Chemical Physics Letters 461 (2008) 1–8

Contents lists available at ScienceDirect

Chemical Physics Letters

journal homepage: www.elsevier.com/locate/cplett

FRONTIERS ARTICLE

Downhill dynamics and the molecular rate of protein folding

Feng Liu [a], Martin Gruebele [a,b,c,*]

[a] Center for Biophysics and Computational Biology, USA
[b] Beckman Institute, Department of Physics, University of Illinois, Urbana, IL 61801, USA
[c] Department of Chemistry, University of Illinois, Urbana, IL 61801, USA

ARTICLE INFO

Article history:
Received 4 April 2008
In final form 16 April 2008
Available online 22 April 2008

ABSTRACT

Proteins are held together by many weak contacts, each corresponding to a local reaction coordinate. The activation barrier for folding is distributed along a resultant global folding coordinate. Hence folding barriers are low, and could even become comparable to the thermal energy *kT*. In that case, proteins become downhill folders, with folding times in the microsecond region. Small barriers allow the diffusion of population along the reaction coordinate – the molecular rate – to be observed directly. Five simple free energy building blocks can explain all experimentally observed fast folding data, revealing a range of behaviors from low barrier crossings to completely downhill folding.

1. Introduction

Protein folding is a rich biological, physical and chemical problem: protein sequences are subject to evolution for function; their folding is subject to basic physical constraints, from excluded volume to loop entropies; the many weak chemical interactions, such as hydrogen bonds, or water–sidechain interactions leading to the hydrophobic effect, produce a complex energy landscape [1].

Protein folding can be viewed as a unimolecular reaction connecting the unfolded and folded states, but the existence of a single folding coordinate is less obvious for proteins than for small molecules because many distributed weak interactions occur. Nonetheless, a small number of important global coordinates must emerge, or folding reactions would not be as fast (μs to hours) as they are empirically found to be.

Energy landscape models first suggested that the entropy and enthalpy of folding could nearly cancel at each step along a global folding coordinate, resulting in downhill folding [2]. Proteins are held together by many weak interactions with strengths of a few $kT \sim 1$–10 kJ/mol. So it is possible that the most important reaction coordinate is composed of many small steps, each of which decreases the entropy of the solvated chain and the contact enthalpy among residues by small, compensating amounts. The net result is a series of small free energy barriers along the global reaction coordinate termed 'residual roughness'. Folding along such a reaction coordinate is a hindered *diffusion* process, rather than an *activated* barrier crossing, when the largest barrier is on the order of *kT*. We say that such a protein has a strong bias towards the native state, or folds downhill towards the native state.

When a protein is stressed, one or more of these small barriers could grow beyond a few *kT*. Stressing results from tuning the temperature above or below the temperature of maximal stability in Fig. 1, adding denaturants, or mutating amino acid residues to decrease protein stability. Such destabilizing residues can be introduced by evolution for protein function or by evolution against protein aggregation [1,3]. The consequence of larger barriers is folding through a sequence of intermediates, or even folding over a single dominant barrier, so-called two-state folding [4]. The 'intermediates' scenario has been accepted for decades, and many examples have been studied [5]. The two-state scenario is more recent, but ample evidence exists that it adequately describes the folding of many small proteins [4]. The downhill scenario is the youngest and most controversial. One reason for the controversy is that in the past, protein folding was mainly studied by stressing proteins, in the process creating activation barriers that obscure downhill folding (Fig. 1). As discussed later on, applying such stresses systematically, for example by tuning the temperature, can be used to identify downhill folding.

The comparison of activated and diffusional time scales required by downhill protein folding creates an interesting conundrum. Modern unimolecular rate models based on the theory of Kramers [6] posit an activated reaction rate

$$k_a = k_m(\eta_{loc})e^{-\Delta G^\dagger/RT}. \quad (1)$$

The activated rate coefficient k_a is split into a molecular rate coefficient k_m dependent on local viscosity η_{loc} [7,8], and a Boltzmann factor that accounts for reduced population at the free energy barrier $\Delta G^\dagger$. When the folding process is downhill ($\Delta G^\dagger = 0$), a molecular phase with rate coefficient k_m is observed, corresponding to the rate of diffusion between folded and unfolded states; complex diffusion processes may not even be characterized by a single rate coefficient k_m, and are often fitted by stretched exponentials [9,10]. When the

* Corresponding author. Address: Department of Chemistry, University of Illinois, Urbana, IL 61801, USA. Fax: +1 217 244 3186.
E-mail address: gruebele@scs.uiuc.edu (M. Gruebele).

0009-2614/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cplett.2008.04.075

