 2003).

The solvent response of the unfolded state is poorly understood. It often is not representative of structure acquisition, because it is rarely associated with hydrogen bond formation (Robertson and Baldwin 1991; Gladwin and Evans 1996; Sosnick et al. 1997; Krantz et al. 2002). Residual CD_{222nm} signals may reflect helix formation for some specific sequences, but could otherwise result from a shift in the average backbone dihedral angle distribution from PPII to random coil or helical conformers.

Despite the evidence, attributions of a burst phase to an obligate, on-pathway intermediate persist, and in some sense this is correct. Any unfolded molecule at high denaturant concentration must respond to the new solvent conditions before folding. However, accumulating evidence suggests that this response can be attributed to the requirement that many experiments, even those with nanosecond resolution, induce folding of a population of unfolded molecules through a change in thermodynamic conditions. If so, an accurate description of burst phase events may be tangential to the study of protein folding under native conditions.

1.5 Reaction rate theories

Demonstration that many small, single-domain proteins fold in a kinetically two-state manner with a single energy barrier and obey the chevron formalism suggests that rate theories developed for descriptions of chemical reactions may also be applied to protein folding. Justification of this application rests on a few assumptions, the most important of which being that there are no significant barriers between conformations within the same energy wells, enabling molecules of various conformations to rapidly equilibrate relative to equilibration of the entire system. Creighton observed that the experimental evidence is consistent with this scheme (Creighton 1988), and Zwanzig describes how rapid equilibration can be explained by a few reasonable assumptions about rate constants and folding thermodynamics (Zwanzig 1997).

One element common to virtually all treatments of chemical kinetics is that the reaction rate depends on the height of the energy barrier E_A, and on a probability factor analogous to the frequency with which the system "attempts" to cross the barrier. Thermal motion enables passage over the barrier, resulting in a reaction rate k proportional to a pre-exponential attempt frequency A and a Boltzman factor $\exp(-E_A/RT)$ according to the Arrhenius relation:

$$k = A \exp(-E_A/RT) \tag{1.5.1}$$

Transition state theory applied in the context of covalent bond formation assumes that the ground and excited states exist in a pseudothermodynamic equilibrium, such that:

$$k = (k_B T/h) \exp(-\Delta G^{\ddagger}/RT) \quad (1.5.2)$$

k_B and h are Boltzman's and Planck's constants, respectively, and $\Delta G^{\ddagger}$ is the Gibbs free energy of activation. Because transition state theory was developed for reactions of small molecules in the gas phase, the quantum mechanical prefactor, $k_B T/h \sim 10^{13}$ s^{-1}, overestimates the attempt frequency of a polypeptide in high-friction aqueous solvent by many orders of magnitude. Parameters that describe solvent effects can be used to modulate the predicted rates, but quantification of these factors has proven elusive for protein folding (Schlitter 1988). Transition state theory nevertheless remains attractive because differences ($\Delta\Delta G^{\ddagger}$) caused by mutation enable characterization of the transition state, although the absolute values of the attempt frequency and activation free energy may be unreliable,

An alternative formalism designed specifically for reactions in solution is Kramers' theory (Kramers 1940). This treatment models a chemical reaction as a diffusive passage over an energy barrier, and was later extended to multiple dimensions. For a reaction proceeding from substrate S to product P through an activated state, Kramers' relation can be written as:

$$k = \tau_{SX}^{-1} \exp(-\Delta U/RT) \quad (1.5.3)$$

where ΔU is the height of the potential energy barrier and τ_{SX} is the time constant in the absence of the barrier. The primary disadvantage of this formalism lies in the preexponential factor, which in many cases is poorly understood. However, because it is based on diffusional motion in a viscous solvent, thus accounting for the chain

dynamics required for exploration and barrier crossing, it is likely to be better suited for modeling and simulating the folding kinetics of proteins (Maik 1999).

1.6 Diffusion-collision

Kramers' theory forms the basis of the diffusion-collision model (D-C model) for protein folding proposed by Karplus and Weaver in 1976 (Karplus and Weaver 1976; Karplus and Weaver 1979; Karplus and Weaver 1994). It views the protein as composed of several parts (microdomains) each short enough to search though all conformational alternatives rapidly relative to the timescale of whole molecule folding. Successful folding is initiated by collision between structured microdomains. Folding rates are thus determined by several independent, but rapid and simultaneous conformational searches, and diffusion. This model can be considered to be representative of a whole class of models where pieces of secondary structure form in isolation.

Further refinement of the D-C model became possible with the development of helix prediction algorithms (Munoz and Serrano 1994; Lacroix, Viguera and Serrano 1998) using experimentally derived parameters and based on the helix-coil transition theory first postulated by Zimm & Bragg (Zimm and Bragg 1959). These algorithms predict the helical behavior of individual residues within a peptide sequence considering short range interactions, but excluding long-range and tertiary interactions. Since tertiary contacts are of prime importance for determination of protein structure, this method is unsuitable for prediction of secondary structure in

folded proteins, but ideal for application to unfolded proteins, where local interactions dominate.

The output of these algorithms can be interpreted as the probability that any particular stretch of residues within an unfolded chain will assume a helical geometry at any given time. This knowledge enables calculation of a folding rate when the appropriate diffusion rate and success frequency are known. In the simplest case, folding of a dimeric, fully helical molecule, the equation reduces to:

$$k_f = k_2 P_{1h} P_{2h} \theta \qquad (1.6.1)$$

where k_f is the folding rate, k_2 is the 2nd order, Smoluchowski-derived, bimolecular collision rate, P_{1h} and P_{2h} are the probabilities that the first and second chains are helical, and θ is the success frequency of collisions between helical segments.

The D-C model benefits from its aesthetics, the knowledge that the initiation of helix formation occurs much more quickly than even the fastest folding rates measured for entire proteins (Williams et al. 1996; Thompson, Eaton and Hofrichter 1997; Huang et al. 2002), that unfolded proteins often contain significant helical structure (Islam, Karplus and Weaver 2002; Mayor et al. 2003), and evidence that helix is present at the rate limiting step of folding for helical proteins (Krantz et al. 2002). In addition, a number of studies have shown consistency between the D-C model and observed folding rates, but explicit tests of whether helix forms prior to collision are difficult, because two-state folding behavior precludes identification of the order of events leading to the transition state.

1.7 Diffusion-collision and experiments

Two fundamental principles of the D-C mechanism are its reliance on diffusion rates and partial formation of secondary structure in unfolded proteins. Interpretation of the effects of heightened solvent viscosity are complicated by the fact that the viscogenic agents added to increase solvent viscosity also affect protein stability. Several experiments have attempted to compensate for this effect by adding denaturants until stabilities under all viscosity conditions are equal, resulting in the expected inverse relationship between folding rates and viscosity for two proteins (Plaxco and Baker 1998; Bhattacharyya and Sosnick 1999; Jacob et al. 1999). However, this approach suffers from its own problems, chiefly that no detailed compilation of data has been published regarding the viscosity of denaturant/viscogen mixtures. Plaxco & Baker used a method proposed by Tanford to estimate the viscosity of solutions containing denaturants with buffers and low concentrations of salts (Kawahara and Tanford 1966), but Sato *et al.* (Sato, Sayid and Raleigh 2000) demonstrated that this method cannot be extended to highly viscous conditions. Doing so significantly underestimates the viscosity of highly viscous solutions, leading to an overestimate of the viscosity dependence of folding.

Transient secondary structure formation in unfolded proteins is the second tenet of the D-C model. Though the presence of a burst-phase in kinetic experiments is sometimes asserted as evidence of this behavior, the argument is perilous due to the caveats described in *section 1.4*. Helix prediction algorithms described above have been used to predict the probability of helix formation in unfolded helical proteins

subjected to mutation, and the predicted positive correlation between helicity and the folding rates of helical proteins has been observed (Zitzewitz et al. 2000). However, as discussed more fully in chapter 2, the same correlation could be attributed to decreased activation energy because enhanced helicity stabilizes all conformations containing helical structure, including the transition state. The studies just described cannot deconvolute these two effects.

Sosnick *et al.* evaluated the applicability of the D-C model to folding of a fully helical coiled coil derived from yeast transcription factor GCN4 (Sosnick et al. 1996). Glycine mutations at several surface residues decrease helix propensity with little effect on folding rates, leading to the conclusion that helix was not necessary at the moment of collision. Myers and Oas reinterpreted this data (Myers and Oas 1999), demonstrating that the use of a helix prediction algorithm and the D-C formalism predicted folding rates consistent with observed values. However, their analysis assumed a uniform helical propensity belied by the helix prediction algorithm, which attributes more than 85% of the helicity to the C-terminal third of the protein. This error seriously overestimates the probability of encountering long helical stretches in isolated monomers, and underestimates the effects of mutation in the C-terminal region. As a consequence, their calculated rate is too fast, and the formalism fails to predict the results of mutations (Moran et al. 1999).

Moran *et al.* also uncovered multiple folding pathways and showed that single mutations cause a redistribution of the fraction of folding events employing each pathway. This redistribution could explain the insensitivity of folding rates to

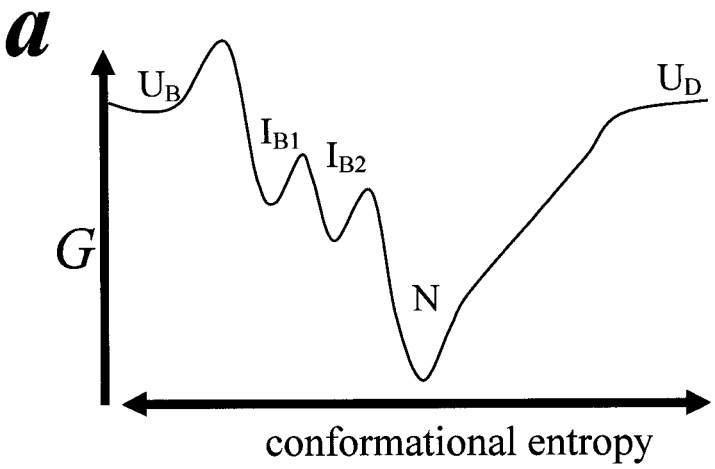
mutation, but the same study also shows that the major fraction of folding events need not begin from the most helical region, refuting the validity of the D-C model once again for the folding of this protein.

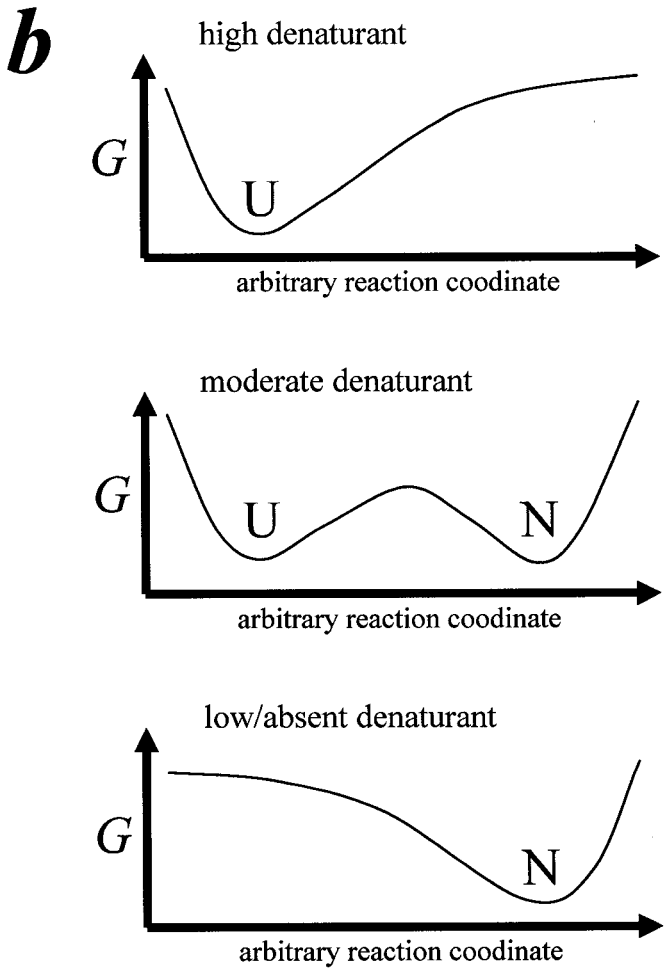
1.8 Landscape Theory

Experimentalists tend to view the transition state as a single, homogenous entity traversed by all molecules as they fold (Fersht et al. 1994; Martinez, Pisabarro and Serrano 1998; Grantcharova, Riddle and Baker 2000). The discovery of multiple folding pathways in GCN4, and later in ProtG and Prot L (Grantcharova et al. 2001), illustrates that this is not always the case, and provides evidence in support of a view preferred by theoreticians – that protein folding is better modeled by a statistical free energy landscape or funnel (Bryngelson et al. 1995). A landscape allows individual molecules to pass over any of many structurally distinct, but energetically similar activation barriers as they fold, and even describes scenarios where the barriers encountered are lower in energy than the unfolded state.

One goal of landscape theory is to map folding on the high dimensionality of the energy surface to diffusion along a one-dimensional reaction coordinate (Socci, Onuchic and Wolynes 1998). Toward this end, illustrations of energy landscapes and funnels often express the number of conformations accessible to a protein (in two dimensional figures, the width along the x-axis) relative to its free energy (y-axis) (Fig 1.3 ***a***). Energy is often expressed in units of RT and referred to as "temperature" for this reason. High temperatures allow the protein to assume a cosmologically large

Figure 1.3 *Folding free energy surfaces.* ***(a)*** The energy of a system (y-axis) is sketched against its conformational entropy (width along the x-axis). At high energies, individual molecules within a population assume the largest number of possible conformations. As the energy of the system is lowered, fewer and fewer conformations are accessible. At the lowest energy, all molecular conformations converge on the native state. Unfolded molecules of conformation U_D can fold to the native state in a downhill manner. Folding from U_B will be barrier-limited, unless the barrier height between U_B and I_{B1} is greater than that between U_B and U_D. ***(b)*** The dominant shape of a folding free energy surface is influenced by experimental conditions such as temperature or here, denaturant concentration. Under highly destabilizing conditions (e.g. high denaturant), the free energy surface contains a single minimum corresponding to the unfolded state (*top panel*). Under moderately destabilizing conditions, a second minimum similar to the native state develops, the system is bimodal, and folding is barrier limited (*middle panel*). Under conditions of extreme stability the system again has only one minimum, this time corresponding to the native state (*bottom panel*). Folding on this surface is downhill; however, proteins of sufficient stability to fold in this regime have yet to be found.


a
U_B
I_{B1}
I_{B2}
U_D
G
N
conformational entropy


b
high denaturant
G
U
arbitrary reaction coodinate
moderate denaturant
G
U
N
arbitrary reaction coodinate
low/absent denaturant
G
N
arbitrary reaction coodinate

number of conformations. As the temperature of the system is reduced, the molecule accesses fewer and fewer conformations, until only the native fold remains. Thus, the unfolded state exists at the top of the funnel, and the native state in the energy well at the bottom.

The free energy landscape has been described with three different shapes relative to an arbitrary folding reaction coordinate (Fig 1.3 ***b***). At unstable conditions (e.g. high denaturant concentration), the free energy function has a single minimum, corresponding to the unfolded state, and folding does not occur. As conditions become increasingly stable, a second minimum very similar to the native structure develops, and the two minima are separated by an energy barrier. Folding on a free energy surface of this shape is referred to as Type I, or barrier-limited. Finally, under extremely stable conditions, there is again a single minimum representative of the native structure. This last surface corresponds to Type 0 or "downhill" protein folding.

1.9 Downhill folding

The search for downhill folding is spurred by more than just the lure of the unconventional, and desire to assess the accuracy of theoretical landscape models. It may also allow characterization of intermediate structures which do not accumulate in two-state, barrier limited systems. The unfolded state in downhill folding is analogous to the energy barrier of a barrier-limited system in that it occupies the position of highest free energy on the reaction coordinate, yet all intermediates lie toward its

native side. Thus, the techniques employed by Englander's group to identify partially unfolded conformations on the native side of the energy barrier to Cyt c folding (described in *section 1.11*) may also allow characterization of as yet invisible intermediate states if applied to a downhill system.

It is perhaps self-evident that a barrier-free process such as downhill folding must be extremely fast, but just how fast remains subject to some conjecture. Downhill folding behavior is predicted to occur when folding rates approach the value of the attempt frequency of the barrier crossing process, $k_f = k_{attempt} e^{\Delta G^{\ddagger}/RT} \sim k_{attempt}$ when $\Delta G^{\ddagger} \sim 0$. Since the attempt frequency is likely to be no faster than the slowest elementary steps of folding, such as collapse or helix initiation, the measured rates for these elementary steps reviewed in *section 1.3* are often used to estimate the maximum rate of folding. Rates measured for formation of structural elements in peptide fragments suggest a minimum folding time constant of about 0.5 μsec for very short helical proteins, while proteins containing a high proportion of β-sheet structure could fold as much as ten-fold slower.

Kramers' reaction rate theory would enable calculation of the barrier-free time constant, τ_{SX}, from observed rates if the barrier height can be estimated accurately. Theoretical calculation of the barrier height by Munoz and Eaton (Munoz and Eaton 1999) and Portman *et al.* (Portman 2001), yield minimum folding times of 10 and 0.4 μsec, respectively. Theories of polymer collapse yield no absolute rate, but demonstrate that the rate will scale with length, and the increased ruggedness of energy landscapes for stable proteins suggests an additional correction (Kubelka,

Hofrichter and Eaton 2004). There are caveats to all these estimates, but both experiment and theory yield a similar time constant of ~N/100 μsec where N is the number of residues, for downhill folding of single domain proteins.

1.10 Downhill folding experiments

Experimental evidence of downhill folding is meager at best. Non-exponential folding kinetics have been reported in nanosecond temperature-induced refolding measurements of a cold-denatured Ub mutant (Sabelko, Ervin and Gruebele 1999), which the authors cite as evidence for downhill folding. This result is puzzling, since Ub folds much more slowly than would be expected for a barrier-free process. Folding traces are exponential at 2 °C (after accounting for a solvent-dependent response of the unfolded state). More complex kinetics are observed at 8 °C because a fraction of the –8 °C, cold-denatured ensemble is postulated to assume a configuration residing on the native-side of the rate-limiting barrier on the new +8 °C reaction surface after the temperature jump. In effect, this sub-population folds downhill, but not from the unfolded state, so the reaction surface remains consistent with a barrier-limited system. More fundamentally, Hagen (Hagen 2003) and Bryngelson (Bryngelson et al. 1995) describe scenarios in which downhill folding could be exponential and barrier-limited behavior non-exponential, so the observation of non-exponential kinetics alone can neither confirm nor deny downhill behavior.