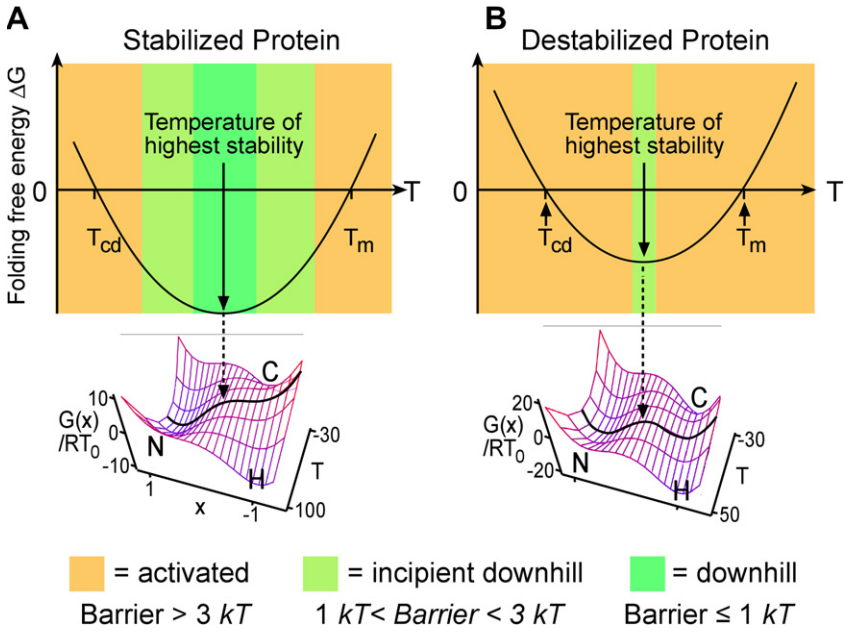

Fig. 1. Correlation between downhill folding and protein stability. (*Top:*) Orange to green signifies the shift from activated to downhill folding throughout the paper. Near the heat and cold denaturation transitions (T_m and T_{cd}) where the folding free energy $\Delta G = 0$, folding is generally an activated process (orange). When stress on the protein is reduced, ΔG reaches a minimum at the point of maximal stability. Very stable proteins (left) will reach the green downhill zone because the free energy surface (left bottom) is highly biased towards the native state. Less stable proteins (right) never achieve strong enough native bias, and never reach the green downhill zone. N, H and C indicate the locations of the native, heat- and cold denatured states of the protein along the reaction coordinate.

folding barrier is large, the molecular rate coefficient k_m is simply the prefactor for activated folding. For activated processes, only k_a can be measured directly, leaving $\Delta G^{\dagger}$ dependent on the choice of k_m. k_m in turn depends on the choice of reaction coordinate, which is not as obvious for proteins as for small molecules. Consider a local coordinate: a particular rotation of an amino acid side chain may have $k_m \sim 1\ ps^{-1}$. The problem with such local coordinates is that many are required to describe folding even approximately. By coarse-graining the coordinates, one eventually arrives at global coordinates that are more efficient descriptors of folding. Such global coordinates, like formation of several tertiary contacts that define the folded shape of the protein, will have much slower molecular time scales, typically on the order of $k_m^{-1} = 0.05–1\ \mu s$. The essential idea of defining a free energy surface for protein folding, as opposed to the full protein-solvent potential energy surface, is to find a good coarse grained global coordinate.

This is a recurring theme in chemical physics: how do we scale our understanding of a large number of fast, localized processes, to account for a much slower, global process? The idea of slaving of slow protein motions to fast solvent degrees of freedom is a 'top–down' example of this problem [11]. Studying local conformational transitions of small peptides to infer information about how the much larger protein might behave, is a 'bottom–up' example [12]. Fortunately, the theory of reaction dynamics provides us with two very general criteria for choosing good global reaction coordinates [7]:

- A good set of global coordinates spans the transition region between folded and unfolded states with a minimal number of recrossings, thus yielding a minimal k_m.
- A good set of global coordinates is of minimal size while projecting out maximal information about the reaction (location of local minima and saddle points traversed with high probability during folding).

From an experimental point of view, different spectroscopic probes $S(x)$ represent different coordinates and monitor different protein populations [13]: the IR spectrum at 1650 cm^{-1} monitors extent of helical hydrogen bonding; tryptophan fluorescence quenching and FRET (Förster resonant energy transfer) monitor local contacts; SAXS (small angle X-ray scattering) monitors the radius of gyration; an NOE (nuclear Overhauser effect) peak monitors the proximity of two sidechains. All these probes are fairly coarse-grained already. The k_m values for these probes, if they could be measured, yield upper limits for the average molecular rate along the best single (or at least the best few) global reaction coordinates needed to describe folding. The observed rate k_a may still turn out to be much slower than k_m for any choice of probe, in which case a genuine free energy barrier exists for the overall folding reaction.

By allowing the direct observation of protein populations everywhere along the reaction coordinate, downhill folding makes a connection between local dynamics and global kinetics not possible with activated folding: k_m can be measured separately from k_a. Fig. 2 illustrates how this works, and makes the crucial point, occasionally overlooked in the computational folding literature [14,15], that identifying downhill folding kinetically or thermodynamically requires stressing of the protein [8,10,16–18]: The protein free energy surface must be tuned, for example by changing temperature as shown in Figs. 1 and 2. Upon tuning, downhill free energy surfaces predict robust correlations between kinetic amplitudes and rates that would require inordinate fine-adjustment of parameters to be explained by activated kinetics via intermediates [13].

One such prediction is a smooth transition from slow exponential kinetics, to a slow exponential plus fast phase, to just a fast phase when a protein of the type shown in Fig. 1A is stabilized by lowering its temperature from T_m to the temperature of maximal stability. Under stress (Fig. 2, top), the protein ensemble samples only the folded and denatured wells, and a slow single exponential unimolecular reaction is expected. When stress is lowered, the folding protein population samples the entire reaction coordinate with significant probability. Now the unimolecular reaction kinetics are preceded by a fast molecular phase with rate coefficient k_m sampling diffusion of the protein population. This

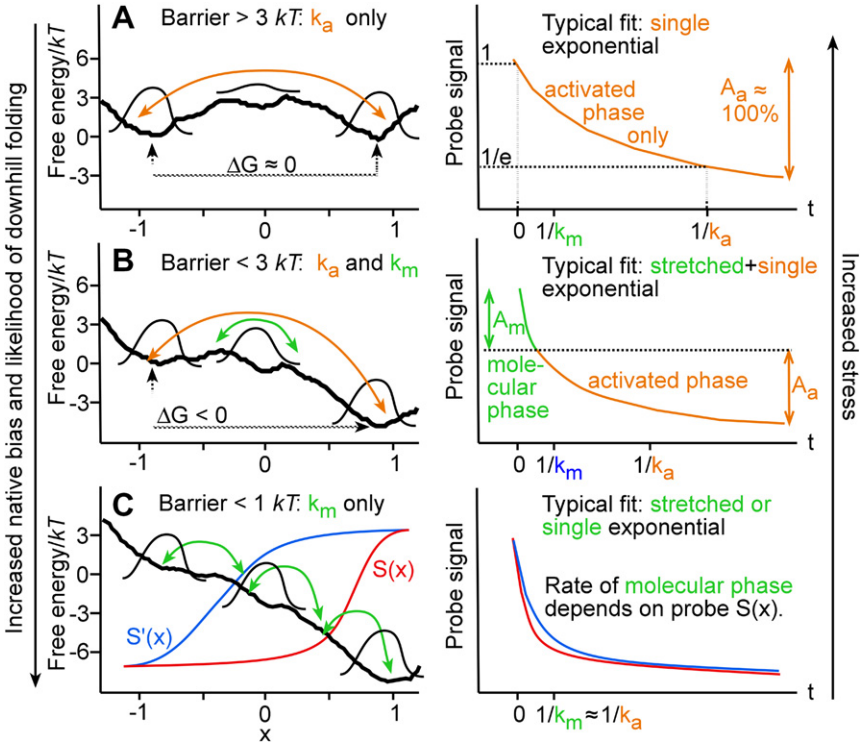

Fig. 2. Kinetic and thermodynamic studies of downhill folding rely on smooth tuning of protein behavior when stress on the protein is changed by temperature-, solvent- or mutation-tuning. (*Top:*) high stress ($\Delta G \approx 0$, unfolded and folded free energies are equal): the activation barrier is large, the barrier-top population is small, and single-exponential activated kinetics (orange) are observed. All probes yield the same reaction rate. Middle: stress is reduced ($\Delta G < 0$), and the barrier is low enough so a significant barrier top population exists. The fast molecular phase (green) and the residual activated phase (orange) coexist. (*Bottom:*) minimal stress results in a merger of activated (k_a) and molecular (k_m) rates for particularly stable proteins. Probes with different coordinate dependence $S(x)$ (red, blue) yield different reaction rates.

becomes observable even in experiments with moderate signal-to-noise ratio when the bias towards the native state drops activation barriers below 3 *kT*, assuring $\geqslant$5% protein populations in the transition region between unfolded and folded states (Fig. 2, middle). Finally, if the protein is sufficiently strongly biased towards the native state (Fig. 2, bottom), only the fast molecular phase remains. It has been observed as either a single exponential relaxation with rate coefficient k_m, or as stretched exponential relaxation, depending on the nature of downhill diffusion process [16,19]. According to a recent proposal, some proteins remain in the downhill limit even when they are stressed [17]. We will term folding with barriers above 3 *kT* as activated folding, between 1 and 3 *kT* as incipient downhill folding, and below 1 *kT* as downhill folding. These values are somewhat arbitrary, but it is fair to say that they correspond to zones where activated rate theory fails only slightly, or significantly, or completely [20].

2. A brief history of downhill folding

Downhill folding was first proposed by Bryngelson, Onuchic and Wolynes in 1995 [2], and termed the 'type 0' scenario of folding (bottom of Fig. 2). In their scenario, a given native protein fold can be achieved without a barrier under optimal sequence and solvent conditions. If the protein is stressed, energy-entropy compensation is lost, and one or more significant barriers appear (type 1 folding, also known as two-state folding; top of Fig. 2). More recently, a scenario having a single well at all temperatures has been proposed by Muñoz and coworkers [17]. Such a protein retains the ability to fold downhill even under thermal stress.

k_m has been estimated by diffusion measurements on unfolded peptides, and was found to be about $(1\ \mu s)^{-1}$ for a chain length of about 50 residues [21]. Downhill folding to a compact state under reduced thermal stress was first observed by Sabelko et al. in 1999 [10]. They measured an exponential to non-exponential transition of kinetics when the protein was stabilized, attributed to appearance of the molecular phase k_m (Fig. 2). This and subsequent work determined a molecular rate of $(20–35\ \mu s)^{-1}$ for the 415 residue enzyme phosphoglycerate kinase. Dyer and coworkers [22], and more recently Tokmakoff and coworkers [23], have used temperature jumps to switch the location of the activation barrier along the reaction coordinate, such that downhill folding from the barrier top could be observed. In several papers since 2003, Yang, Liu and Gruebele have reported a complete exponential–non-exponential–exponential kinetics transition of the type summarized in Fig. 2, with $k_m \approx (1–2.5\ \mu s)^{-1}$, and k_a ranging from 2.5 μs to 2 ms [8,16,24]. They also showed that the molecular rate coefficient k_m, but not the activated rate coefficient k_a, is viscosity-dependent (Eq. (1)). The groups of Gai, Eaton, and others [25–29] have observed several more fast folding proteins and peptides, some of which are fast enough to be completely in the downhill limit, while others show non-exponential kinetics which we interpret in terms of incipient downhill folding as outlined in Fig. 2. Very recently, a family of 35 WW domains (named after two tryptophan residues = WW) has been investigated, uncovering four incipient downhill folders and establishing a quantitative correlation between protein stability and downhill folding [20].

Although not the main topic of this Frontiers article, folding thermodynamics have also been scrutinized in detail. In thermodynamic folding experiments, the temperature or solvent is tuned until a change of state (between U = unfolded, N = native, or I = intermediate) is observed. For example, one refers to 'melting' the protein when the temperature is increased so N → U (at T_m in Fig. 1). Probes such as infrared, fluorescence, SAXS, or NMR, already discussed above, are used to monitor the state change.