Garcia-Mira and coworkers have recently claimed to observe downhill folding for *E. coli* BBL protein, a small, 2½ -helix protein (Garcia-Mira et al. 2002). They

demonstrated that the equilibrium folding is non-cooperative. In conjunction with its likely μsec folding timescale (no kinetic experiments were conducted), the authors assert that the folding reaction is energetically downhill. However, equilibrium folding experiments traverse the protein's thermodynamic midpoint, so the huge thermal driving force required for downhill folding is not present under these experimental conditions, and may not exist under any conditions for this unstable molecule.

This requirement for highly stabilizing conditions is perhaps the most problematic for investigations of downhill folding. Traditional folding experiments measure relaxation rates as a population of molecules changes from predominantly unfolded state to predominantly folded state or vice versa in response to changes in solvent conditions. Such protocols generally necessitate the use of denaturants (Pascher et al. 1996) or highly destabilizing mutants (Sabelko, Ervin and Gruebele 1999). The net result is that the final stability of the protein is low. Similarly, NMR line-broadening measurements (Burton et al. 1996) can only be conducted under marginally stable conditions because the technique requires that the protein fold and unfold at comparable rates during the course of the pulse sequence and FID acquisition, i.e. $0.05 < k_f/k_u = K_{eq} < 20$. Nanosecond laser-induced temperature jumps can only increase temperatures by ~15 °C, causing only a mild shift from marginally unstable to marginally stable conditions (e.g. -8°C to +8°C). Since downhill folding requires that the native state be strongly favored, it is unlikely to be observed in these experiments.

1.11 Native state hydrogen exchange

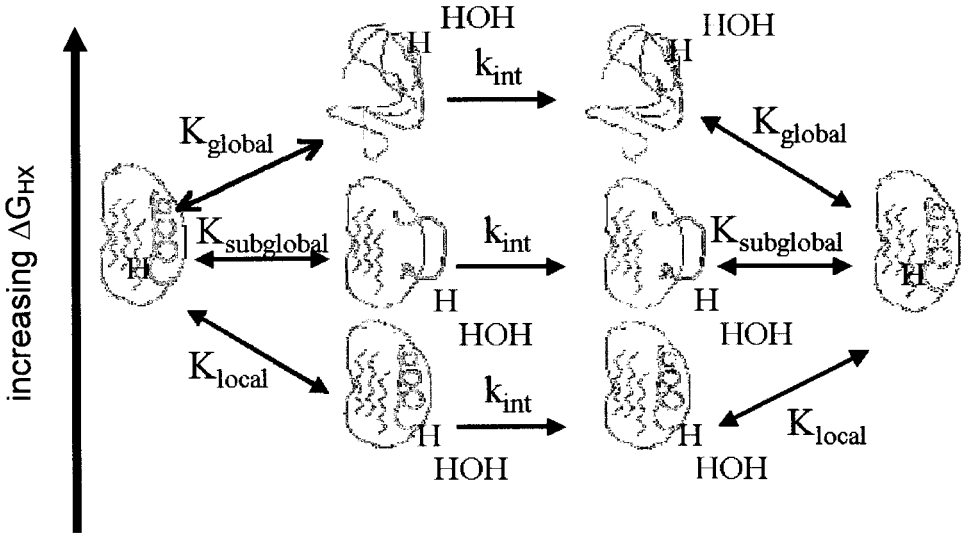
Native state hydrogen exchange can monitor spontaneous protein dynamics under strongly native conditions. Its utility rests on the observation that the amide protons of a peptide backbone can exchange with those of solvent molecules only when they are not involved in a intramolecular hydrogen bond. Hence, the exchange of protons involved in a native, H-bonded structural element signals structural change. The analysis of HX rates in relationship to protein dynamics is discussed in Chapter 3.

The energetics of these structural changes can be usefully grouped into three different regimes (Fig 1.4 ***a***). Low energy events require only small, local fluctuations such as the looping out of a single helical turn. These occur with little change in exposed surface area, so exchange rates at these positions are relatively insensitive to denaturant. Protons that require a regional loss of structure exchange with a moderate denaturant dependence, because of the greater energetic cost and additional surface area exposure required for regional unfolding. Finally, the exchange of a few protons may require complete unfolding of the molecule. Since global unfolding is energetically costly and exposes the maximum amount of surface area, exchange of these protons is very slow with a denaturant dependence similar to that calculated from whole-molecule denaturation profiles.

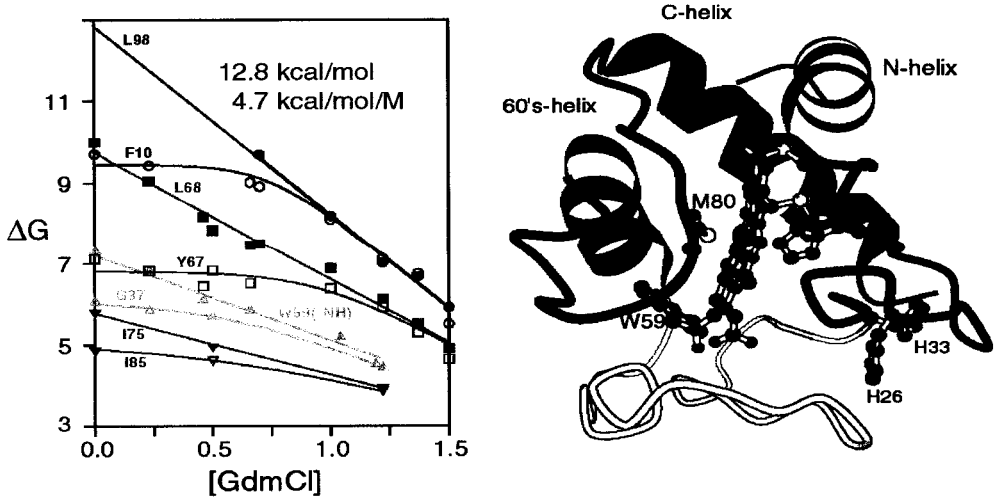
In 1995, Bai and coworkers (Bai et al. 1995) monitored the exchange behavior of Cyt c as a function of denaturant concentration using native state

Figure 1.4 Native state hydrogen exchange. ***(a)*** The utility of hydrogen exchange rests on the observation that the amide protons of a peptide backbone can exchange with those of solvent molecules, but not when they are involved in a hydrogen bond, so the exchange of protons involved in a native, H-bonded structural element signals structural change. The energetics of these structural changes can be usefully grouped into three different regimes. Low energy events require only small, local fluctuations, such as the looping out of a single helical turn and occur at high frequency, so exchange rates at these positions are extremely fast. Protons that require a regional loss of structure exchange more slowly because of the greater energetic cost of regional unfolding. Finally, the exchange of a few protons may require complete unfolding of the entire protein. Since global unfolding is the most energetically costly, exchange of these protons is the slowest. ***(b)*** Denaturant dependence of NsHX for representative protons of the four exchange regimes in Cyt c identified by Bai *et al.* (Bai et al. 1995). The 70-to-85 loop (represented by I75 and I85 in red), exchanged with the lowest free energy and no denaturant dependence at low concentrations, followed by the 36-to-61 loop (G37 and W59, yellow), the 20-to-35 loop and 60s helix (Y67 and L68, green), and lastly, the amino- and carboxy-terminal helices (F10 and L98, blue). Thus, energetically uphill unfolding begins from the 70-to-85 loop, progresses through the 36-to-61 loop, then the 20-to-35 loop and 60s helix, and culminates at the terminal helices. Folding proceeds in the reverse order (*figures adapted from Bai et al. 1995*).

a


increasing ΔG$_{HX}$
HOH
H
k_{int}
K_{global}
$K_{subglobal}$
K_{local}

b


L98
12.8 kcal/mol
4.7 kcal/mol/M
F10
L68
Y67
I75
I85
ΔG
[GdmCl]
C-helix
N-helix
60's-helix
M80
W59
H33
H26

hydrogen exchange (Fig 1.4 ***b***). Protons from different regions of the molecule exchanged in four regimes, each with its own characteristic free energy of exchange and denaturant dependence. Each region was assigned a color (red, yellow, green, or blue, in increasing order). The 70 to 85 loop exchanged with the lowest energy and denaturant dependence (red), followed by the 36 to 61 loop (yellow), the 20 to 35 loop and 60s helix (green), and lastly, the amino- and carboxy-terminal helices (blue). Thus, energetically uphill unfolding proceeds through a sequence of partially unfolded forms

$$N_{RYGB} \leftrightarrow P_{YGB} \leftrightarrow P_{GB} \leftrightarrow P_{B} \leftrightarrow U$$

where the subscripts indicate the colored regions that retain structure. Since frequent unfolding events must be matched by equally frequent refolding steps, and infrequent with infrequent, in order to maintain the microscopic equilibrium concentration of each species, the principle of microscopic reversibility requires that folding occurs in reverse order.

The uniqueness of native state hydrogen exchange is that it does not require the experimentally-induced unfolding and refolding of large populations of molecules, and can thus be performed under conditions that strongly favor the protein's native state. In addition to Bai's groundbreaking study, this method has since been used monitor the folding of Ub (Sivaraman 2001), RNase H and barnase (Takei, Chu and Bai 2000) and the μsec-sec fluctuations of ~20 amide protons in OMTKY3 (Arrington and Robertson 2000).

CHAPTER 2

FAST FOLDING OF A HELICAL PROTEIN INITIATED BY THE COLLISION OF UNSTRUCTURED CHAINS

All the data discussed in this chapter is my own, but I wish to express my gratitude for the contributions of many others. In addition to those acknowledged in the beginning of this volume, I also thank Giridher Reddy of the University of Chicago for peptide synthesis, and Professors Martin Karplus of Harvard University, Neville Kallenbach of New York University, and S. Walter Englander of the University of Pennsylvania for comments and discussions specific to this chapter.

2.1 Abstract

In order to examine whether helix formation necessarily precedes chain collision, we have measured the folding of a fully helical coiled coil which has been specially engineered to have negligible intrinsic helical propensity, but high overall stability. The folding rate approaches the diffusion-limited value and is much faster than possible if folding is contingent upon pre-collision helix formation. Therefore, the collision of two unstructured chains is the initial step of the dominant kinetic pathway, while helicity exerts its influence only at a later step. Folding from an unstructured encounter complex may be efficient and robust, which has implications for any biological process that couples folding to binding.

2.2 Introduction

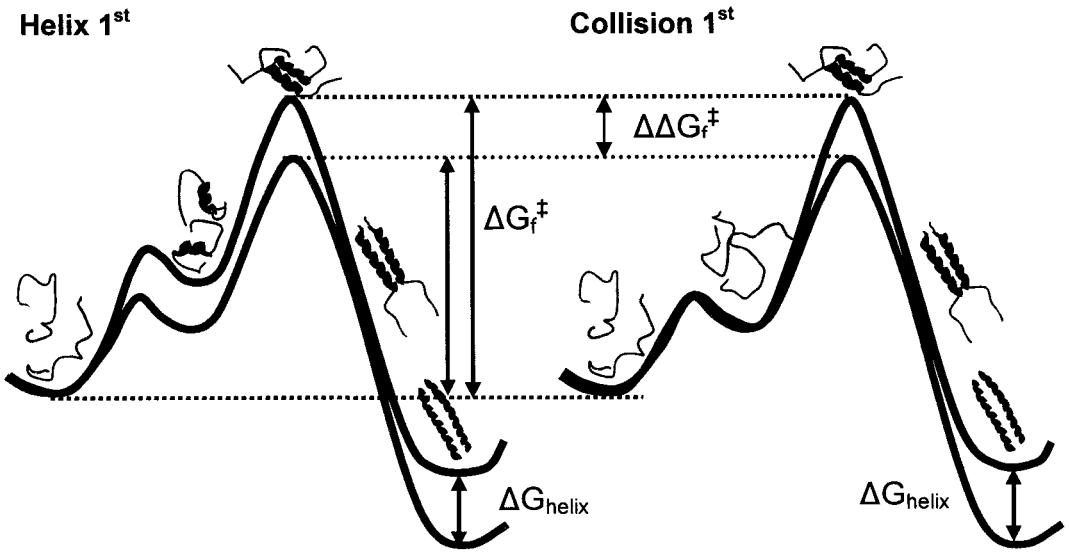
One of the most debated issues in protein folding concerns the earliest folding events leading up to the transition state (Zitzewitz et al. 1995; Sosnick et al. 1996; Moran et al. 1999; Myers and Oas 1999; Zitzewitz et al. 2000; Bosshard et al. 2001; Ibarra-Molero, Makhatadze and Matthews 2001; Krantz and Sosnick 2001; Mayor et al. 2003). For helical proteins, the earliest productive folding steps often are postulated to involve the collision of two pre-formed, but not necessarily stable, helical elements, rather than collision of unstructured chains. This "Diffusion-Collision Model" (D-C Model) (Karplus and Weaver 1976; Karplus and Weaver 1994) is supported by the observation that helix formation occurs faster than overall folding rates (Williams et al. 1996; Thompson, Eaton and Hofrichter 1997; Huang et

al. 2002). This broadly accepted view also is based on an increase in folding rates with an increase in helical propensity.

However, this correlation can support an opposing model where unstructured chains first collide, and the enhanced helicity increases the success frequency (or transmission coefficient) of each encounter (Fig. 2.1). Generally, the highly cooperative (two-state) folding behavior of most small proteins precludes identifying the order of events leading up to the kinetic barrier. As a result, the demonstration of helical structure in the transition state cannot by itself resolve whether helix formation or chain collision occurs first.

This obstacle can be overcome by studying a system with minimal helical propensity and composed of more than one chain, so that the rate of collision can be varied (Wagman, Dobson and Karplus 1980). These properties enable comparisons between observed folding rates and the maximum rate consistent with a model where pre-collision helix formation is required. Our investigation uses this strategy in conjunction with a specially engineered dimeric coiled coil protein having negligible intrinsic helicity but high stability. This protein folds at nearly the diffusion limit, and certainly much faster than possible if helix must form prior to collision. Thus, an unstructured encounter complex can successfully initiate rapid folding, with helix formation occurring at a later step. The collision-first route sets a high basal level for the folding rate of any protein.

Figure 2.1 Acceleration of folding rates through helix stabilization does not identify whether helix formation comes before or after collision. The wild-type protein's free energy surface (red line) is shifted up (blue line) upon helix destabilization, ΔG_{helix}. For both scenarios, this destabilization slows folding identically through an increase in activation energy, $\Delta\Delta G_f^{\ddagger} = \Delta G_{helix}$. In the Diffusion Collision Model, the slower folding rate is due to a reduction in the probability of helix formation prior to collision. The same rate reduction, ΔG_{helix}, is observed when helix forms after collision, because access to the transition state is thermodynamically less favorable. Hence, slower folding alone cannot distinguish the order of events leading up to the transition state. However, the use of a dimeric system with negligible intrinsic helicity, along with an analysis of folding rates, can distinguish the two models when folding is extremely fast, as is the case for GCN4-E_9G_4.


Helix 1st
Collision 1st
ΔΔGf‡
ΔGf‡
ΔGhelix
ΔGhelix

2.3 Material and methods.

Peptides. GCN4-E_9G_4 was prepared and characterized as described in (Moran et al. 1999).

Equilibrium measurements. CD measurements used a Jasco 715 spectropolarimeter, in equilibrium mode with pathlength of 1 cm. All experiments were carried out in 20 mM sodium phosphate, 200 mM sodium chloride, except the pH-dependence of kinetics was determined in 20 mM sodium citrate and 200 mM sodium chloride

Stopped-flow spectroscopy. Rapid mixing fluorescence experiments used a Biologic SFM-400 stopped-flow apparatus connected via a fiber optic cable to a PTI A101 arc lamp. Fluorescence spectroscopy used excitation and emission wavelengths of 280-290 nm and 300-400 nm, respectively. CD relaxation measurements were conducted at 2 nm resolution with a path length of 0.8 cm using a Biologic SFM-4 interfaced with CD spectropolarimeter.

Data analysis. The diffusion constant for unfolded chains is calculated according to the Einstein-Sutherland Equation, $D = kT/(6 \pi \eta R_H)$, where η is the solvent viscosity, and R_H is the hydrodynamic radius obtained from diffusion measurements of unfolded chains (Wilkins et al. 1999). Encounter rates are calculated according to Smoluchowski's Equation $k_2 = 8\pi N_A R_{ij} D$, using a contact radius R_{ij}=10 Å. The pH dependence of stability, folding, and unfolding was fit to the following equations for a model that accommodates differential binding affinities in the unfolded, native and transition state conformations:

$$K_{obs}=K_{eq}((1+10^{pKaN-pH})/(1+10^{pKaU-pH}))^{n} \quad (2.3.1)$$

$$k_{f(obs)} = k_f((1+10^{pKaN-pH})/(1+10^{pKaU-pH}))^{nF} \quad (2.3.2)$$

$$k_{u(obs)} = k_f((1+10^{pKaU\text{-}pH})/(1+10^{pKaN\text{-}pH}))^{nU} \quad (2.3.3)$$

where *pKaN* and *pKaU* are the pK_as of the glutamic acid side chains in the native and unfolded states, respectively, *n* is the number of protons bound during folding, *nF* and *nU* the number of protons bound before and after the transition state. The value for *pKaU* is fixed at the intrinsic pKa for glutamic acid (4.4), and the value of *pKaN* determined by the equilibrium fit is used as a fixed input parameter in the fit to the kinetic data.