A key problem in such experiments is that one has to distinguish baselines (due to the intrinsic temperature dependence of fluorescence intensity, for example) from real changes in signal caused by folding. In 1999, Parker and Marqusee pointed out that downhill U → I transitions could mimic activated two-state folding if one allows for baselines when probes are tuned [30]. In 2002 Sadqi et al. showed that downhill folding could lead to melting point discrepancies and inconsistent baselines when different probes are compared [31]. Their work was subsequently extended to a whole array of probes. Other very fast folding peptides and proteins have also been shown to have thermodynamic anomalies, such as inconsistent melting points when interrogated by several different probes [32]. As in the kinetic measurements, tuning of parameters such as temperature is key to identifying downhill folding.

Downhill folding kinetics have been examined critically in the literature. It has been pointed out that non-exponential or probe-dependent kinetics observed at a single temperature are not a unique signature of downhill folding, and that activated three-state fits could also account for the results [14]. These arguments have been answered by observation of a complete exponential–non-exponential–exponential transition as outlined in Fig. 2, by direct thermodynamic and kinetic comparison of an activated protein with a similar downhill protein, by the temperature- and viscosity-dependence of probe-dependent kinetics, and by a survey of many fast folding proteins [18–20,24,33], showing that the appearance of downhill folding is always correlated with increased stabilization of the protein Fig. 1. Computational work has indicated that some of the hotly debated examples can fold downhill (<1 kT barrier), incipiently downhill, or over a significant barrier, depending on the exact details of sequence and solvent used [34]. Other proteins have been engineered to fold to the same structure by many different mechanisms, showing that the same folded structure can support multiple mechanisms [3,16,19,20,24,35,36]. At the very least, all of these results indicate that barriers are not much greater than the intrinsic energy landscape roughness, estimated by theory and experiments between 0.8 and 1.3 kT [16,32,37,38].

3. Molecular rates and activated kinetics from five free energy building blocks

We previously applied the Langevin dynamics model to folding along one or two reaction coordinates, generalizing Eq. (1) to low/multiple barriers [13]. Here we review this model and its key features, so we can use it to interpret the fast folding data in the current literature in the next section.

Rather than describing motion along the free energy surfaces in Fig. 2 by a discretized set of states and a kinetic master equation, we describe the dynamics of a protein trajectory by continuous motion along the reaction coordinate,

$$-\frac{\partial G(x,T,P)}{\partial x}-\gamma\frac{\mathrm{d}x}{\mathrm{d}t}+\xi(t)=0. \qquad (2)$$

The derivative is the force acting on a protein as a result of the free energy surface G, the velocity term represents frictional damping, and the random force term ξ, related to the frictional term γ by the fluctuation–dissipation theorem, describes Brownian dynamics of the protein in the implicit solvent. In a relaxation experiment such as a temperature jump, G is suddenly switched by switching T. The resulting out-of-equilibrium distribution relaxes back to equilibrium according to Eq. (2). A more complete discussion, including higher dimensional reaction coordinates, is given elsewhere [13].

The key ingredient is the free energy surface $G(x, T, P)$ as a function of reaction coordinate x. For proteins, G is made of five basic components that can be assembled to match different folding scenarios. Fig. 3A–E illustrates these five building blocks, and shows how they are altered when the native bias (the relative stability of the native state) is changed.