2.4 Results and Discussion.

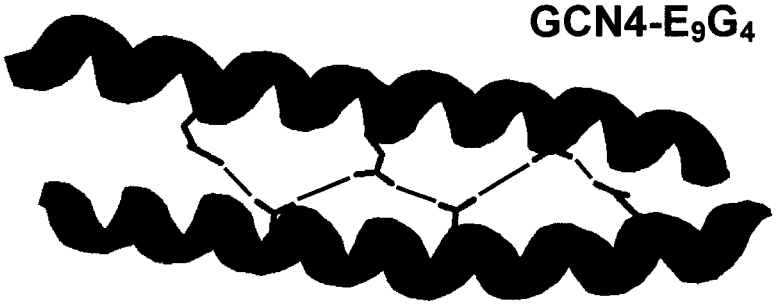
2.4.1 Design of a stable coiled coil with negligible intrinsic helicity.

Dimeric coiled coils are an ideal system to test the D-C Model because they are composed of only two docked helices. The coiled coil presently investigated is derived from GCN4-p1, the leucine zipper region of yeast transcription factor GCN4 (Fig. 2.2 *a*). We have engineered a 30-residue version, GCN4-E_9G_4, which has near-zero intrinsic helicity but wild-type stability (ΔG^o=10 kcal mol^{-1}). To reduce helicity, four glycines are substituted at solvent exposed positions (eight total in the dimer). As a result, the isolated chains are devoid of helical content under aqueous conditions (Fig. 2.2 *b*, inset). Helix-coil calculations predict a maximum helicity of only 2.7% at any single position, and less than 1% averaged across the isolated chain at 35 °C (Fig. 2.2 *c*) (Munoz and Serrano 1994; Lacroix, Viguera and Serrano 1998).

To offset the stability lost due to low helical propensity, nine glutamic acids are introduced in each chain at positions next to the hydrophobic interface (Durr, Jelesarov and Bosshard 1999). Below pH 4, the carboxylic side chains protonate

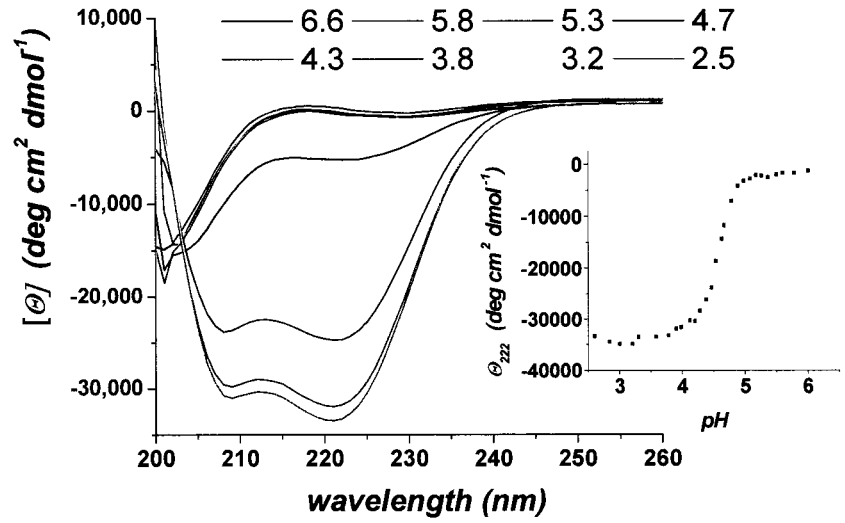
Figure 2.2 Design properties of stable coiled coil with negligible intrinsic helicity. ***(a)*** Coiled coils consist of heptad repeats $(abcdefg)_n$. Hydrophobic residues are located at the interfacial positions *a* and *d*. Charged residues are often located at the *e* and *g* positions and form ionic interactions that bridge the two helices. The *b*, *c*, and *f* positions are the most solvent exposed. The parent GCN4-p1 sequence was altered to include four glycines in positions *b*, *c* or *f*, and nine glutamic acids in *e* or *g* positions. Possible intermolecular hydrogen bonds between glutamic acid side chains are illustrated for one face of the molecule. An equivalent number on the second side are left out for clarity. ***(b)*** Secondary structure content of native (pH 3.2) and denatured (pH 5.3) GCN4-E_9G_4 are characteristic of a fully folded and a fully unfolded helical protein, respectively. (*inset*) The midpoint of the pH titration is near the intrinsic pK_a of glutamic acid side chains (4.4) while the steepness indicates that folding is driven by the protonation of 7 ± 1 glutamic acids. ***(c)*** Intrinsic helix propensity for isolated monomers of GCN4-E_9G_4, independent of tertiary interactions at 35° C (Munoz and Serrano 1994; Lacroix, Viguera et al. 1998). The absence of helical structure in the monomer above pH 5.3 (green line) is accurately reproduced by helix-coil calculations and insensitive to decreased pH. The glutamic acid side-chains are too far apart to interact with each other in a helical geometry in the monomers. For GCN4-E_9G_4 (red line), the maximum probability for one, two and just under four turns (the amount actually present at the transition state) is 2.4, 0.8% and 0.6%, respectively. For these amounts of pre-collisional helix formation, the D-C Model predicts maximum folding rate constants of 1.6×10^6, 1.8×10^5, and 1.0×10^5 $M^{-1} s^{-1}$ respectively, which the observed rate exceeds by 25-400 fold.

a

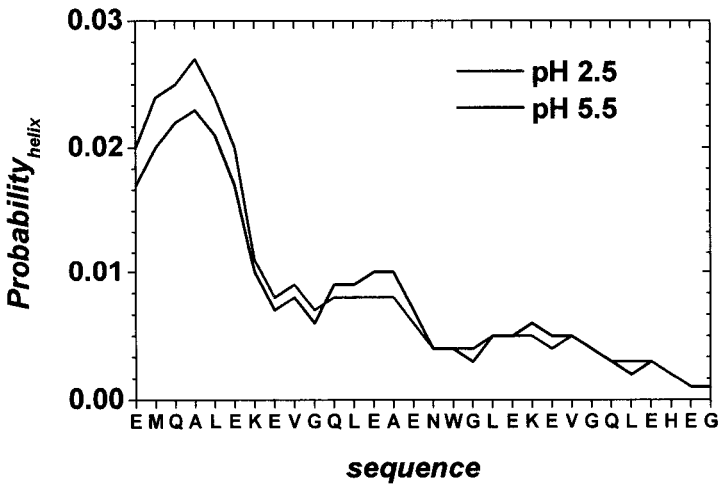


GCN4-p1: G MKQLEDK VEELLSK NWHLENE VARLKKL VGE
GCN4-E_9G_4: E MQALEKE VGQLEAE NWGLEKE VGQLEHE G
heptad position *abcdefg*

b



c



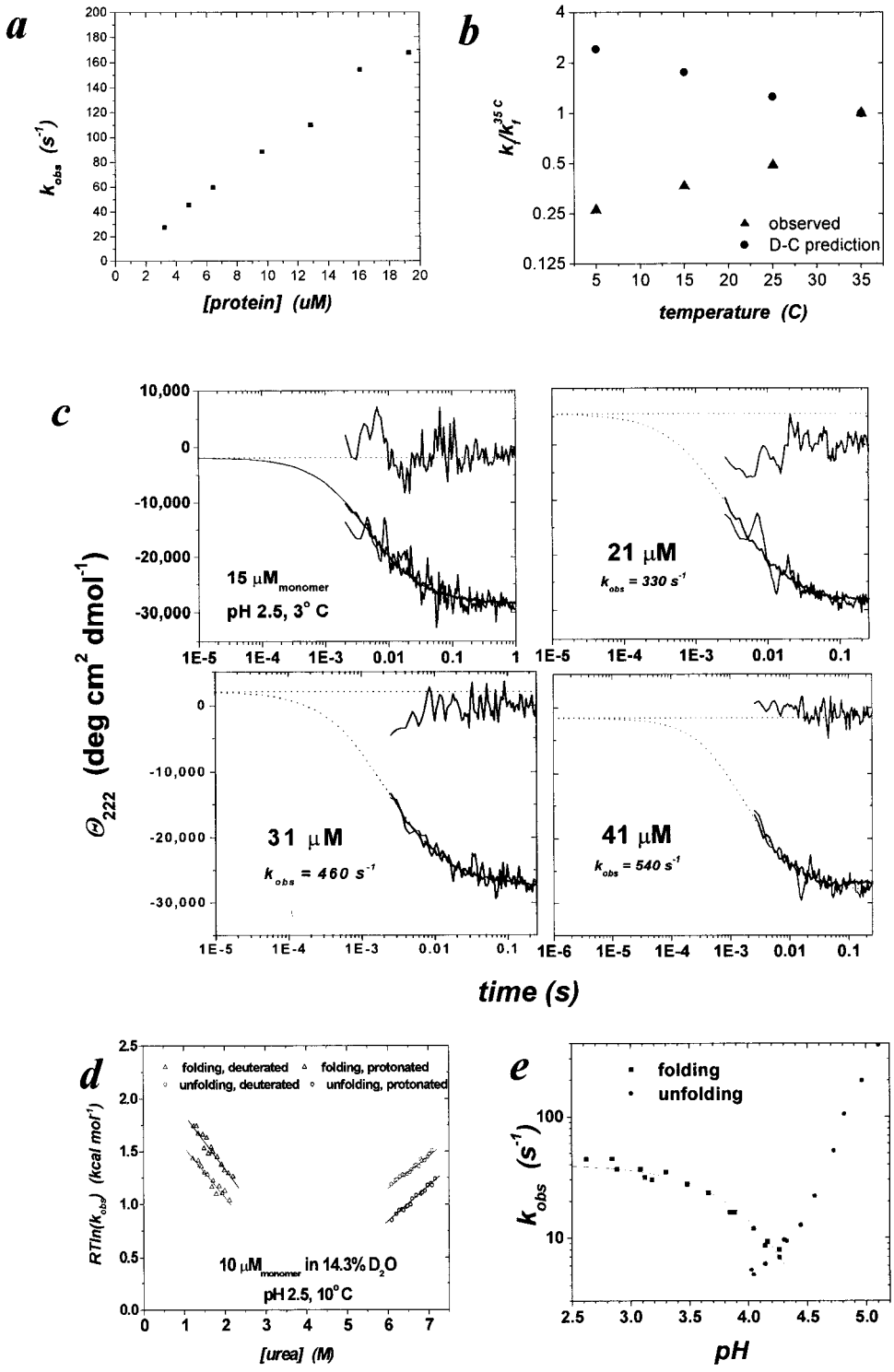
and form stabilizing tertiary interactions across the dimer interface - "glutamic staples" - via either hydrogen bonds or hydrophobic interactions. The protein, which is unfolded at neutral pH, folds to a stable, fully helical structure upon acidification (Fig. 2.2 *b*). The midpoint of the equilibrium pH titration is 4.5, very near the intrinsic pK_a of the glutamic acid side chains.

2.4.2 Two-state folding kinetics

The folding kinetics of GCN4-E_9G_4 are well described by a 2(Monomer) $\leftrightarrow$ Dimer bimolecular reaction (Fig. 2.3). The observed folding rate is proportional to protein concentration, $k_{obs} = k_f \bullet$[protein] (Fig. 2.3 ***a***), while the unfolding process is first order and independent of protein concentration. The second order folding rate constant is temperature dependent, increasing from $k_f = 1 \times 10^7\ M^{-1}\ s^{-1}$ at 5° C to $4 \times 10^7\ M^{-1}\ s^{-1}$ at 35° C (Fig. 2.3 ***b***). Rates are independent of whether folding is initiated by a decrease in pH or the dilution of denaturant. Additionally, the same rate constant is observed by a global probe sensitive to helical content (circular dichroism at 222 nm) or a local probe (total fluorescence of the sole tryptophan), consistent with the two-state nature of the reaction (Fig. 2.3 *b*).

The highly cooperative, two-state folding behavior is confirmed through a "chevron analysis" with a linear dependence of the equilibrium and activation free energies for folding (*f*) and unfolding (*u*) on GdmCl concentration (Matthews 1987) (Fig. 2.3 ***d***).

Figure 2.3 Folding behavior of the designed coiled coil GCN4-E_9G_4. ***(a)*** A two-fold increase in the protein concentration doubles the rate, while the intercept is near zero ($1 \pm 5\ s^{-1}$), confirming that folding is limited by a step involving the collision of two monomers (performed at 10 °C, pH 2.5). ***(b)*** Observed and predicted temperature dependence of folding rates. Decreasing temperature has two antagonistic effects on folding rates. The decrease in temperature from 35 to 5° C results in a two-fold decrease in the bimolecular collision rate but a two fold increase in helix propensity. As $k_f \propto k_2 P_{helix}^2$, the rate predicted by the D-C Model (blue line) should increase as temperatures drop. Measured rates (red line) exhibit the opposite trend, indicating that the D-C Model does not correctly predict the temperature dependence. ***(c)*** Rapid mixing folding measurements monitoring $CD_{222\ nm}$ (—) and fluorescence (—, scaled to the CD trace). The CD trace extrapolated to t=0 s (—, using $k_{fluor} = 210\ s^{-1}$) matches the value at its initial, unfolded CD value at pH 5.5 (—) indicating that the unfolded monomers have negligible helical content prior to collision and productive folding. ***(d)*** The denaturant dependence of the activation energies for folding (left side) and unfolding (right side) indicates that 61 ± 6% of the net surface area is buried at the transition state (m_f/m_o Matthews 1987). Similarly, the change in folding activation free energy upon backbone deuteration relative to the change in equilibrium stability $\Delta\Delta G^{\ddagger}_{f}{}^{D\text{-to-}H}/\Delta\Delta G_{eq}{}^{D\text{-to-}H}$ (Moran et al. 1999), indicates that 49 ± 2% of helical hydrogen bonds are formed in the transition state. ***(e)*** The pH dependence of folding (black) and unfolding (red) demonstrate that 46 ± 11% of the glutamic acid side chains are protonated in the transition state (eq. 2.3.2 and 2.3.3).


a
k_{obs} (s^{-1})
[protein] (uM)
b
$k_f/k_f^{35\,C}$
observed
D-C prediction
temperature (C)
c
Θ_{222} (deg cm^2 dmol^{-1})
15 μM monomer
pH 2.5, 3° C
21 μM
k_{obs} = 330 s^{-1}
31 μM
k_{obs} = 460 s^{-1}
41 μM
k_{obs} = 540 s^{-1}
time (s)
d
folding, deuterated
folding, protonated
unfolding, deuterated
unfolding, protonated
RTln(k_{obs}) (kcal mol^{-1})
10 μM monomer in 14.3% D_2O
pH 2.5, 10° C
[urea] (M)
e
folding
unfolding
k_{obs} (s^{-1})
pH

$$\Delta G^{o}([\text{GdmCl}]) = \Delta G^{o}_{\text{H2O}} - m^{o}\,[\text{GdmCl}] \qquad (2.4.1)$$

$$\Delta G^{\ddagger}_{f}([\text{GdmCl}]) = -\,\text{RT ln}\, k_f^{\,H2O} - m_f\,[\text{GdmCl}] + \text{Constant} \qquad (2.4.2)$$

$$\Delta G^{\ddagger}_{u}([\text{GdmCl}]) = -\,\text{RT ln}\, 2k_u^{\,H2O} - m_u\,[\text{GdmCl}] + \text{Constant} \qquad (2.4.3)$$

When folding is effectively two-state, the equilibrium values for the change in free energy and surface burial can be calculated from kinetic measurements according to: $-\Delta G^{\circ}_{H2O} = \Delta G^{\ddagger}_{u} - \Delta G^{\ddagger}_{f}$ and $m^{\circ} = m_u - m_f$,. The equivalence of thermodynamically and kinetically determined values for ΔG^{o}_{H2O} and m^{o} demonstrates the applicability of a two-state model for GCN4-E_9G_4 folding. In particular, the folding arm of the chevron remains linear under all conditions. This linearity indicates that no surface is buried in the burst phase upon transfer from high to low denaturant.

In addition, no helical structure is formed in a submillisecond "burst-phase" at temperatures as low as 3° C (Fig. 2.3 *b*), and the predicted helicity at pH 2.5 is equivalent to that at pH 5.5, where the chain is seen to be devoid of helical structure (Fig. 2.2 *b*). Also, the glutamic acids are too far apart to interact with each other in a helical geometry within individual monomers. Thus, multiple experimental measures all indicate that helical structure does not form to an appreciable degree prior to the major folding event.

2.4.3 Folding is faster than predicted by D-C Model

In the D-C model, the folding rate equals the product of three quantities, $k_f = k_2 P_{helix}^2 \theta$ where k_2, P_{helix}^2, and θ are the 2nd order Smoluchowski-derived bimolecular collision rate, the probability that both chains have helix formed at the moment of collision, and the success frequency for each collision, respectively (Karplus and Weaver 1994; Myers and Oas 1999). The model predicts a maximal folding rate that is only 1/25 of the observed value at 35°C. With a collision rate of $k_2 = 2.8 \times 10^9$ M^{-1} sec^{-1} and a 2.4% probability for the formation of a single turn of helix, the predicted folding rate constant is $k_f = 1.6 \times 10^6$ M^{-1} s^{-1}. This calculated rate is an upper bound, because every collision is assumed to be productive ($\theta = 1$). The discrepancy increases to $k^{observed} / k^{predicted} > 100$, if either the success frequency includes the probability that the two hydrophobic faces collide in the correct orientation, or the amount of pre-collision helical structure is increased to two turns – the minimum we envision necessary for a chain to be considered helical.

The D-C Model considers the collision of two preformed helices to be rate-limiting and common throughout the ensemble of conformations comprising the folding transition state. Kinetic amide isotope measurements described below indicate that 49 ± 2% of the helical content is formed in the transition state. If this amount of helical structure is present at the moment of collision, the observed rate is 400-fold faster than can be accommodated by the Model, even assuming $\theta = 1$.

Each chain has a net charge of +2, and no negatively charged residues. Thus, electrostatic considerations, which can enhance bimolecular rates for oppositely

charged molecules ("electrostatic guidance" Getzoff et al. 1992), would only further reduce the collision rate for GCN4-E_9G_4. Thus, the upper bound for the D-C Model is at least an order of magnitude slower than the observed folding rate, and more reasonable estimates produce a much more drastic discrepancy.

The D-C model also predicts the wrong temperature dependence for folding rates (Fig. 2.3 *b*). A reduction in temperature from 35 to 5° C increases the helical content in isolated chains by two-fold, or a four-fold increase in P_{helix}^2. At the same time, the second order collision frequency, k_2, is halved due to a proportionate increase in solvent viscosity. These opposing effects combine to produce a net two-fold increase in k_f for the 30 °C reduction in temperature. In fact, a four-fold decrease is observed. Hence, folding cannot be contingent upon pre-formed helical structure, unless the success frequency decreases by nearly an order of magnitude over the same temperature range.

Finally, the folding rate of GCN4-E_9G_4 is about four-fold faster than that of the wild-type sequence, despite having no more than one-tenth the intrinsic helicity across any given 4-10 residue stretch. Thus, the DC Model predicts, based on the decreased helicity and assuming a similar success frequency, that folding of the wild-type sequence should be 100-10,000 fold faster than GCN4-E_9G_4. This inconsistency, in conjunction with folding being too fast to be contingent upon pre-collision helical structure, and its temperature dependence being antithetical to that predicted by the D-C Model, leads us to conclude that the initial step of the dominant kinetic pathway is the collision of two unstructured chains.

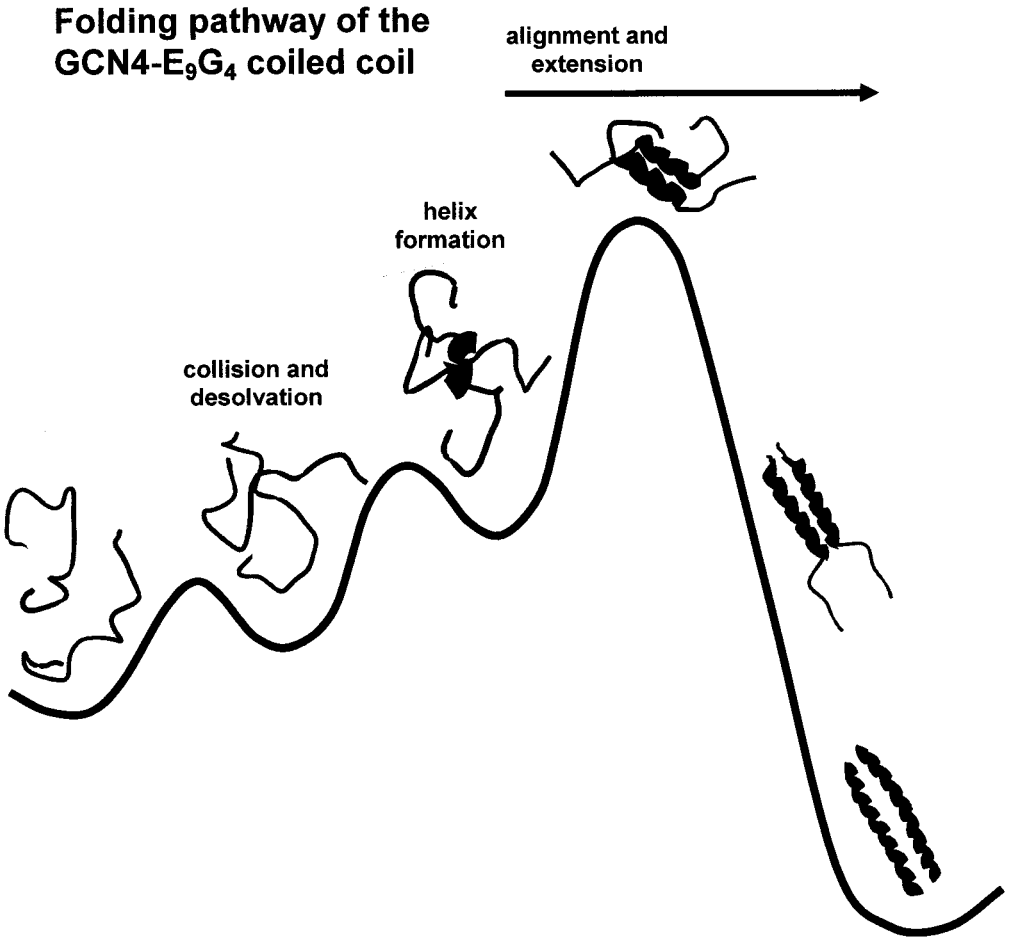
2.4.4 GCN4-E_9G_4 Folding pathway

Properties of the transition state point to a folding pathway which proceeds through a proportional build-up of secondary and tertiary structure. The denaturant dependence of k_f indicates that 61±6% of the total surface area is buried in the folding transition state (m_f/m^o) (Fig. 2.3 *d*). Likewise, the pH dependence of folding rates indicates that on average 46±11% of the "glutamic staples" are formed in the transition state ensemble (Fig. 2.3 *e*).

Backbone kinetic isotope effects provide a powerful method to directly monitor the helical hydrogen bond content in the transition state (Kentsis and Sosnick 1998; Krantz et al. 2000; Krantz et al. 2002). When conducted under the same bulk solvent conditions, the change in folding activation free energy upon backbone deuteration relative to the change in equilibrium stability, $\Delta\Delta G^{\ddagger}_{f}{}^{D\text{-to-}H}/\Delta\Delta G_{eq}{}^{D\text{-to-}H}$, is proportional to the fraction of hydrogen bonds formed. These quantities are obtained from the chevron plots for the protonated and deuterated coiled coil, conducted in a uniform 14.3% D_2O solvent condition (Fig. 2.3 *d*). Accordingly, we find that 49 ± 2% of helical hydrogen bonds are formed in the transition state.

The similarity in the degree of surface burial, and the fraction of staples and hydrogen bonds formed in the transition state produces the following picture for the folding pathway of GCN4-E_9G_4 (Fig. 2.4). Chain collision results in hydrophobic contacts that partially desolvate the polypeptide backbone. Exclusion of water molecules induces hydrogen bond formation and small helical regions develop at points of contact. These helical elements are then stabilized in the native register by

Figure 2.4 *Folding pathway of a designed coiled coil.* Unstructured chains collide which results in partial desolvation of the polypeptide backbone. Intra-chain hydrogen bonds form in response to the loss of solvent hydrogen-bond partners, forming isolated helices at the point of contact. Tertiary contacts stabilize these helices, bring them into alignment, and further desolvate the backbone, thereby promoting the extension of helical structure incrementally up to the half-helical transition state, from which folding is energetically downhill to the native structure.


Folding pathway of the
GCN4-E9G4 coiled coil
alignment and
extension
helix
formation
collision and
desolvation

tertiary interactions, both from hydrophobic residues at the interface, as well as the glutamic staples. This process continues incrementally up to the transition state.

2.4.5 The role of helicity in folding

The incompatibility of GCN4-E_9G_4 folding with the D-C Model may lead to the speculation that its folding behavior is somehow unique. However, its folding properties are very similar to that of its parent (Moran et al. 1999), as well as most similarly sized proteins. At the transition state, the degree of surface burial is typical, both in relative and absolute terms, while the same one-to-one correspondence between surface burial and hydrogen bond formation is found for a dozen other proteins (Krantz et al. 2000; Krantz et al. 2002). Thus, all the observed properties of the folding transition state of the designed coiled coil are typical of globular proteins.

The collision-first route may be likened to a rudimentary mechanism that sets a basal level on folding rates for any protein. Unexpectedly, the folding of our construct with low helical content demonstrates that this option can be very fast. It may be considered that the speed is due in part to the glutamic acids. However, mutational studies of the GCN4-p1' parent have shown that folding can begin from the least helical region, contrary to the expectation of the D-C model (Moran et al. 1999) and this parent molecule lacks these staples and even has regions of high helical propensity. Nevertheless, some proteins with very high helical propensity are likely to have regions structured at the moment of collision, and obey the D-C model (Mayor

et al. 2003). Generally, which of these two limiting scenarios dominates may vary for proteins with limited amounts of residual structure.

The rapid folding of our construct indicates that an unstructured encounter complex can still be very productive, as recently proposed (Shoemaker, Portman and Wolynes 2000). In GCN4-E_9G_4, the even distribution of glutamic staples and other tertiary interactions further ensures the productivity of multiple folding routes. This mechanism may apply to growing number of "natively unfolded" proteins (Dunker et al. 2002), whose folding is coupled to binding (Dyson and Wright 2002).

What, then, is the required role of helicity along the pathway? Many experiments have demonstrated that enhancing intrinsic helicity accelerates folding. Such strengthening stabilizes all conformations containing helix, including the transition state (Fig. 1). Furthermore, kinetic isotope measurements indicate that hydrogen bond formation is intimately linked with surface burial at the transition state (Krantz et al. 2000; Krantz et al. 2002). Thus, secondary structure is critical for the transition state to be thermodynamically accessible, although the amount of secondary structure present before collision may be minimal, as found for GCN4-E_9G_4.

CHAPTER 3

INVESTIGATING BARRIER-FREE PROTEIN FOLDING USING NATIVE-STATE HYDROGEN EXCHANGE METHODS

All the data discussed in this chapter is my own, but I wish to express my gratitude for the contributions of many others. In addition to those acknowledged in the beginning of this volume, I thank Dr. Liam Moran for preliminary work, and both Dr. Josh Kurutz of the Biological Sciences Division NMR facility at the University of Chicago, and Dr. Shohei Koide, assistant professor of the same institution, for their aid in obtaining and implementing the CLEANEX pulse sequence.

3.1 Abstract

Folding experiments are conducted to test whether the covalently cross-linked GCN4 coiled coil folds so fast that the process is no longer limited by a free energy barrier. This protein is very stable and topologically simple, needing merely to "zipper up" while having an extrapolated folding rate of $k_f \sim 2x10^5$ s^{-1}. These properties make it likely to attain the elusive "downhill folding" limit where a series of intermediates can be characterized. Near the denaturant melting transition where $k_f < 300$ s^{-1}, stopped-flow measurements reveal barrier-limited kinetics. To measure the ultra-fast kinetics in the absence of denaturant, we apply NMR and hydrogen exchange methods under EX1 and EX2 conditions to measure global stability along with folding rates. The stability and its denaturant dependence for the hydrogen bonds in the central part of protein equal the values calculated for whole molecule unfolding. Likewise, their closing and opening rates indicate that these most stable hydrogen bonds are broken and reformed in a single cooperative event representing the folding transition from the fully unfolded state to the native state. Additionally, folding rates for these hydrogen bonds agrees with the extrapolated barrier-limited rate. Therefore, even in the absence of denaturant, where $\Delta G_{eq} \sim 6$ kcal mol^{-1} and $\tau_f \sim 6$ μsec, folding remains cooperative and barrier limited. Given that this prime candidate for downhill folding fails to do exhibit such behavior, we propose that protein folding will remain barrier-limited for proteins that fold cooperatively.

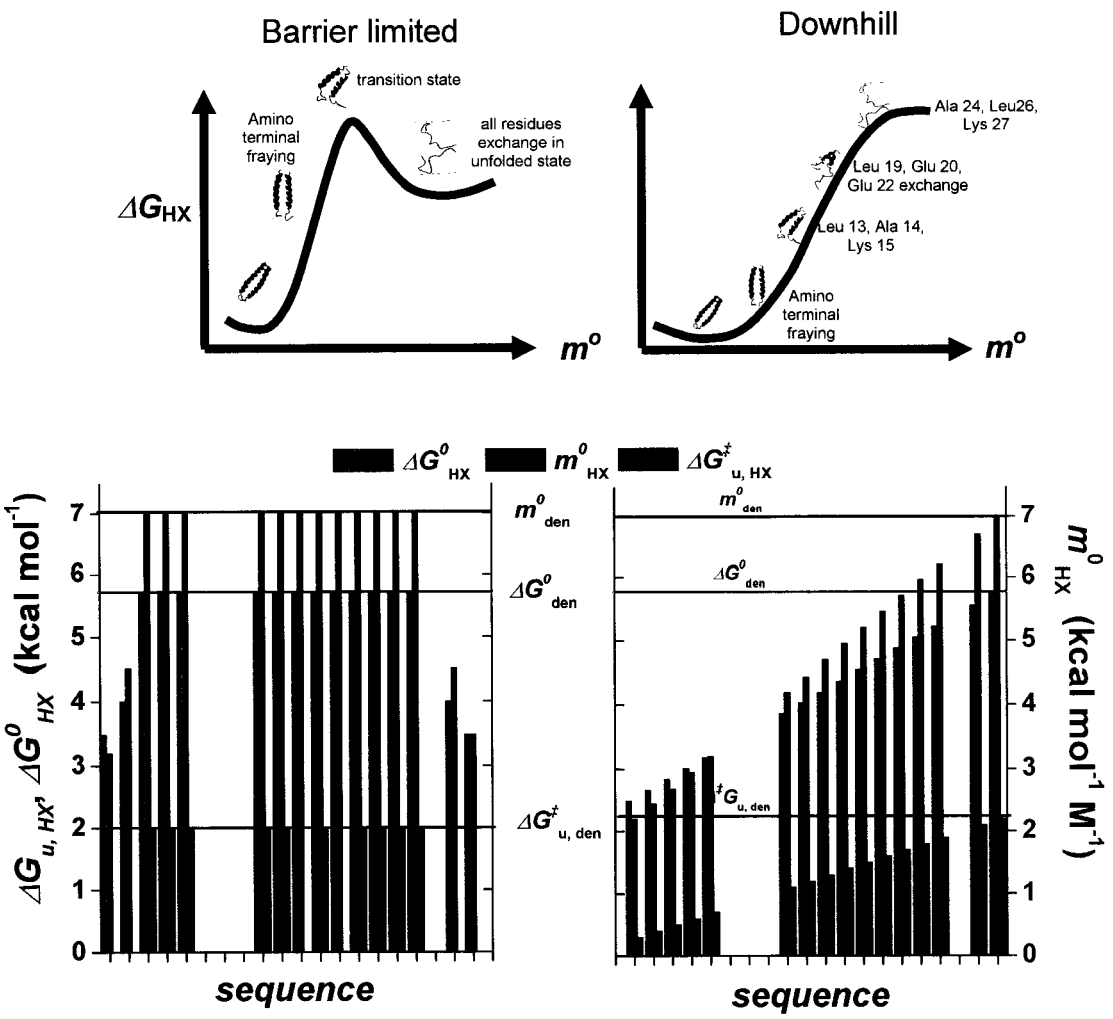
3.2 Introduction

The two-state approximation of protein folding, in which the unfolded and native states are separated by a single free energy barrier, and no other species accumulate to a significant degree, appears adequate for most small proteins (Jackson 1998; Krantz et al. 2002). It is nevertheless unsatisfying, because it generally precludes identifying the individual steps involved in the folding process. To overcome this problem, experiments (Sabelko, Ervin and Gruebele 1999; Garcia-Mira et al. 2002; Cavalli et al. 2003; Yang and Gruebele 2003; Kubelka, Hofrichter and Eaton 2004) have pursued the elusive theoretical prediction of downhill folding (Bryngelson et al. 1995; Socci, Onuchic and Wolynes 1998; Eaton 1999) (Fig. 3.1). If attainable, a downhill energy surface may allow the experimental identification of intermediate states along the folding pathways that must exist, but as yet have largely been "seen" only in computer simulations.

Downhill folding behavior is predicted to occur when folding rates approach the value of the attempt frequency of the barrier crossing process, $k_f = k_{attempt} e^{\Delta G^{\ddagger}/RT} \sim k_{attempt}$. Here, the barrier height makes no contribution to the rate, and $\Delta G^{\ddagger} \sim 0$. This limit may be likened to extreme Hammond behavior in which the folding transition state moves so far to the starting condition in response to heightened stability, that the transition state coincides with the unfolded state.

A barrier-free reaction must be extremely rapid, but how fast is fast enough? Estimates from measured rates of helix, loop, and hairpin formation, reaction-rate and polymer collapse theories, and folding simulations suggest that the minimum possible

Figure 3.1 *Barrier-limited versus downhill folding.* For a two-state folding process (*left panels*), a free energy barrier separates the folded and unfolded states whereas in downhill folding (*right panels*), no such barrier exists. Native state hydrogen exchange measures the free energy and surface area exposed in opening events where amide hydrogen bonds are broken. On a barrier-limited pathway (lower left panel), positions on the native side of the barrier exchange at lower free energies and surface area exposure, but all other positions exchange from the unfolded state with the free energy (red), surface area exposure (blue), and rate of hydrogen bond breakage (unfolding free energies, green) as global unfolding. In downhill folding (lower right panel) a ladder of openings with increasing free energy, surface area exposure and unfolding free energies starts from the amino terminus.


Barrier limited
Downhill
transition state
Amino terminal fraying
all residues exchange in unfolded state
ΔG_HX
m^o
Ala 24, Leu26, Lys 27
Leu 19, Glu 20, Glu 22 exchange
Leu 13, Ala 14, Lys 15
Amino terminal fraying
m^o
ΔG^o_HX
m^o_HX
ΔG‡_u, HX
m^o_den
ΔG^o_den
ΔG‡_u, den
ΔG_u, HX, ΔG^o_HX (kcal mol^-1)
m^o_HX (kcal mol^-1 M^-1)
sequence
sequence

folding time constant is on the order of one μsec (Kubelka, Hofrichter and Eaton 2004). Because of the increased ruggedness of the energy landscape, downhill folders of high stability would have proportionally slower folding rates, as would those with longer primary sequences and greater β-sheet content. After correcting for length and stability, thirteen proteins with predicted barrier-free time constants less than 100 μsec have been identified as potential downhill folders (Kubelka, Hofrichter and Eaton 2004).

The fastest folding and most stable of these proteins is the covalently cross-linked variant of the dimeric yeast transcription factor GCN4 coiled coil. Its extrapolated folding time constant is τ_f ~10 μsec, and less than 1 μsec when normalized for the molecule's high stability (Kubelka, Hofrichter and Eaton 2004). Intuition suggests that this fast folding rate may be due to a simple topology that only requires the unfolded helices to "zipper-up" into the native structure. Taken together, these qualities place it among the most likely candidates for downhill folding.

Here, we investigate whether this prime candidate does, in fact, exhibit such behavior. By combining traditional stopped-flow and denaturation experiments with native state hydrogen exchange, we find that even in the absence of denaturant, where folding rates approach those predicted for downhill behavior and stability is 6 kcal mol^{-1}, folding remains cooperative and barrier-limited.