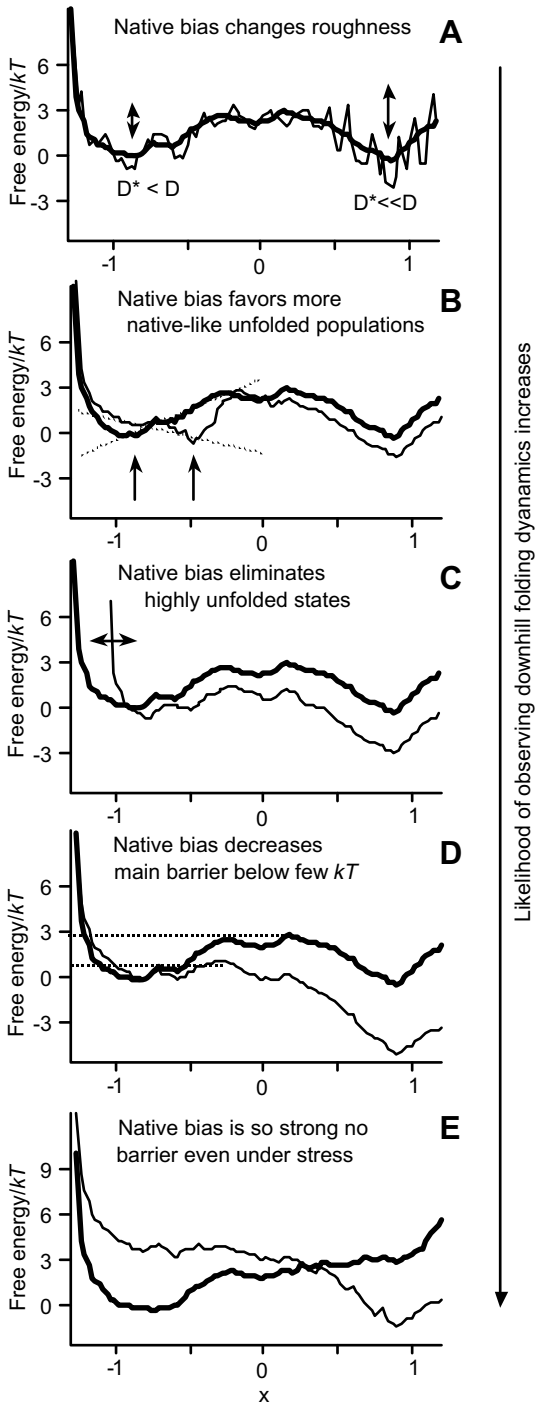

Fig. 3. Free energy building blocks in one dimension that contribute to observed dynamics of fast folding. The thick free energy indicates a protein under higher stress (e.g. temperature near the melting point T_m), the thin free energy curve indicates a protein under lower stress (e.g. temperature below the melting point). (A) Roughness can alter the coarse-grained diffusion coefficient, or create local minima deep enough to be labeled long-lived intermediates; (B) the unfolded state local minimum becomes more native-like when stress is reduced; (C) the relative free energy of highly unfolded states increases when stress is reduced; (D) the free energy is biased towards the native state, reducing the activation barrier by Hammond's principle; (E) native bias is so strong that melting occurs without formation of an activation barrier.

Simulations and experiments report a residual roughness of about 1 kT for proteins (Fig. 3A) [16,38]. The roughness depends

on solvent and sequence. For example, salt bridges made by amino acids of opposite charge can increase roughness through non-native pairwise interactions. Increased native bias can also lead to a more compact protein with increased self-friction and hence more roughness [39]. Paradoxically, the resulting increased roughness keeps the folding rate constant as the protein begins to fold downhill. Fig. 4 shows how the effective diffusion coefficient scales with roughness for surfaces with an average barrier of 2 *kT*. At very low temperatures, the hydrophobic effect is eventually reduced, and roughness decreases with the onset of cold denaturation.

When the roughness increases to several *kT* in scenario 3A, the resulting deep local minima can be identified as a long-lived intermediates, making a connection with that classic folding scenario. The protein folding kinetics then show at least a double exponential decay, but both phases have a folding time longer than several μs, and their amplitudes and rate coefficients are not correlated in any particular way, unlike the downhill folding scenario in Fig. 2.

The residual native structure of unfolded proteins increases with more native bias. This has been verified directly by single molecule experiments that observed unfolded proteins under native conditions [40]. Such pre-existing structure in the unfolded state can greatly speed up folding. In Fig. 3B, the unfolded free energy minimum shifts towards the native state. In scenario 3C, the highly unfolded states are less populated as their free energy becomes large. This results in a steep free energy gradient $\partial G/\partial x$ during relaxation experiments, and hence a very fast molecular phase. Scenario C can lead to biexponential kinetics with a ratio of $k_m/k_a > 50$ and a molecular phase amplitude $A_m > 10\%$. The unfolded population immediately after the jump in free energy folds downhill very rapidly (k_m), and then undergoes a slower barrier crossing (k_a). The fast phase k_m will be faster than the diffusive motion at the barrier, which has a less steep slope. Thus the observed k_m sets an upper limit on the diffusive dynamics in the transition region.

Even when the unfolded structure does not shift very much, increasing the native bias decreases the activation barrier according to Hammond's principle [41]: the transition state does not have all the favorable contacts of the native state, but it does have some of them; stabilizing the native state thus also stabilizes the transition region. This trickle-down effect from native state to transition region leads to the case shown in Fig. 3D. As the native state is stabilized and the activation barrier is lowered, relaxation kinetics shift from a slow single exponential (k_a only), to stretched + exponential (molecular phase and k_a), to fast stretched kinetics (molecular phase only). If downhill diffusion is normal, the molecular phase may be characterized by a single rate coefficient k_m instead of a stretched exponential, as assumed in Eq. (1). In scenario 3D, the amplitude A_m of the fast molecular phase is significant only if $k_m/k_a < 15$, and amplitude and ratio are strongly correlated as $A_m/(A_a+A_m) = 0.5–0.025(k_m/k_a-1)$ [13,24]. This smooth correlation of amplitudes and rate coefficients has been used to identify downhill folding.

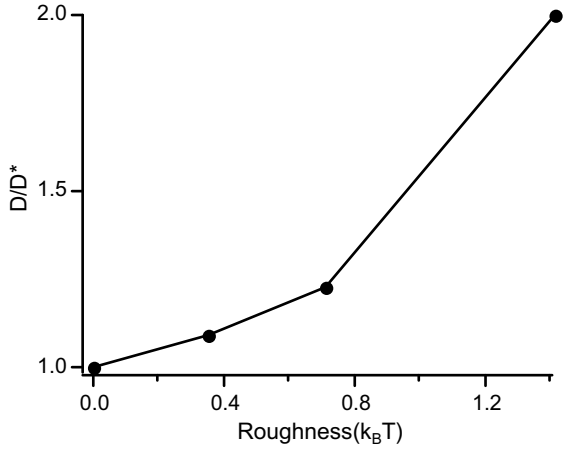

Fig. 4. Computed dependence of the inverse effective diffusion constant $1/D^*$ on the root-mean-square roughness of a free energy surface with a minimal barrier of 2 *kT*. A similar result is obtained for net barriers in the 0–3 *kT* range, with diffusion decreasing slightly at larger roughness.

Fig. 3E shows the scenario of downhill folding with the strongest native bias. There is no barrier left even under stress, and the single well simply shifts from the unfolded state to native state gradually as native bias increases. We expect to see ultrafast stretched- or single-exponential folding kinetics, depending on whether diffusion is anomalous or normal, but no exponential–non-exponential–exponential transition. The folding rate on such a surface is less temperature-dependent, with only changes in diffusion coefficient and roughness on a *kT* energy scale contributing to changes in the observed rate.