3.3 Materials and methods

Peptides. GCN4p2C was synthesized as described in (Moran et al. 1999) and includes a C-terminal extension consisting of two glycines and a cysteine (Ac-RMKQLEGKVEELLAKNWHLENEVARLKKLVGERGGC). Bubbling of oxygen for 1 hr at pH 9 in 20 mM borate buffer crosslinked the two monomers at the C-terminal extension, producing monomeric GCN4p2C. To facilitate HX measurements using saturation transfer methods, D7G and S14A substitutions which accelerate unfolding and folding rates, respectively (Moran et al. 1999; Krantz and Sosnick 2001) were also included. A Y17W substitution provides a fluorescent folding probe. The sequence is ^{15}N labeled at positions Leu 12, Leu 13, Ala 24, and Leu 26, to reduce overlap in NMR spectra. Product identity and purity was confirmed by mass spectroscopy. All experiments were conducted in 0.2M NaCl, 10 mM phosphate, 10 mM borate buffer at 40 C and the indicated pH. Peptide concentrations were determined using $\varepsilon_{280nm}^{1\ cm}$=11,200 $M^{-1}cm^{-1}$.

Equilibrium measurements. Experiments were performed with a Jasco 715 spectropolarimeter, in equilibrium mode. CD measurements utilized a pathlength of 1 cm, and fluorescence spectroscopy used excitation and emission wavelengths of 280-290 nm and 300-400 nm, respectively.

Stopped-flow spectroscopy. Rapid mixing fluorescence experiments used a Biologic SFM-400 stopped-flow apparatus connected via a fiber optic cable to a PTI A101 arc

lamp. Fluorescence spectroscopy used excitation and emission wavelengths of 280-290 nm and 300-400 nm, respectively.

Hydrogen exchange theory. Spontaneous structural fluctuations break and reform hydrogen bonds at an unfolding rate k_u and a folding rate k_f. Once broken, and when the amide is exposed to solvent above pH ~ 4, HX is base catalyzed and occurs according to (Englander and Kallenbach 1984; Englander et al. 1997):

$$\text{folded} \underset{k_f}{\overset{k_u}{\rightleftharpoons}} \text{unfolded} \xrightarrow{k_{int}[OH^-]} \text{exchange} \qquad (3.3.1)$$

where k_{int} is the intrinsic exchange rate, which depends upon the identity of the amino acid, its amino-terminal neighbor, and temperature, as previously calibrated (Bai et al. 1993; Connelly et al. 1993) and found to be accurate to within a factor of two. Provided k_f and k_u are pH independent, the observed exchange rate is given by the expression

$$k_{HX} = \frac{k_u k_{int}[OH^-]}{k_u + k_f + k_{int}[OH^-]} \qquad (3.3.2)$$

For a stable protein in the so-called EX1 limit where $k_u < k_f << k_{int}$ [OH-], every unfolding event leads to exchange, and the observed HX rate reduces to $k_{HX} = k_u$. In the EX2 limit where $k_f >> k_{int}$ [OH^-], the protein establishes an equilibrium between the two states. The HX rate reduces to the fraction of time the protein is exchange competent multiplied by the rate of exchange in that state, $k_{HX} = (k_u/k_f)\, k_{int}$ [OH^-] = $K_{eq}\, k_{int}$ [OH^-]. Because K_{eq} equates to the unfolding equilibrium constant, the

free energy required to break the involved hydrogen bond is accessible from measurements taken in the EX2 regime, $\Delta G_{HX} = -RT \ln (k_{HX}/(k_{int} [OH^-]))$.

At lower pH where the EX2 condition occurs, k_{int} $[OH^-]$ and therefore k_{HX}, increases logarithmically with pH. At higher pH, where exchange is governed by the EX1 mechanism, k_{HX} is independent of k_{int} $[OH^-]$ and reaches a maximum equal to k_u. Measuring the HX rate over a range of pH's spanning the EX2 to EX1 behavior enables calculation of k_f ,k_u, and K_{eq}, according to Eq. 3.3.2, provided these parameters are pH independent.

NMR spectroscopy. Spectra were taken on a Varian Unity Inova 600 MHz spectrometer at 0.5-2 mM protein concentration. Amide proton assignment has been made previously (Goodman and Kim 1991), and were confirmed with 2D NOESY and TOCSY spectra. Exchange rates were determined with 1D-spectra using two sequences. Above pH 9.5, where exchange occurs faster than 0.5 s^{-1}, rates were measured using the CLEANEX pulse sequence (Hwang, van Zijl and Mori 1998). This sequence acts by dephasing the protein signal while maintaining coherence of solvent protons. As exchange occurs, amide protein peaks reappear and the heights of these peaks after exchange periods of different lengths are subjected to initial slope analysis to determine k_{HX}.

Accuracy of the CLEANEX sequence is limited to rates greater than ~0.5 s^{-1}. Below pH 9.5, EX2 exchange often occurs more slowly, so these rates were measured by dephasing solvent protons and fitting the resulting exponential decrease in the amide peaks as they undergo chemical exchange with solvent proteins as described in

(Bai et al. 1993; Connelly et al. 1993). Manipulation of the HSQC phase cycling in this second sequence allowed relaxation curves of ^{15}N- and ^{14}N- protons to be obtained independently. However, unlike the CLEANEX sequence, this method does not account for possible NOE contributions to relaxation, which could result in a slight overestimate of k_{HX}, and hence underestimate of ΔG_{HX} for residues that have an NOE with a solvent molecule.

3.4 Results and discussion.

3.4.1 Standard equilibrium and kinetic measurements.

A combination of equilibrium denaturation and chevron analysis identifies the folding of GCN4p2C as cooperative and barrier-limited under conditions near the denaturant melting transition at 40 °C.. The results of denaturation melts (Fig. 3.2 ***a***) and standard (msec) stopped-flow experiments (Fig. 3.2 ***c***) on GCN4p2C are analyzed according two-state folding transition, with ΔG_{eq} , $\Delta G^{\ddagger}_{f}$, and $\Delta G^{\ddagger}_{u}$ being linearly dependent upon denaturant concentration (Moran et al. 1999; Krantz and Sosnick 2001).

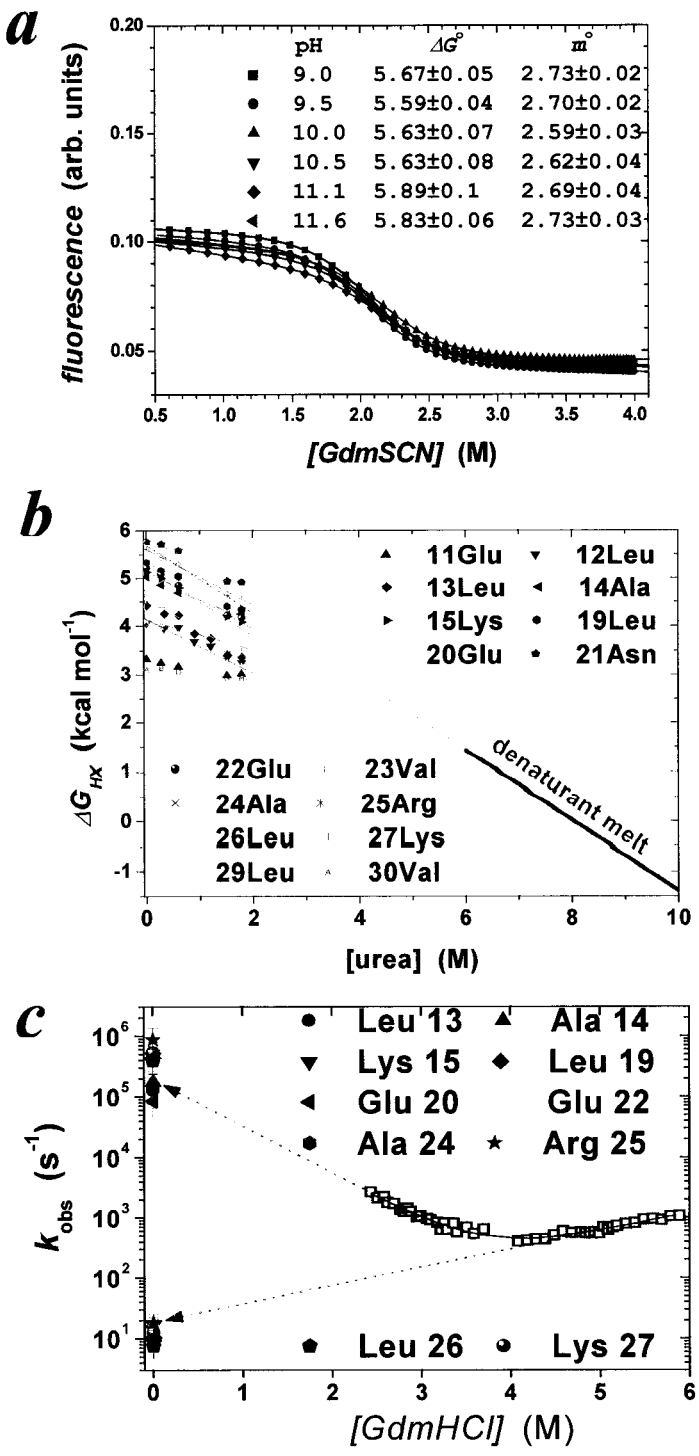
$$\Delta G^{o}([\text{denat}]) = \Delta G^{o}_{H2O} - m^{o}\,[\text{denat}] \qquad (3.4.1)$$

$$\Delta G^{\ddagger}_{f}([\text{denat}]) = -\,\text{RT ln}\, k_f^{H2O} - m_f\,[\text{denat}] + \text{Constant} \qquad (3.4.2)$$

$$\Delta G^{\ddagger}_{u}([\text{denat}]) = -\,\text{RT ln}\, k_u^{H2O} - m_u\,[\text{denat}] + \text{Constant} \qquad (3.4.3)$$

The slopes, m^{o}, m_f and m_u represent the difference in the amount of denaturant sensitive surface area buried between the initial and final state for the transition under consideration. The folding and unfolding rates extrapolated to zero denaturant are

Figure 3.2 *Equilibrium and kinetic measurements.* ***(a)*** Equilibrium free energy profiles measuring fluorescence as a function of gdmSCN concentration at six different pH conditions confirm that protein stability is invariant between pH 9 and 11.6. ***(b)*** Free energies of exchange for 16 amide protons were determined at five to seven urea concentrations below 2 M. The low denaturant activity of urea requires that denaturation profiles with this denaturant be taken in a constant background of 1.2 M gdmHCl in order to completely unfold the protein. In this solvent, at pH 10.5, ΔG^0_{den} = 3.84 ± 0.1 kcal mol^{-1}, consistent with that determined in gdmHCl after correcting for the background gdmHCl concentration, and m^0_{den} = 0.71 ± 0.03 kcal mol^{-1} M^{-1}. For 12 amide protons, $\Delta G^0_{HX} = \Delta G^0_{den}$ and $m^0_{HX} = m^0_{den}$. ***(c)*** Unfolded protein in high concentrations of gdmHCl is mixed with low gdmHCl buffer or folded protein at low gdmHCl concentration is mixed with high gdmHCl buffer to determine the denaturant dependence of folding (open red squares) and unfolding (closed red squares) free energies. The respective slopes report on the degree of surface area burial during the folding process. When folding is kinetically two-state and barrier-limited, the ratio of the extrapolated folding (k_f = 1.7 X 10^6 s^{-1}) and unfolding rates (k_u = 18 s^{-1}) recapitulates the eqilibrium folding free energy and the difference in the slopes is equal to the denaturant dependence of the protein stability, $m^0=m_u-m_f$. Here, $\Delta G^0_{den} = \Delta G^0_{kin}$ (5.91 ± 0.07 and 5.7 ± 0.2 kcal mol^{-1}, respectively) and $m^0_{den}=m^0_{kin}$ (1.57 ± 0.02 and 1.52 ± 0.05 kcal mol^{-1} M^{-1}, respectively) confirming that folding is barrier-limited near the equilibrium transition.


a
pH ΔG° m°
9.0 5.67±0.05 2.73±0.02
9.5 5.59±0.04 2.70±0.02
10.0 5.63±0.07 2.59±0.03
10.5 5.63±0.08 2.62±0.04
11.1 5.89±0.1 2.69±0.04
11.6 5.83±0.06 2.73±0.03
fluorescence (arb. units)
[GdmSCN] (M)
b
11Glu 12Leu
13Leu 14Ala
15Lys 19Leu
20Glu 21Asn
22Glu 23Val
24Ala 25Arg
26Leu 27Lys
29Leu 30Val
denaturant melt
ΔG_HX (kcal mol^-1)
[urea] (M)
c
Leu 13 Ala 14
Lys 15 Leu 19
Glu 20 Glu 22
Ala 24 Arg 25
Leu 26 Lys 27
k_obs (s^-1)
[GdmHCl] (M)

k_f =170,000 s^{-1} and k_u = 18 s^{-1}. The fractional surface area burial at the energy barrier is $m_f/m^0 = 0.71$.

The equilibrium properties can be obtained from the kinetic parameters according to $\Delta G^o_{H2O} = -RT \ln k_f^{H2O} / k_u^{H2O}$ and $m^o = m_u - m_f$. The equilibrium values of the stability and its denaturant dependence are in agreement with the kinetically determined values (ΔG^o = 5.91 ± 0.07 versus 5.7 ± 0.2 kcal mol^{-1} and m^o_{GdmHCl} = 1.57 ± 0.02 versus 1.52 ± 0.05 kcal mol^{-1} M^{-1}, respectively). This demonstration is widely accepted as a demonstration of the applicability of a two-state, barrier-limited folding scenario.

Fluorescence measurements at six different pHs between 9.0 and 11.6 (Fig. 3.2 ***a***) confirm that stability and surface burial do not vary significantly over this pH range (at high pH, solvent absorbance increases and accurate CD measurements could not be conducted). The insensitivity of ΔG_{eq} to pH simplifies the subsequent analysis of HX data taken over this pH range.

3.4.2 HX measurements

The equilibrium and kinetic measurements just described indicate that the coiled coil folds in a cooperative, two-state manner under conditions near the unfolding transition, approximately 3.5 M GdmHCl. Of central interest is whether this barrier-limited behavior persists even in the absence of denaturant, where the extrapolated stability and folding rate are 6 kcal mol^{-1} and k_f = 170,000 s^{-1}, respectively. These parameters along with k_u can be obtained under fully native

conditions by monitoring HX rates in the EX2 and EX1 limits (see Methods). The ability to obtain this information under these highly stable conditions is generally impossible with other rapid measurement techniques such as pressure or T-jumps where only a limited thermodynamic perturbation is obtainable. Furthermore, HX monitors individual amide protons, providing us with the stability and folding rate for about a dozen sites across the whole protein in which to test whether folding remains barrier limited.

The linear pH dependence of log k_{HX}, suggests that exchange is occurring in the EX2 regime below about pH 10 (Fig. 3.3 ***c*** and ***d***). In this regime, the observed HX rate for an amide proton is proportion to K_{eq} of the associated hydrogen bond. For the eleven resolvable residues between positions 13 and 27, their HX-determined stabilities ΔG_{HX} equal the ΔG^{o} for whole molecule unfolding (Fig. 3.4). Furthermore, measurements taken across 5-7 urea concentrations indicate that for these residues, $m^{o}_{HX} = m^{o}$ for the entire molecule. That is, these eleven hydrogen bonds exchange in a reaction with the same free energy and surface exposure as the entire N↔U reaction. Therefore, the exchange of these amide protons monitors the global unfolding process.

A complete analysis of the pH dependence provides the corresponding kinetic parameters. At pH ~ 11, HX shifts to the EX1 limit. Here, k_f is slower than k_{int} in the exposed state, so that every opening results in exchange, and $k_u=k_{HX}$. The transition from EX2 to EX1 occurs when $k_f \sim k_{int}$, which is different for each amide proton. Fitting the HX data taken from pH 10-12 to Eq. 2 for the globally exchanging amide

Figure3. 3 *Native state hydrogen exchange.* ***(a)*** Representative data taken during the first second of exchange (the inset shows the entire trajectory) below pH ~ 9.5 with the saturation transfer sequence and used to determine ΔG^{o}_{HX} and m^{o}_{HX}, here performed at pH 9.2 in the absence of denaturant. Because T_1 relaxation competes with exchange, trajectories are fitted to the equation: $I(t_{HX}) = I^0\ (k_{HX}/(k_{HX} + 1/T_1)\ \exp\{-(k_{HX} + 1/T_1)\ t_{HX}) + (1/[(k_{HX} + 1/T_1)T_1)]\}$, where k_{HX} is the observed exchange rate, T_1 is T_1 relaxation, t_{HX} is the exchange time, and I^0 is the peak intensity when $t_{HX} = 0$. ***(b)*** Representative data taken with the CLEANEX pulse sequence above pH ~ 9.5 and used to determine the pH dependence of HX, here performed at pH 11.6. The traces are subjected to initial slope analysis to determine k_{HX}. ***(c*** and ***d)*** The pH dependence of HX for 10 resolvable protons and their calculated intrinsic exchange rates are fitted to *eq.*2 to determine the kinetics of hydrogen bond formation for each amide hydrogen. The dotted diagonal line has a slope = 1, illustrating that below pH ~10, exchange occurs in the EX2 regime.