In real proteins, the five building blocks A–E will not contribute in isolation, but can be combined to explain the observed fast folding kinetics. Different proteins will have observed dynamics controlled by different free energy building blocks.

4. A survey of experimental downhill candidates

The observed fast folders fall into several categories, based on the dominant building blocks of their free energy surfaces. Our Langevin dynamics simulations quantitatively fit the experimentally observed behavior, and in some cases, allow assignment of previously unassigned kinetic phases. Fig. 5 illustrates four classes of downhill folding or incipient downhill folding proteins and peptides. Representative free energy curves as a function of temperature are shown on the left, and calculated relaxation kinetics following an instantaneous temperature jump between the two free energy curves are shown on the right. Generally, a smooth bias of the free energy towards the native state as the temperature is lowered suffices to fit the experimental data.

The first category of proteins Fig. 5A makes a gradual transition from activated to downhill folding upon reducing stress. In most studies stress was reduced by lowering the temperature below the melting temperature. When folding is monitored by spectroscopic probes whose signal changes slowly across the transition region ($S'(x)$ in Fig. 2C), k_m reflects diffusional dynamics governed by residual roughness in the transition region. This corresponds to Fig. 3D. Typical representatives of this category of proteins are mutants of lambda repressor fragment 6–85 (λ_{6-85}) [8,16,19,24], the artificial three-helix bundle α_3D [25], and mutants of the beta sheet protein WW domain [20]. The wild type of (λ_{6-85}) is a fast two-state folder and folds with single exponential kinetics at all temperatures. By successively combining mutations that stabilize the native state, such as Y22W, G46A, and G48A, and D14A, a smooth transition from activated, to incipient downhill (stretched + exponential), to downwhill folding has been observed (Fig. 5A). In the incipient downhill folding region (barrier between 1 and 3 *kT*), the simultaneously observed molecular and activated phases are strongly correlated irrespective of mutation, temperature, or solvent additives. The correlation obeys the formula predicted by Langevin dynamics simulations on downhill surfaces as discussed in the previous section. The molecular phase has $k_m \approx (1–2\ \mu s)^{-1}$ [8,16,24]. Additional experimental clues pointing towards downhill folding include probe-dependent kinetics (Fig. 2C) [19], and different viscosity scaling of molecular and activated phase [16]. No model other than downhill folding has been able to explain all of these observations quantitatively. The 73-residue three-helix bundle protein α3D also shows a signature of incipient downhill folding [25]. IR and fluorescence probes both reveal very fast

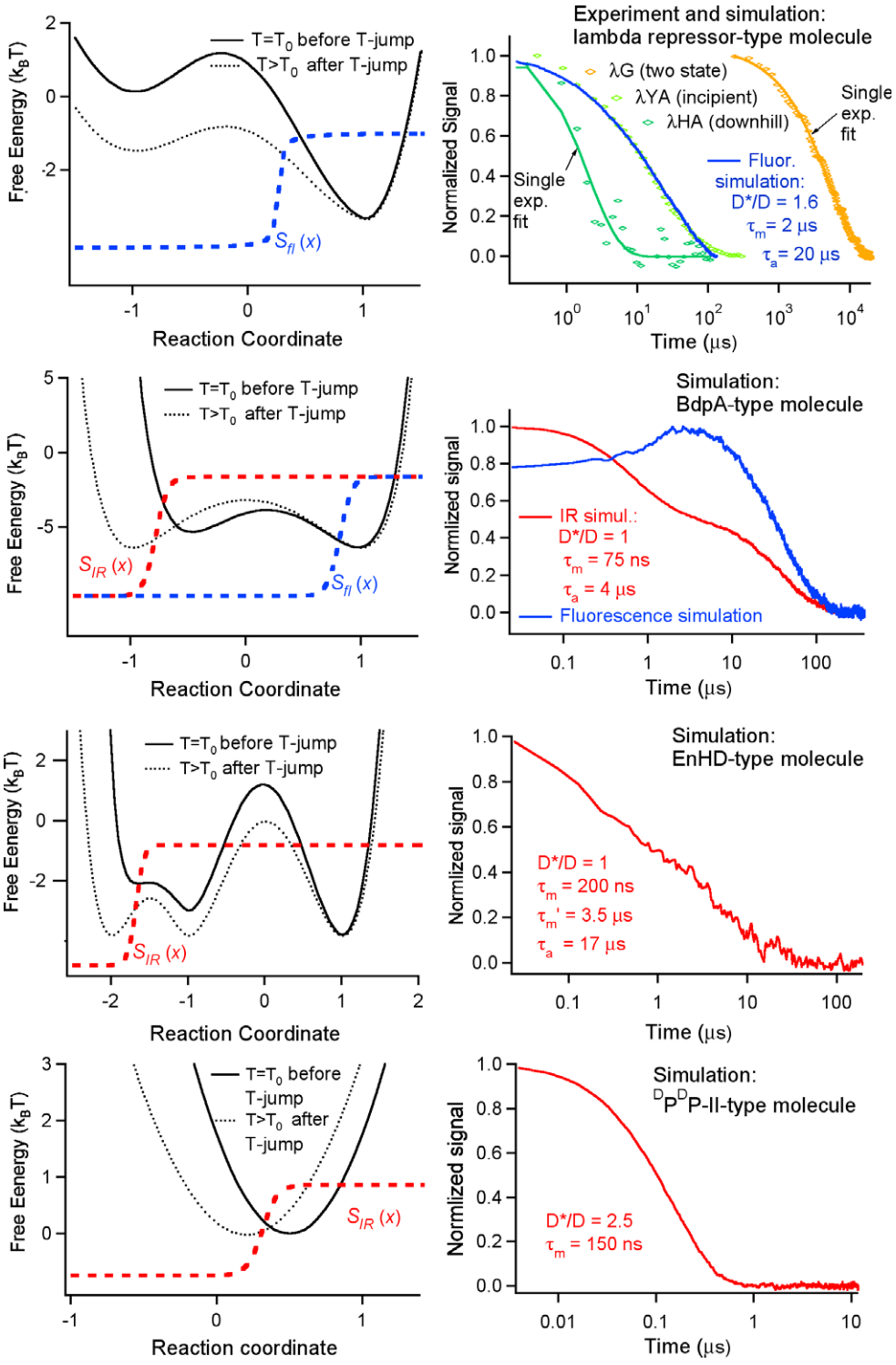

Fig. 5. Free energy surfaces (left) and Langevin simulations of downhill relaxation dynamics (right) constructed from the basic free energy building blocks. The dotted surface is the one under higher stress = less native bias. Changes in roughness are accounted for here by changing the ratio (D^*/D) compared to the nominal diffusion coefficient $D = 10^{-4}$ nm/ns^2. τ_m is defined as k_m^{-1} (Top row:) proteins like lambda repressor smoothly lose their barrier when native bias is increased. The experimental data ranges from fully downhill folding (green, fast single exponential) to activated folding (orange, slow single exponential). Incipient downhill folding is fitted by a Langevin simulation combining both fast molecular and slow activated phases. Protein sequences and solvent conditions in order of increasing stability are: λG (Y22WA47,49 G mutant in 0.5 M GuHCl denaturant at 34 °C), λYA (Y22WQ33YG46,48A mutant in 68 °C phosphate buffer) and λHA (Y22WQ33HG47,48A at 44 °C in phosphate buffer [8,16,24]). (Second row:) Proteins like BdpA have a probe that emphasizes changes in residual unfolded state structure when the native bias changes. This results in a faster molecular phase (smaller τ_m). Large differences in probe $S(x)$ can even lead to non-monotonic kinetics (compare our simulation to data in fig. 1 of Ref. [42]). (Third row:) Proteins like engrailed homeodomain retain a small (here, 4 kT) free energy barrier to the native state, but unfolded state structure forms downhill ($\leqslant 1$ kT barrier) and there is a loss of highly unfolded states under native conditions. Langevin dynamics show the resulting three phases, a very fast molecular phase analogous to BdpA, a slower molecular phase analogous to lambda repressor, and an even slower activated phase (compare our simulation to data in fig. 4 of Ref. [43]). (Bottom row:) Proteins like the designed beta-strand protein ^{D}P^{D}P-II remain biased downhill even under non-optimal conditions (e.g. higher temperature). A fast single exponential decay is observed in Langevin dynamics, or a fast stretched exponential if diffusion is higher-dimensional or anomalous (compare our simulation to data in fig. 3 of Ref. [27]). These four scenarios and similar fits for other fast folders explain all current fast folding data in terms of incipient or complete downhill folding.

2–3 μs folding kinetics, but with a different temperature dependence for each probe (Dumont, Zhu, Gai & Gruebele, unpublished results). Several incipient downhill folders have been identified among stabilized mutants of 35-residue variants of WW domain [20]. Only proteins with a melting temperature higher than a threshold of 50 °C show stretched + exponential kinetics with a molecular rate $k_m \approx (1–5\ \mu s)^{-1}$. The model depicted in Fig. 1A accurately predicts this threshold temperature observed in kinetics measurements [20].

A second category of proteins (Fig. 5B) undergoes a rapid population shift, driven by the large gradient $\partial G/\partial x$ of the unfolded state free energy illustrated in Fig. 3C. The corresponding k_m sets an upper limit on the diffusional time scale in the transition region, where such large gradients are absent. Such proteins have $k_m/k_a >> 15$ and large amplitudes for the molecular phase. The activated rate coefficient k_a can also be fast, indicating downhill or near-downhill folding. HP35 (chicken Villin headpiece, a 35-residue helical three-helical bundle) [28,29] and BdpA (B domain of staphylococcal protein A, a 58-residue three-helical bundle) [42] are examples of this category. The molecular rate coefficient is > $(100\ ns)^{-1}$ and the slow phase ranges from several μs for mutants with a low activation barrier, to sub-μs for mutants with activation barriers <2 *kT*. For HP35 and BdpA, k_m probably corresponds to helix-coil equilibration. Fig. 5B shows free energy surfaces and Langevin dynamics simulations for the latter case, which accurately fit the experimental data. An interesting observation can be made for the fast fluorescence phase of BdpA. The nonmonotonic fluorescence observed experimentally was attributed to laser-induced cavitation during the temperature jump experiment. While this is entirely possible, it is somewhat unlikely that a transmission experiment (IR) would be unaffected by cavitation, while detection of more isotropically distributed fluorescence would be. Our LD simulation data perfectly fit the non-monotonic fluorescence data and monotonic IR data simultaneously, simply by switching their probe function $S(x)$ at different position along the reaction coordinate (Fig. 5B).