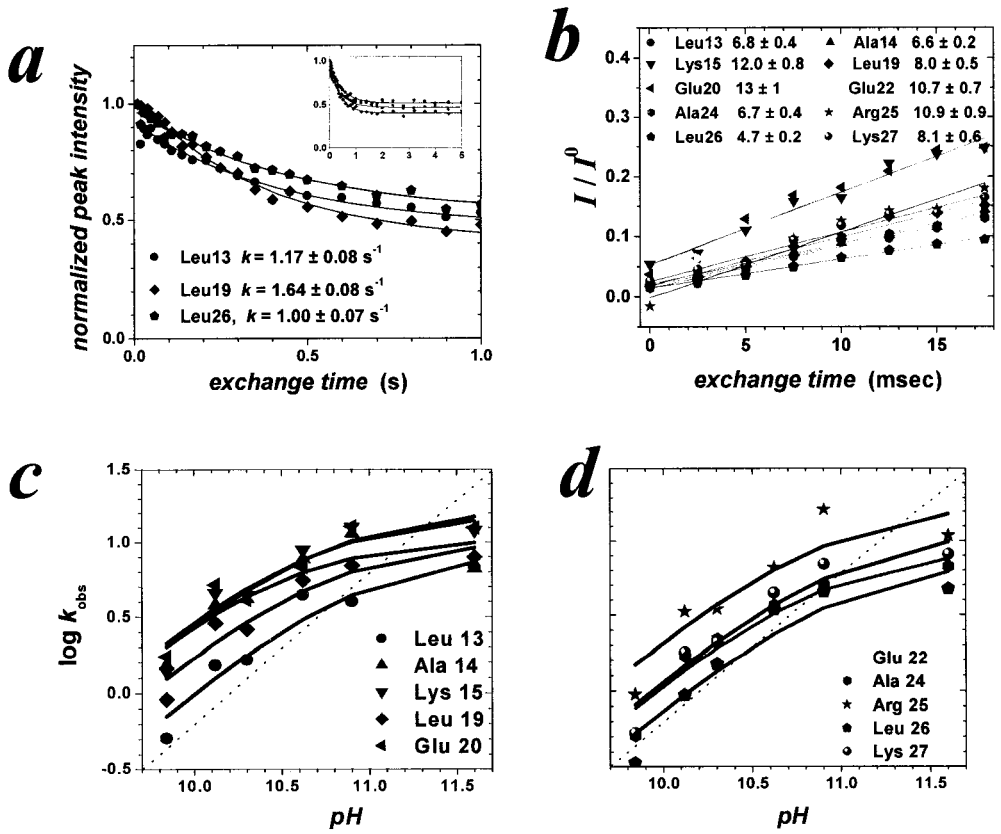
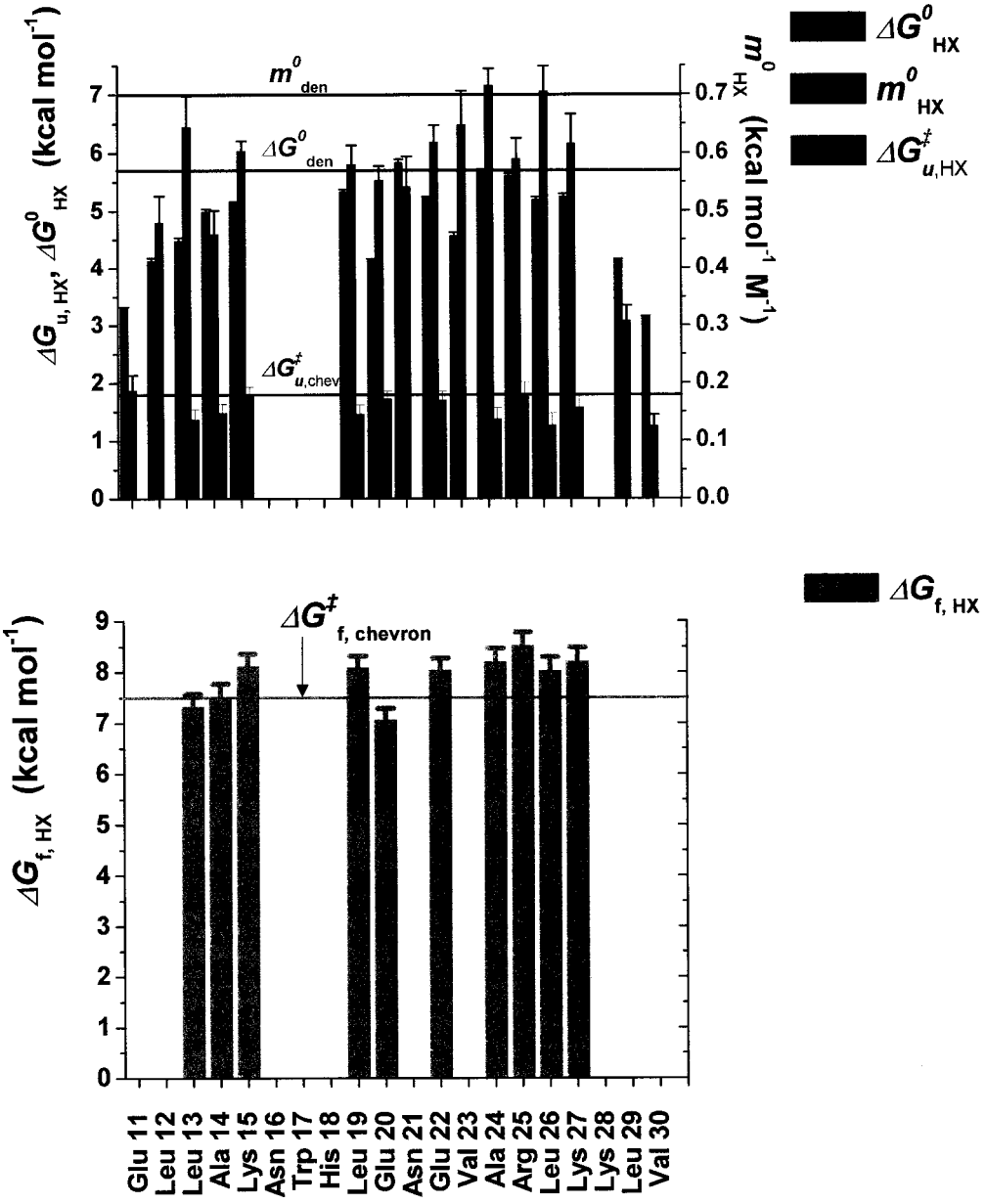

a
normalized peak intensity
exchange time (s)
Leu13 k = 1.17 ± 0.08 s-1
Leu19 k = 1.64 ± 0.08 s-1
Leu26, k = 1.00 ± 0.07 s-1
b
I / I0
exchange time (msec)
Leu13 6.8 ± 0.4
Ala14 6.6 ± 0.2
Lys15 12.0 ± 0.8
Leu19 8.0 ± 0.5
Glu20 13 ± 1
Glu22 10.7 ± 0.7
Ala24 6.7 ± 0.4
Arg25 10.9 ± 0.9
Leu26 4.7 ± 0.2
Lys27 8.1 ± 0.6
c
log kobs
pH
Leu 13
Ala 14
Lys 15
Leu 19
Glu 20
d
pH
Glu 22
Ala 24
Arg 25
Leu 26
Lys 27

Figure 3. 4 *μsec folding is barrier-limited.* HX data displayed in Fig 3.2***b*** and ***c*** are replotted for comparison to the barrier-limited and downhill scenarios shown in Fig 3.1. For 10 – 12 centrally located amide protons, $\Delta G^{o}_{HX} = \Delta G^{o}_{den}$, $m^{o}_{HX} = m^{o}_{den}$, $\Delta G_{u,HX} = \Delta G^{\ddagger}_{u}$, chevron, and $\Delta G^{\ddagger}_{f}$, HX $= \Delta G^{\ddagger}_{f,\ chevron}$, demonstrating that these residues exchange simultaneously in a global unfolding event. Cooperativity and agreement with measurements taken near the equilibrium midpoint, where folding is known to be two-state, indicate that the folding of this molecule remains barrier limited, even in the absence of denaturant. In addition, since roughly half the native amount of surface area burial is present in the transition state, incorporating about half of the 36 residues in each helix, it might be inferred that the involved residues comprise the stretch between Leu 13 and Lys 27.


ΔG^0_{HX}
m^0_{HX}
$\Delta G^{\ddagger}_{u,HX}$
$\Delta G_{f,HX}$
m^0_{den}
ΔG^0_{den}
$\Delta G^{\ddagger}_{u,chev}$
$\Delta G^{\ddagger}_{f,chevron}$
$\Delta G_{u,HX}$, ΔG^0_{HX} (kcal mol^{-1})
m^0_{HX} (kcal mol^{-1} M^{-1})
$\Delta G_{f,HX}$ (kcal mol^{-1})
Glu 11
Leu 12
Leu 13
Ala 14
Lys 15
Asn 16
Trp 17
His 18
Leu 19
Glu 20
Asn 21
Glu 22
Val 23
Ala 24
Arg 25
Leu 26
Lys 27
Lys 28
Leu 29
Val 30

protons indicates that each of the associated hydrogen bonds break and reform at the chevron-extrapolated unfolding and refolding rates of the entire molecule (Fig. 3.4).

3.4.3 Folding remains barrier-limited

The HX measurements indicate that the folding behavior in the absence of denaturant is qualitatively the same as at the melting transition where folding is barrier-limited. From the equilibrium perspective, the exchange of the centrally located amide protons occurs in the same, cooperative global unfolding process. Likewise, their closing and opening rates indicate that the most stable hydrogen bonds are broken and reformed in a single kinetic event representing the transition from the fully unfolded state to the native state. In addition, the folding and unfolded rates agrees with the barrier-limited rate (extrapolated to zero denaturant). Therefore, even in the absence of denaturant, where $\Delta G^{o} \sim 6$ kcal mol^{-1} and $\tau_f \sim 6$ μsec, folding remains barrier limited.

The conclusion is reinforced when the results for amide protons located at opposite ends of the protein are considered in light of the known folding pathway. Folding begins near the cross-linked carboxy-terminus and proceeds toward the amino-terminus (Krantz and Sosnick 2001). Were folding to occur on a downhill landscape, hydrogen bonds toward the carboxy-terminus would unfold last, and thus require the most energy and surface area exposure to exchange (Fig. 3.1). Likewise, their opening rates would be slower than for bonds located at the opposite end. Neither of these two behaviors associated with downhill folding are observed, further amplifying the barrier-limited nature of the GCN4-p2C folding reaction.

3.4.4 Denaturant does not alter folding behavior

Denaturants exert their destabilizing influence by preferentially interacting with the increased surface area found in the denatured state. For this reason, even moderate concentrations of denaturant may stabilize the unfolded state such that it is more stable than an intermediate position on the folding pathway, thereby creating an energetic barrier that does not exist under fully native conditions (Yang and Gruebele 2003). However, native state HX data taken in the absence of denaturant identifies folding rates consistent with the chevron extrapolation. Because the extrapolation assumes barrier-limited behavior with the same denaturant dependence as observed near the melting transition, the values of m_f , m_u and m^o measured in denaturant persist in its absence. Thus, not only is the barrier present in fully native conditions, the surface burial character of unfolded and transition states are unchanged.

3.4.5 Cooperativity and downhill folding

Barrier-limited folding near the melting transition has been observed for many variants of GCN4 coiled coil, and confirmed here for this construct. Cooperativity is a second, and often overlooked, property associated with barrier-limited behavior. Whenever a system can be described with two thermodynamic states, by necessity, a free energy barrier must separate the two wells. From this perspective, a demonstration of cooperativity even in equilibrium mode is sufficient to preclude downhill folding. Munoz and co-workers used this reasoning in reverse, arguing that by itself, the non-cooperative equilibrium folding of the E. coli BBL protein implies a downhill energy landscape (Garcia-Mira et al. 2002).

Other ultra-fast folding proteins (Zhu et al. 2003; Dimitriadis et al. 2004) have been successfully analyzed in terms of a two-state process (e.g. $K_{eq}=k_f/k_u$), implying that a cooperative, barrier crossing reaction is an appropriate description in these systems as well. Most single domain proteins fold cooperatively. Although larger proteins may have stable intermediates, a barrier still separates the unfolded and intermediates states, precluding the downhill folding scenario. Generally, proteins fold cooperatively, either at the global, domain, or subglobal level (Bai et al. 1995; Bai and Englander 1996; Rumbley et al. 2001). Therefore, downhill folding is likely to be limited to a few, atypical systems, such as "molecular rheostats" (Garcia-Mira et al. 2002) or possibly, designed proteins with an unusually high hydrophobic content (Scalley-Kim and Baker 2004).

3.4.6 Early folding steps are uphill

Cooperative folding requires that early folding steps be uphill in free energy relative to the unfolded state. The unfolded state, however, can depend on the solvent condition. For example, we found that burst phase CD signals observed upon a jump from high to low denaturant can result in backbone of some residues undergoing a mild average shift from a polyproline II geometry to a helical geometry (although not necessarily representing authentic helix formation which requires a stretch of four residues) (Jacob et al. 2004). This backbone relaxation is likely non-cooperative, and hence may appear downhill, although the individual residue shifts from one Ramachandran well to another may be barrier limited (Bolhuis, Dellago and Chandler

2000; Zaman et al. 2003). Even if the solvent adjustment results in authentic helix formation, helix initiation is uphill in free energy.

Downhill folding requires conditions where initial chain-chain interactions are stronger than those between the chain and solvent. For two-state systems satisfying the chevron criteria $m^o = m_u\text{-}m_f$, no denaturant sensitive surface is buried prior to the (sole) kinetic barrier. This result implies that no additional protein-protein contacts are formed in the jump to native conditions. Consistently, we found using small-angle scattering that two small proteins, Ubiquitin and ctAcp, do not undergo any measurable collapse upon a jump from high to low denaturant (Jacob et al. 2004). This behavior is echoed in studies of non-folding versions of lysozyme (Hoshino et al. 1997) and RNase A (J. Jacob and T. Sosnick, unpublished), created by reduction of their four disulfide bonds. Therefore, protein-protein interactions are weak compared to those with solvent early in the folding process.

It is appreciated that hydrophobic interactions are enhanced upon the shift to aqueous conditions. Yet, such contacts are largely inhibited by the loss of conformational entropy, both backbone and side-chain, and the desolvation of main chain polar groups. Empirically, water is a sufficiently good solvent that generic intra-chain contacts do not outweigh protein-solvent interactions. Based upon these and other considerations, we proposed that formation of a *stable* collapsed species is the intrinsic rate-limiting step in protein folding (Sosnick, Mayne and Englander 1996).

Certainly, these issues related to the early folding steps being uphill are only part of the origin of cooperativity, but they are the portions that pertain to the likelihood of observing downhill folding.

3.5 Conclusion

With its simple topology, high stability, and extremely fast folding rate, GCN4p2C is one of the most likely candidates for downhill folding. Nevertheless, we have found that even in the absence of denaturant, its folding behavior retains all the barrier-limited characteristics observed in moderate denaturant concentrations. While faster-folding proteins may yet be found and merit consideration as candidate downhill folders, the results displayed herein suggest that the search may prove difficult in biologically relevant systems. Tautologically, proteins which fold cooperatively cannot do so in a downhill manner. Nearly all single-domain proteins and subunits of larger proteins fold cooperatively, either globally or in parts, and therefore, are barrier limited. Downhill folding requires that initial protein-protein interactions be stable. With naturally-occurring proteins, sufficiently strong interactions may be possible only in unusual solvent conditions, such as in sodium sulfate; however, such conditions also are likely to produce aggregation and non-native species.

CHAPTER 4

CONCLUSIONS AND FUTURE DIRECTIONS

4.1 Summary of conclusions

This thesis investigates early folding events and behavior in the ultra-fast limit. Its focus incorporates both the classical view's pathway analogy and new view's preference for statistical energy landscapes. Folding models and scenarios widely accepted by adherents of both views are tested and found lacking.

4.1.1 Collision between unstructured chains can induce folding

The diffusion-collision model was proposed by Karplus and Weaver in 1976 as a kinetic mechanism by which unfolded proteins can exhaustively search conformational space, and yet quickly find the native fold. A long protein chain is treated as a series of much shorter segments (microdomains) which transiently form

their native secondary structure very quickly, and collision events between these preformed units trigger a successful folding event. It is an intuitively attractive folding mechanism with wide general acceptance, but difficult to test because the two-state folding behavior of most small proteins precludes the identification of intermediates, and hence the order of events, along the folding pathway.

The protein used in this investigation was engineered from the GCN4 coiled coil template. It possesses so little intrinsic helicity that barely 1 out of every 2000 collision events occurs between molecules with even a singe turn of helix at 35 °C. When combined with the bi-molecular collision rate calculated from Smoluchowski's formalism, the maximum folding rate consistent with this model can be no greater than 1.6×10^6 M^{-1} s^{-1} and much less if orientational factors and more reasonable amounts of helix are included. The observed rate exceeds these predictions by at least 25- and as much as 400-fold, yet the surface area burial and helical content present in the folding transition state is half that of the native state, just as in the parent molecule. Thus, folding proceeds through proportional buildup of secondary and tertiary interactions, rather than from preformed units of secondary structure.

These results demonstrate that collisions between unstructured chains can lead to successful folding, and secondary structure forms concurrently with surface burial, at least for this system. Unstructured collision events followed by folding provide a rudimentary mechanism available to all molecules and, as evidenced here, is potentially very fast. Folding from an unstructured encounter complex may have increased efficiency and robustness because it allows passage through the transition state via a wide range of pathways starting from multiple contact points. Thus, even

in systems with significant amounts of residual structure in the unfolded state, collisions between unstructured chains, though rare, may still be productive. If so, pre-collision structure may be ancillary to folding, even when present.

4.1.2 The μsec folding of a stable protein is barrier-limited under native conditions

Treatment of the protein folding question using statistical energy landscapes rather than a classical approach based on transition state theory have grown in popularity, especially among theoreticians. By this analogy, Levinthal's simplistic calculation yields a landscape that resembles the flat green of a golf course, where all possible conformations are of equal free energy and equally likely to be populated, rendering the cup impossible to find. A more accurate calculation produces a landscape of hills and valleys, the deeper of which are much more likely to be populated. The free energy landscape of some proteins may be such that a route from the unfolded to the native state may proceed through successively deeper valleys and encounter no hills of higher elevation than the unfolded state. This scenario describes downhill folding.

The study in chapter 3 investigated the downhill folding scenario using a protein whose topology, folding rate, and stability place it among the most likely candidates for downhill folding. The protein is subjected to a combination of traditional stop-flow experiments and native state hydrogen exchange. The stabilities, denaturant dependence of those stabilities, and folding rates measured for at least ten amide protons in the absence of denaturant with native state hydrogen exchange are self-consistent, and match the values calculated for the molecule as a whole obtained

at moderate denaturant concentrations. Thus, folding behavior in the absence of denaturant is qualitatively the same as at the melting transition where folding is barrier-limited.

These results demonstrate that folding of this molecule is barrier limited, even under conditions where downhill folding would be expected to occur. While faster folding and more stable proteins may yet be found, it seems unlikely in biologically relevant molecules because most, if not all, fold cooperatively – a second, and often overlooked quality that signals barrier limited behavior. Cooperativity, a hallmark of protein folding, and many other biological processes, is very difficult to model theoretically, and was not in the original models that generated downhill folding.

4.2 Future Directions

During the course of the two investigations described herein, properties of the molecules utilized suggest three future studies that could address as yet unanswered questions.

4.2.1 Success frequency of collisions

The diffusion-collision equation includes a poorly understood parameter called the success frequency (θ, in eq. 1.6.1), which represents the fraction of collisions between helical structures that progress to the native fold. This parameter also factors into the association rate of any two biomolecules, making determination of its value of general interest in biological recognition. Orientation of the two helices relative to each other are believed primarily responsible for this factor being less than

unity, and calculated estimates suggest it is related to the fraction of molecular surface area buried at the interface, but no true measurement exists.

Mutation of the GCN4-E_9G_4 to enable formation of two or more lactam (Gong et al. 1997) or dichloro acetone (A. Shandiz, unpublished data) bridges, between i and $i + 4$ of the helix could freeze each monomer into a helical geometry, making $P^2 = 1$. The bimolecular collision rate is known, and folding rates could be measured with a jump from high pH, where the negatively charged monomers repel each other, to low pH, where they can collide and fold. This leaves θ as the only unknown quantity, and thus easily calculated.

4.2.2 Size vs. orientation in the folding transition state

Based on an assessment of amide NOEs, native CD, and surface burial, the D7G mutant of GCN4 used in chapter 3 appears to have lost helical structure relative to wild-type in the two most N-terminal turns. However, the fraction of surface area burial present in the transition state increases from about 0.5 for wild-type to 0.7 for the mutant so the total amount of structure present in the transition state appears unchanged. If that structure extends from the crosslink as expected, it includes the intermolecular hydrogen bond between Asn 16 residues of each helix.

These observations suggest that the folding transition state of GCN4 requires either a specific amount of hydrophobic surface area burial to be energetically accessible, or that the intermolecular contact between Asn 16 residues must form to assure proper alignment of the helices. Movement of this contact to position 23 would have no effect on the fractional surface area if, as in the former case, the defining

characteristic of the transition state is energetic in nature. If, however, it requires only the structural alignment of the two helices, then the fractional surface area burial present at the transition state would be significantly lower in this mutant.

4.2.3 Design of a downhill folder

To date, most of the focus of downhill folding has been on molecules of low stability, because they are amenable to direct rate measurements through ensemble measurements such as laser temperature jumps. However, the theoretical literature that describes downhill folding emphasizes the stability of the system. Its occurrence "would require a huge thermal driving force" (Bryngelson et al. 1995). Despite possessing high stability and the folding rate predicted for downhill folding, the crosslinked GCN4 construct still folds in a barrier-limited manner, casting doubt upon hopes that any naturally occurring protein will fold in a downhill fashion. In light of the new appreciation for cooperativity as a necessary consequence of barrier-limited behavior, this is perhaps not surprising. Cooperativity seems likely to aid in the avoidance of mis-folding and aggregation, and may therefore have been selected for by evolution.

For these reasons, downhill folding seems more likely to be observed in a designed protein, and the GCN4-E_9G_4 construct utilized in chapter 2 may provide guidance in this endeavor. The enhanced tertiary interactions provided by the protonated glutamic acids can overcome the destabilizing effects of eight glycine residues (more than 10% of the sequence) to produce a fast folding and stable

molecule. Those same interactions may induce downhill folding in a sequence with more intrinsic stability.

REFERENCES

Akiyama, S., S. Takahashi, T. Kimura, K. Ishimori, I. Morishima, Y. Nishikawa and T. Fujisawa (2002). "Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering." Proc. Natl. Acad. Sci. U S A **99**(3): 1329-34.

Anfinsen, C. B. (1973). "Principles that govern the folding of protein chains." Science **181**: 223-230.

Arrington, C. B. and A. D. Robertson (2000). "Microsecond to minute dynamics revealed by EX1-type hydrogen exchange at nearly every backbone hydrogen bond in a native protein." J. Mol. Biol. **296**(5): 1307-17.

Bai, Y. and S. W. Englander (1996). "Future directions in folding: the multi-state nature of protein structure." Proteins **24**(2): 145-51.

Bai, Y., J. S. Milne, L. Mayne and S. W. Englander (1993). "Primary structure effects on peptide group hydrogen exchange." Proteins **17**(1): 75-86.

Bai, Y., T. R. Sosnick, L. Mayne and S. W. Englander (1995). "Protein folding intermediates studied by native state hydrogen exchange." Science **269**: 192-197.

Bhattacharyya, R. P. and T. R. Sosnick (1999). "Viscosity dependence of the folding kinetics of a dimeric and monomeric coiled coil." Biochemistry **38**(8): 2601-9.

Bhuyan, A. K. and J. B. Udgaonkar (2001). "Folding of horse cytochrome c in the reduced state." J Mol Biol **312**(5): 1135-60.

Bieri, O., J. Wirz, B. Hellrung, M. Schutkowski, M. Drewello and T. Kiefhaber (1999). "The speed limit for protein folding measured by triplet-triplet energy transfer." Proc. Natl. Acad. Sci. USA **96**(17): 9597-601.

Bolhuis, P. G., C. Dellago and D. Chandler (2000). "Reaction coordinates of biomolecular isomerization." Proc. Natl. Acad. Sci. USA **97**(11): 5877-82.

Bosshard, H. R., E. Durr, T. Hitz and I. Jelesarov (2001). "Energetics of Coiled Coil Folding: The Nature of the Transition States." Biochemistry **40**(12): 3544-3552.

Brandts, J. F., H. R. Halvorson and M. Brennan (1975). "Consideration of the possibility that the slow step in protein denaturation reactions is due to cis trans isomerism of proline residues." Biochemistry **14**: 4953-4963.

Bryngelson, J. D., J. N. Onuchic, N. D. Socci and P. G. Wolynes (1995). "Funnels, pathways, and the energy landscape of protein folding: a synthesis." Proteins **21**(3): 167-195.

Burton, R. E., G. S. Huang, M. A. Daugherty, P. W. Fullbright and T. G. Oas (1996). "Microsecond protein folding through a compact transition state." J Mol Biol **263**(2): 311-22.

Cavalli, A., U. Haberthur, E. Paci and A. Caflisch (2003). "Fast protein folding on downhill energy landscape." Protein Sci **12**(8): 1801-3.

Chen, L., K. O. Hodgson and S. Doniach (1996). "A lysozyme folding intermediate revealed by solution X-ray scattering." J. Mol. Biol. **261**(5): 658-71.

Connelly, G. P., Y. Bai, M.-F. Jeng, L. Mayne and S. W. Englander (1993). "Isotope effects in peptide group hydrogen exchange." Proteins **17**: 87-92.

Creighton, T. E. (1978). Prog. Biophys. Mol. Biol. **33**: 231-297.

Creighton, T. E. (1985). J. Phys. Chem. **89**: 2452-2459.

Creighton, T. E. (1988). "Toward a better understanding of protein folding pathways." Proc. Natl. Acad. Sci. USA **85**(14): 5082-6.

Dimitriadis, G., A. Drysdale, J. K. Myers, P. Arora, S. E. Radford, T. G. Oas and D. A. Smith (2004). "Microsecond folding dynamics of the F13W G29A mutant of the B domain of staphylococcal protein A by laser-induced temperature jump." Proc Natl Acad Sci U S A **101**(11): 3809-14.

Dunker, A. K., C. J. Brown, J. D. Lawson, L. M. Iakoucheva and Z. Obradovic (2002). "Intrinsic disorder and protein function." Biochemistry **41**(21): 6573-82.

Durr, E., I. Jelesarov and H. R. Bosshard (1999). "Extremely fast folding of a very stable leucine zipper with a strengthened hydrophobic core and lacking electrostatic interactions between helices." Biochemistry **38**(3): 870-880.

Dyson, H. J. and P. E. Wright (2002). "Coupling of folding and binding for unstructured proteins." Curr. Opin. Struct. Biol. **12**(1): 54-60.

Eaton, W. A. (1999). "Searching for "downhill scenarios" in protein folding." Proc. Natl. Acad. Sci. U S A **96**(11): 5897-9.

Eliezer, D., K. Chiba, H. Tsuruta, S. Doniach, K. O. Hodgson and H. Kihara (1993). "Evidence of an associative intermediate on the myoglobin refolding pathway." Biophys. J. **65**: 912-917.

Englander, S. W. and N. R. Kallenbach (1984). "Hydrogen exchange and structural dynamics of proteins and nucleic-acids." Q. Rev. Biophys. **16**: 521-655.

Englander, S. W., L. C. Mayne, Y. Bai and T. R. Sosnick (1997). "Hydrogen exchange: the modern legacy of Linderstrom-Lang." Protein Sci. **6**: 1101-1109.

Fersht, A. R., L. S. Itzhaki, N. F. elMasry, J. M. Matthews and D. E. Otzen (1994). "Single versus parallel pathways of protein folding and fractional formation of structure in the transition state." Proc. Natl. Acad. Sci. USA **91**(22): 10426-9.

Garcia, A. E. (2004). "Characterization of non-alpha helical conformations in Ala peptides." Polymer **45**(2): 669-676.

Garcia-Mira, M. M., M. Sadqi, N. Fischer, J. M. Sanchez-Ruiz and V. Munoz (2002). "Experimental identification of downhill protein folding." Science **298**(5601): 2191-5.

Garel, J.-R. and R. L. Baldwin (1975). "The heat-unfolded state of ribonuclease A is an equilibrium mixture of fast and slow refolding species." J. Mol. Biol.: 611-620.

Getzoff, E. D., D. E. Cabelli, C. L. Fisher, H. E. Parge, M. S. Viezzoli, L. Banci and R. A. Hallewell (1992). "Faster superoxide dismutase mutants designed by enhancing electrostatic guidance." Nature **358**(6384): 347-51.

Gladwin, S. T. and P. A. Evans (1996). "Structure of very early protein folding intermediates: new insights through a variant of hydrogen exchange labelling." Fold. Des. **1**(6): 407-17.

Goodman, E. M. and P. S. Kim (1991). "Periodicity of amide proton exchange rates in a coiled-coil leucine zipper peptide." Biochemistry **30**: 11615-11620.

Grantcharova, V. P., D. S. Riddle and D. Baker (2000). "Long-range order in the src SH3 folding transition state." Proc. Natl. Acad. Sci. U S A **97**(13): 7084-7089.

Hagen, S. J. (2003). "Exponential decay kinetics in "downhill" protein folding." Proteins **50**(1): 1-4.

Hagen, S. J. and W. A. Eaton (2000). "Two-state expansion and collapse of a polypeptide." J Mol Biol **301**(4): 1019-27.

Hoshino, M., Y. Hagihara, D. Hamada, M. Kataoka and Y. Goto (1997). "Trifluoroethanol-induced conformational transition of hen egg-white lysozyme studied by small-angle X-ray scattering." FEBS Lett. **416**(1): 72-6.

Huang, C. Y., Z. Getahun, Y. Zhu, J. W. Klemke, W. F. DeGrado and F. Gai (2002). "Helix formation via conformation diffusion search." Proc. Natl. Acad. Sci. U S A **99**(5): 2788-93.

Hwang, T. L., P. C. van Zijl and S. Mori (1998). "Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme." J. Biomol. NMR **11**(2): 221-6.

Ibarra-Molero, B., G. I. Makhatadze and C. R. Matthews (2001). "Mapping the energy surface for the folding reaction of the coiled-coil peptide GCN4-p1." Biochemistry **40**(3): 719-31.

Islam, S. A., M. Karplus and D. L. Weaver (2002). "Application of the diffusion-collision model to the folding of three-helix bundle proteins." J Mol Biol **318**(1): 199-215.

Jackson, S. E. (1998). "How do small single-domain proteins fold?" Fold. Des. **3**(4): R81-91.

Jackson, S. E. and A. R. Fersht (1991). "Folding of chymotrypsin inhibitor 2. 2. Influence of proline." Biochemistry **30**: 10436-10443.

Jacob, J., B. Krantz, R. S. Dothager, P. Thiyagarajan and T. R. Sosnick (2004). "Early Collapse is not an Obligate Step in Protein Folding." J. Mol. Biol. **338**(2): 369-82.

Jacob, M., M. Geeves, G. Holtermann and F. X. Schmid (1999). "Diffusional barrier crossing in a two-state protein folding reaction." Nat. Struct. Biol. **6**(10): 923-6.

Karplus, M. and D. L. Weaver (1976). "Protein folding dynamics." Nature (lond) **260**(5550): 404-406.

Karplus, M. and D. L. Weaver (1979). "Diffusion collision model for protein folding." Biopolymers **18**: 1421-1438.

Karplus, M. and D. L. Weaver (1994). "Protein folding dynamics: the diffusion-collision model and experimental data. [Review]." Protein Science **3**(4): 650-68.

Kawahara, K. and C. Tanford (1966). "Viscosity and density of aqueous solutions of urea and guanidine hydrochloride." Journal of Biological Chemistry **241**(13): 3228-32.

Kentsis, A., M. Mezei, T. Gindin and R. Osman (2004). "Unfolded state of polyalanine is a segmented polyproline II helix." Proteins **55**(3): 493-501.

Kentsis, A. and T. R. Sosnick (1998). "Trifluoroethanol promotes helix formation by destabilizing backbone exposure: Desolvation rather than native hydrogen bonding defines the kinetic pathway of dimeric coiled coil folding." Biochemistry **37**: 14613-14622.

Kramers, H. A. (1940). "Brownian motion in a field of force and the diffusion model of chemical reactions." Physica **7**: 284-304.

Krantz, B. A., L. Mayne, J. Rumbley, S. W. Englander and T. R. Sosnick (2002). "Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding." J. Mol. Biol. **324**(2): 359-71.

Krantz, B. A., L. B. Moran, A. Kentsis and T. R. Sosnick (2000). "D/H amide kinetic isotope effects reveal when hydrogen bonds form during protein folding." Nature Struct. Biol. **7**(1): 62-71.

Krantz, B. A. and T. R. Sosnick (2000). "Distinguishing between two-state and three-state models for ubiquitin folding." Biochemistry **39**(38): 11696-701.

Krantz, B. A. and T. R. Sosnick (2001). "Engineered metal binding sites map the heterogeneous folding landscape of a coiled coil." Nature Struct. Biol. **8**(12): 1042-1047.

Krantz, B. A., A. K. Srivastava, S. Nauli, D. Baker, R. T. Sauer and T. R. Sosnick (2002). "Understanding protein hydrogen bond formation with kinetic H/D amide isotope effects." Nature Struct. Biol. **9**(6): 458-63.

Kubelka, J., J. Hofrichter and W. A. Eaton (2004). "The protein folding 'speed limit'." Curr. Opin. Struct. Biol. **14**(1): 76-88.

Lacroix, E., A. R. Viguera and L. Serrano (1998). "Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters." J Mol Biol **284**(1): 173-91.

Lapidus, L. J., W. A. Eaton and J. Hofrichter (2000). "Measuring the rate of intramolecular contact formation in polypeptides." Proc. Natl. Acad. Sci. U S A **97**(13): 7220-5.

Levinthal, C. (1968). "Are there pathways for protein folding." J. Chim. Phys. **65**: 44-45.

Lumry, R., R. Biltonen and J. F. Brandts (1966). "Validity of the "two-state" hypothesis for conformational transitions of proteins." Biopolymers **4**: 917-44.

Maik, J. a. S., F. X. (1999). "Protein folding as a diffusional process." Biochemistry **38**(42): 13773-14118.

Martinez, J. C., M. T. Pisabarro and L. Serrano (1998). "Obligatory steps in protein folding and the conformational diversity of the transition state." Nature Struct. Biol. **5**(8): 721-9.

Matthews, C. R. (1987). "Effects of point mutations on the folding of globular proteins." Methods Enzymol. **154**: 498-511.

Mayor, U., N. R. Guydosh, C. M. Johnson, J. G. Grossmann, S. Sato, G. S. Jas, S. M. Freund, D. O. Alonso, V. Daggett and A. R. Fersht (2003). "The complete folding pathway of a protein from nanoseconds to microseconds." Nature **421**(6925): 863-7.

Mezei, M., P. J. Fleming, R. Srinivasan and G. D. Rose (2004). "Polyproline II helix is the preferred conformation for unfolded polyalanine in water." Proteins **55**(3): 502-7.

Millet, I. S., S. Doniach and K. W. Plaxco (2002). "Toward a taxonomy of the denatured state: Small angle scattering studies of unfolded proteins." Adv. Protein Chem. **62**: 241-262.

Moran, L. B., J. P. Schneider, A. Kentsis, G. A. Reddy and T. R. Sosnick (1999). "Transition state heterogeneity in GCN4 coiled coil folding studied by using multisite mutations and crosslinking." Proc. Natl. Acad. Sci. USA **96**(19): 10699-10704.

Munoz, V. and W. A. Eaton (1999). "A simple model for calculating the kinetics of protein folding from three-dimensional structures." Proc. Nat. Acad. Sci. USA **96**(20): 11311-6.

Munoz, V. and L. Serrano (1994). "Elucidating the folding problem of helical peptides using empirical parameters." Nat. Struct. Biol. **1**(6): 399-409.

Myers, J. K. and T. G. Oas (1999). "Reinterpretation of GCN4-p1 folding kinetics: partial helix formation precedes dimerization in coiled coil folding." J Mol Biol **289**(2): 205-9.

Myers, J. K., C. N. Pace and J. M. Scholtz (1995). "Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding." Protein Sci. **4**(10): 2138-48.

Noppert, A., K. Gast, D. Zirwer and G. Damaschun (1998). "Initial hydrophobic collapse is not necessary for folding RNase A." Fold. Des. **3**(3): 213-21.

Pappu, R. V. and G. D. Rose (2002). "A simple model for polyproline II structure in unfolded states of alanine-based peptides." Protein Sci **11**(10): 2437-55.

Pascher, T., J. P. Chesick, J. R. Winkler and H. B. Gray (1996). "Protein folding triggered by electron transfer." Science **271**(5255): 1558-60.

Plaxco, K. W. and D. Baker (1998). "Limited internal friction in the rate-limiting step of a two-state protein folding reaction." Proc. Natl. Acad. Sci. U S A **95**(23): 13591-6.

Plaxco, K. W., I. S. Millett, D. J. Segel, S. Doniach and D. Baker (1999). "Chain collapse can occur concomitantly with the rate-limiting step in protein folding." Nature Struct. Biol. **6**(6): 554-6.

Portman, J. J., Takada, S., Wolynes, P. G. (2001). "Microscopic theory of protein folding rates, II. Local reaction coordinates and chain dynamics." Journal of Chemical Physics **114**(11): 5082-96.

Qi, P. X., T. R. Sosnick and S. W. Englander (1998). "The burst phase in ribonuclease A folding and solvent dependence of the unfolded state." Nature Struct. Biol. **5**(10): 882-4.

Robertson, A. D. and R. L. Baldwin (1991). "Hydrogen exchange in thermally denatured ribonuclease A." Biochemistry **30**(41): 9907-14.

Rumbley, J., L. Hoang, L. Mayne and S. W. Englander (2001). "An amino acid code for protein folding." Proc. Natl. Acad. Sci. USA **98**(1): 105-12.

Sabelko, J., J. Ervin and M. Gruebele (1999). "Observation of strange kinetics in protein folding." Proc. Natl. Acad. Sci. USA **96**(11): 6031-6.

Sato, S., D. L. Luisi and D. P. Raleigh (2000). "pH jump studies of the folding of the multidomain ribosomal protein L9: the structural organization of the N-terminal domain does not affect the anomalously slow folding of the C-terminal domain." Biochemistry **39**(16): 4955-62.

Sato, S., Sayid, C. J., Raleigh, D. P. (2000). "The failure of simple empirical relationships to predict the viscosity of mixed aqueous solutions of guanidine

hydrochloride and glucose has important implications for the study of protein folding." Protein Sci. **9**(8): 1601-3.

Scalley, M. L., Q. Yi, H. Gu, A. McCormack, J. R. Yates, 3rd and D. Baker (1997). "Kinetics of folding of the IgG binding domain of peptostreptococcal protein L." Biochemistry **36**(11): 3373-82.

Scalley-Kim, M. and D. Baker (2004). "Characterization of the folding energy landscapes of computer generated proteins suggests high folding free energy barriers and cooperativity may be consequences of natural selection." J. Mol. Biol. **338**(3): 573-83.

Schlitter, J. (1988). "Viscosity dependence of intramolecular activated processes." Chem. Phys. **120**: 187-197.

Shi, Z., R. W. Woody and N. R. Kallenbach (2002). "Is polyproline II a major backbone conformation in unfolded proteins?" Adv. Prot. Chem. **62**: 163-240.

Shoemaker, B. A., J. J. Portman and P. G. Wolynes (2000). "Speeding molecular recognition by using the folding funnel: the fly-casting mechanism." Proc. Natl. Acad. Sci. USA **97**(16): 8868-73.

Sivaraman, T., Arrington C. B., Robertson A. D. (2001). "Kinetics of unfolding and folding from amide hydrogen exchange in native ubiquitin." Nat. Struct. Biol. **8**(4): 331-3.

Socci, N. D., J. N. Onuchic and P. G. Wolynes (1998). "Protein folding mechanisms and the multidimensional folding funnel." Proteins **32**(2): 136-58.

Sosnick, T. R., S. Jackson, R. M. Wilk, S. W. Englander and W. F. DeGrado (1996). "The role of helix formation in the folding of a fully alpha-helical coiled coil." Proteins **24**: 427-432.

Sosnick, T. R., L. Mayne and S. W. Englander (1996). "Molecular collapse: The rate-limiting step in two-state cytochrome c folding." Proteins **24**: 413-426.

Sosnick, T. R., L. Mayne, R. Hiller and S. W. Englander (1994). "The barriers in protein folding." Nature Struct. Biol. **1**: 149-156.

Sosnick, T. R., L. Mayne, R. Hiller and S. W. Englander (1995). The Barriers in Protein Folding. Peptide and Protein Folding Workshop, Philadelphia, PA, International Business Communications.

Sosnick, T. R., M. D. Shtilerman, L. Mayne and S. W. Englander (1997). "Ultrafast signals in protein folding and the polypeptide contracted state." Proc. Natl. Acad. Sci. USA **94**(16): 8545-50.

Sosnick, T. R. and J. Trewhella (1992). "Denatured states of ribonuclease A have compact dimensions and residual secondary structure." Biochemistry **31**(35): 8329-35.

Sreerama, N. and R. W. Woody (2003). "Structural composition of betaI- and betaII-proteins." Protein Sci. **12**(2): 384-8.

Svergun, D. I. and M. H. J. Koch (2003). "Small-angle scattering studies of biological macromolecules in solution." Rep. Prog. Phys. **66**(10): 1735-1782.

Takei, J., R. A. Chu and Y. Bai (2000). "Absence of stable intermediates on the folding pathway of barnase." Proc. Natl. Acad. Sci. U S A **97**(20): 10796-801.

Thompson, P. A., W. A. Eaton and J. Hofrichter (1997). "Laser temperature jump study of the helix-coil kinetics of an alanine peptide interpreted with a 'kinetic zipper' model." Biochemistry **36**(30): 9200-10.

Udgaonkar, J. B. and R. L. Baldwin (1995). "Nature of the early folding intermediate of ribonuclease A." Biochemistry **34**(12): 4088-96.

Wagman, M. E., C. M. Dobson and M. Karplus (1980). "Proton NMR studies of the association and folding of glucagon in solution." FEBS Lett **119**(2): 265-70.

Weissman, J. S. and P. S. Kim (1991). "Reexamination of the folding of BPTI:Predominance of native intermediates." Science **253**: 1386-1393.

Wilkins, D. K., S. B. Grimshaw, V. Receveur, C. M. Dobson, J. A. Jones and L. J. Smith (1999). "Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques." Biochemistry **38**(50): 16424-31.

Williams, S., T. P. Causgrove, R. Gilmanshin, K. S. Fang, R. H. Callender, W. H. Woodruff and R. B. Dyer (1996). "Fast events in protein folding: helix melting and formation in a small peptide." Biochemistry **35**(3): 691-7.

Yang, W. Y. and M. Gruebele (2003). "Folding at the speed limit." Nature **423**(6936): 193-7.

Zaman, M. H., M. Y. Shen, R. S. Berry, K. F. Freed and T. R. Sosnick (2003). "Investigations into sequence and conformational dependence of backbone

entropy, inter-basin dynamics and the Flory isolated-pair hypothesis for peptides." J. Mol. Biol. **331**(3): 693-711.

Zhu, Y., D. O. Alonso, K. Maki, C. Y. Huang, S. J. Lahr, V. Daggett, H. Roder, W. F. DeGrado and F. Gai (2003). "Ultrafast folding of {alpha}3D: A de novo designed three-helix bundle protein." Proc. Natl. Acad. Sci. U S A **100**(26): 15486-15491.

Zimm, G. H. and J. K. Bragg (1959). "Theory of the phase transition between helix and random coil in polypeptide chains." J. Chem. Phys. **31**: 526-535.

Zitzewitz, J. A., O. Bilsel, J. Luo, B. E. Jones and C. R. Matthews (1995). "Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy." Biochemistry **34**(39): 12812-9.

Zitzewitz, J. A., B. Ibarra-Molero, D. R. Fishel, K. L. Terry and C. R. Matthews (2000). "Preformed secondary structure drives the association reaction of GCN4-p1, a model coiled-coil system." J. Mol. Biol. **296**(4): 1105-16.

Zwanzig, R. (1997). "Two-state models of protein folding kinetics." Proc. Natl. Acad. Sci. USA **94**: 148-50.

APPENDIX

CHEVRON ANALYSIS

Chevron analysis is of critical importance to the studies described in chapters 2 and 3, as well as the great majority of protein folding studies. It takes advantage of the linear free energy dependence of protein stability, folding and unfolding rates on denaturant concentration.

$$\Delta G^{o}([\text{denaturant}]) = \Delta G^{o}_{\text{H2O}} - m^{o}\,[\text{denaturant}] \qquad \text{(A.1)}$$

$$\Delta G^{\ddagger}_{f}([\text{denaturant}]) = -\,\text{RT ln}\, k_f^{\,H2O} - m_f\,[\text{denaturant}] + \text{Constant} \qquad \text{(A.2)}$$

$$\Delta G^{\ddagger}_{u}([\text{denaturant}]) = -\,\text{RT ln}\, k_u^{\,H2O} - m_u\,[\text{denaturant}] + \text{Constant} \qquad \text{(A.3)}$$

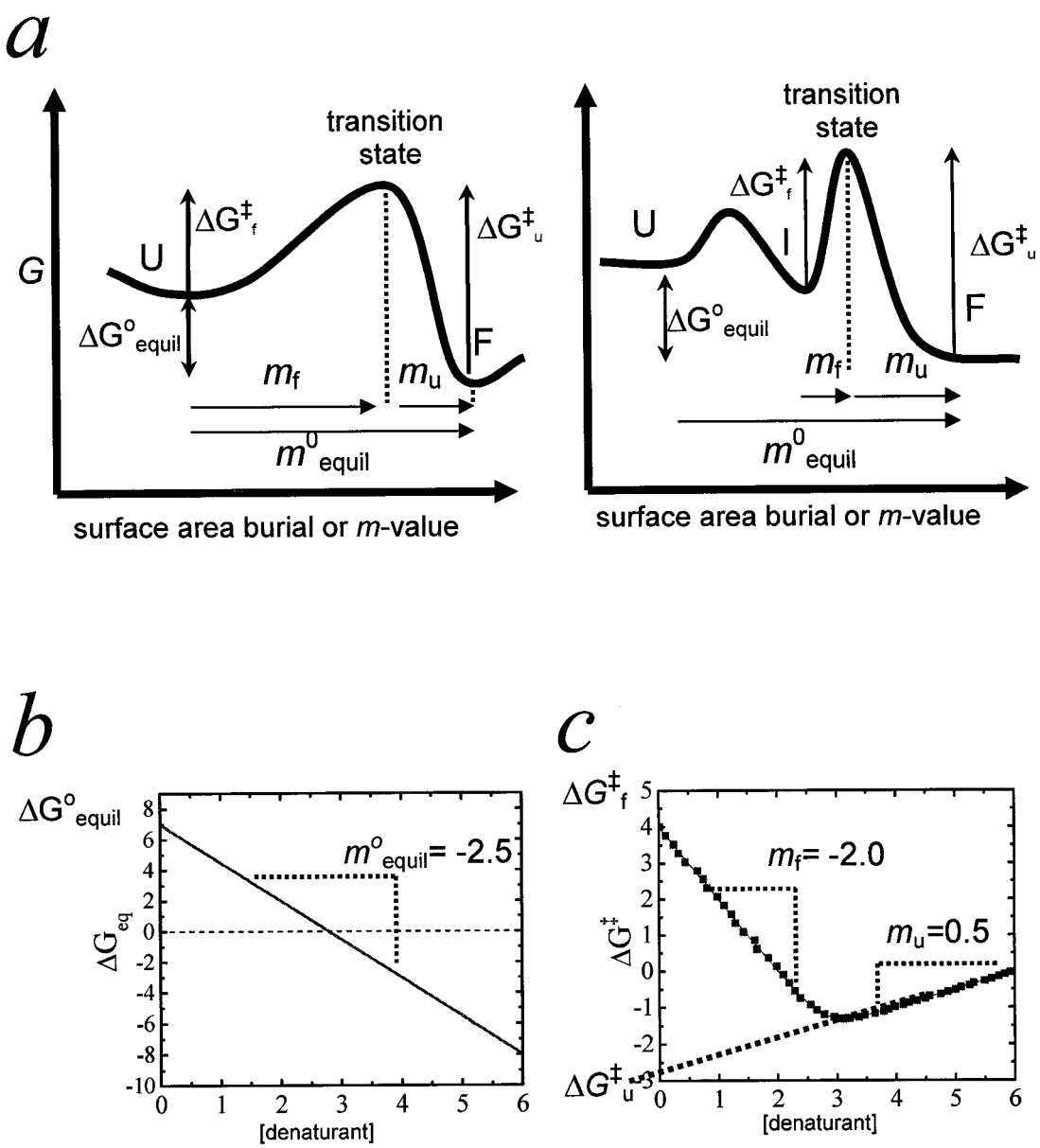
The Constant is proportional to the logarithm of the pre-exponential barrier-crossing attempt frequency (i.e the pre-factor in reaction rate theory), discussed in *section 1.5.*

These relationships allow kinetic free energies in the absence of denaturant to be obtained by extrapolation from measurements at moderate denaturant concentrations. For stable proteins, this extrapolation is often the only method available to determine unfolding and refolding rates under fully native conditions, because most experiments require that populations of molecules be manipulated from conditions where the majority are denatured to where the majority are folded to monitor folding, or vice versa for unfolding (native state hydrogen exchange, the technique employed in chapter 3, is a notable exception). As a result, direct measurements can often only be made near the denaturation midpoint

Chevron analysis compares the results from a denaturation profile to those determined from kinetics. Folding experiments begin with unfolded protein in a high concentration of denaturant that is rapidly diluted to known lower denaturant concentrations, and the relaxation kinetics followed to determine the denaturant dependence of folding (m_f). Unfolding experiments follow the same protocol, but begin with folded protein at low denaturant concentration and end with unfolded protein at high denaturant concentrations to determine the denaturant dependence of unfolding (m_u). These values, and the denaturant dependence of stability (m^0) ,calculated from a denaturation profile, enable extrapolation of their corresponding rates and stability in the absence of denaturant. Moreover, they are also proportional to the amount of surface area buried between the unfolded and transition states, the transition and native states, and the entire folding reaction (m_f, m_u, and m^0, respectively).

Representative two- and three-state, systems are illustrated in Fig. A1*a*. The free energies of folding ($\Delta G^{\ddagger}_{f}$ = RT lnk_f), unfolding ($\Delta G^{\ddagger}_{u}$ = RT lnk_u) and their denaturant dependencies (m_f and m_u, respectively) are measured by kinetic experiments, while molecular stability (ΔG^{0}_{equil}) and its denaturant dependence (m^{0}_{equil}) is determined at equilibrium. In the two-state scenario, both kinetic and equilibrium experiments reflect the complete unfolded-to-folded (U ←→ F) transition, so the observed activation free energies of folding and unfolding and their denaturant dependencies fully account for the molecular stability and surface burial characteristics observed at equilibrium. This contrasts with the three-state system where a stable intermediate populates prior to the major free energy barrier. The intermediate rapidly accumulates during kinetic experiments, so the observed rates are determined by the activation energies and surface burial of the intermediate-to-folded (I ←→ F) transition, falling short of the values observed at equilibrium. For the model data displayed in Fig, A1***b*** and ***c***, the observed kinetic free energies and surface burial parameters fully account for those measured at equilibrium, so there can be no stable intermediate formed prior to the major free energy barrier. This confirms that the system is two-state.

Figure A.1 *Chevron analysis.* ***(a)*** Free energy diagram of two- and three-state folding where the reaction coordinate represents the amount of surface area buried during folding. ΔG^0_{equil} is the free energy difference between the unfolded and folded states, while $\Delta G^{\ddagger}_f$ and $\Delta G^{\ddagger}_u$ are proportional to the height of the energy barrier to folding and unfolding, respectively, and their difference is $\Delta G^0_{kin} = \Delta G^{\ddagger}_f - \Delta G^{\ddagger}_u$. The total amount of surface area buried as the molecule folds is m^0_{equil}, and the observed denaturant dependence of folding and unfolding are m_f and m_u, respectively, their difference being $m^0_{kin} = m_f - m_u$. ΔG^0_{equil} and m^0_{equil} are determined from a denaturation profile while the remaining parameters are measured kinetically. In a two-state system, both equilibrium and kinetic experiments monitor the complete unfolded-to-folded (U ←→ F) transition, so $\Delta G^0_{equil} = \Delta G^0_{kin} = \Delta G^{\ddagger}_f - \Delta G^{\ddagger}_u$ and $m^0_{equil} = m^0_{kin} = m_f - m_u$. However, in a three-state system with a stable intermediate prior to the major energy barrier, the intermediate populates quickly during folding, so kinetic experiments reflect the intermediate-to-folded (I ←→ F) transition. As a result, $\Delta G^0_{equil} > \Delta G^0_{kin} = \Delta G^{\ddagger}_f - \Delta G^{\ddagger}_u$ and $m^0_{equil} > m^0_{kin} = m_f - m_u$. ***(b)*** A model denaturation profile of free energy ($\Delta G = RT\ln K_{eq}$) vs. denaturant concentration. $\Delta G^0_{equil} = 7$ is the free energy extrapolated to the absence of denaturant and the slope is $m^0_{equil} = -2.5$. *(c)* A model kinetic chevron displaying the denaturant dependence of the folding (*left arm*) and unfolding (*right arm*) free energies. Here, $\Delta G^{\ddagger}_f = 4$ and $\Delta G^{\ddagger}_u = -3$ are the folding and unfolding free energies extrapolated to the absence of denaturant, respectively and their difference $\Delta G^0_{kin} = 7$. The slopes of the folding and unfolding arms are $m_f = -2.0$ and $m_u = 0.5$, respectively, and their difference is $m^0_{kin} = m_f - m_u = -2.5$. Thus, for this example, $\Delta G^0_{kin} = \Delta G^0_{equil}$ and $m^0_{kin} = m^0_{equil}$, indicating that the free energy and surface burial characteristics of the kinetic data fully account for that observed at equilibrium, and the system is adequately modeled by the two-state, barrier-limited approximation.


a
G
U
transition state
ΔG‡f
ΔG‡u
ΔG°equil
F
mf
mu
m0equil
surface area burial or m-value
U
I
transition state
ΔG‡f
ΔG‡u
ΔG°equil
F
mf
mu
m0equil
surface area burial or m-value
b
ΔG°equil
m°equil= -2.5
ΔGeq
[denaturant]
c
ΔG‡f
mf= -2.0
mu=0.5
ΔG‡
ΔG‡u
[denaturant]