A third category of proteins Fig. 5C are explained by the building block in Fig. 3B, modulated by roughness from Fig. 3A. EnHD (Engrailed homeodomain, a 61-residue three helix bundle) has more complicated folding kinetics as a result [43]. Three phases are observed, and our model can account for all of them (Fig. 5C). IR shows an ultra fast phase of 100 ns. Both IR and fluorescence show a fast phase of ≈2 μs and a 'slow' phase of ≈20 μs. In our view, a large shift in secondary structure content of the unfolded state results in the ultrafast IR phase, providing an upper limit for k_m. Subsequent 2 μs relaxation in the unfolded free energy well rapidly produces a more compact unfolded state following downhill diffusion over landscape roughness <2 *kT*. Finally, the molecule folds over a small barrier (≈4 *kT*) in 20 μs. Phosphoglycerate kinase, the first protein whose experimental data was analyzed in terms of downhill folding, is another example of an initial downhill step followed by activated folding [10]. It switches from slow exponential, to fast non-exponential, back to slow exponential kinetics, as the temperature is raised from cold denaturation, to maximum stability, to heat denaturation. This was interpreted as downhill formation of a folding intermediate surrounded by two-state formation of the intermediate under stress. The native state of PGK forms in ms to seconds over a barrier of >6 *kT*.

A final category of proteins folds very fast, always with a single or near-single exponential decay over the measured temperature range (Fig. 5D). Peptides and proteins in this category include a mutant of the GA module of an albumin binding domain (K5I/K39 V, folds in 1 μs) [26], the designed Trp-cage (folds in 1 μs) [44], and the artificial beta-strand protein DpDp-II (relaxation time 140 ns) [27]. These proteins almost reach the predicated folding speed limit and show very little temperature dependence of folding rates, a sign of likely downhill folding over the full temperature range (Fig. 3E). But to further confirm the single-well folding scenario, more thermodynamics measurements with different probes need to be tested. So far, only a mutant of lambda repressor has been shown to have probe-dependent thermodynamics at the melting temperature, a separately measured molecular rate of $(1.5\ \mu s)^{-1}$, and weakly temperature-dependent kinetics characteristic of at least incipient downhill folding at the melting transition [24]. Probe-dependent thermodynamics of BBL [17,31] also makes that molecule a likely candidate for fast and only weakly temperature-dependent folding kinetics.

5. Summary and conclusions

Downhill protein folding allows bulk experiments to directly elucidate the complicated protein folding free energy landscape, by revealing both local diffusional dynamics and global activated kinetics. When the overall barrier is reduced below 3 *kT*, the subpopulation outside the unfolded and native wells becomes significant (>5%). The signal from the fast relaxation of these subpopulations can be observed by fast kinetics measurement, such as laser induced T-jump experiments. As a consequence of very low barriers, folding kinetics or mechanisms are very sensitive to the sequence modifications or solvent conditions that stress the protein. The observed folding kinetics also depend on the measurement probes chosen, which may have different response functions $S(x)$ along the reaction coordinate x.

The molecular rate from Eq. (1) (or 'prefactor' when there is a barrier) can be observed directly in downhill folding. It ranges from about 50 nanoseconds to microseconds, depending on the folding coordinate monitored, and the presence or absence of a strong bias on the free energy surface. Lambda repressor and WW domain have molecular rates of several μs, which originate from the diffusional motion over the rugged but not heavily biased transition region between the unfolded and native state. The fastest molecular rate coefficients observed for VHP, BdPA and EnHD, in the range of 70–300 ns, correspond to a strong free energy bias towards more structure in the unfolded well (e.g. helix formation). These time scales represent an upper limit for the molecular rate under weakly biased conditions. Large proteins such as PGK have a much slower molecular rate, corresponding to higher internal friction and longer diffusive loop search. Single molecule experiments on a 66-residue cold shock protein reveals k_m to be at most $(0.4\ \mu s)^{-1}$ [37], confirming the values estimated from ensemble fast kinetics measurements [16,21].

The experimental relaxation kinetics of all current identified incipient downhill and downhill folders can be explained quantitatively with our Langevin dynamics model. One or another combination of five basic free energy building blocks accounts for the molecular and activated folding kinetics, their time scales, and their relative amplitudes, including inverted phases, while avoiding extensive fine-tuning of surface parameters. To identify downhill protein folding, it is necessary to do a series of kinetics or thermodynamics measurements as a function of protein stress.

Acknowledgments

This work was supported by grant MCB 0613643 of the National Science Foundation, and by a William and Janet Lycan Professorship.

References

[1] M. Gruebele, Comptes Rendus Biol. 328 (2005) 701.
[2] J.D. Bryngelson, J.N. Onuchic, N.D. Socci, P.G. Wolynes, Proteins-Struct. Funct. Genetics 21 (1995) 167.

[3] M. Jager, Y. Zhang, J. Bieschke, H. Nguyen, M. Dendle, M.E. Bowman, J.P. Noel, M. Gruebele, J.W. Kelly, Proc. Natl. Acad. Sci. USA 103 (2006) 10648.
[4] S.E. Jackson, A.R. Fersht, Biochemistry 30 (1991) 10428.
[5] A. Ikai, C. Tanford, Nature 230 (1971) 100.
[6] H.A. Kramers, Physica 7 (1940) 284.
[7] B.J. Berne, in: P. Hänggi, G.R. Fleming (Eds.), Activated Barrier Crossing: Applications in Physics, Chemistry and Biology, World Scientific, Singapore, 1993, p. 82.
[8] W.Y. Yang, M. Gruebele, Nature 423 (2003) 193.
[9] R. Metzler, J. Klafter, J. Jortner, Proc. Natl. Acad. Sci. USA 96 (1999) 11085.
[10] J. Sabelko, J. Ervin, M. Gruebele, Proc. Nat. Acad. Sci. USA 96 (1999) 6031.
[11] P.W. Fenimore, H. Frauenfelder, B.H. McMahon, F.G. Parak, Proc. Nat. Acad. Sci. USA 99 (2002) 16047.
[12] T. Zwier, J. Phys. Chem. 110 (2006) 4133.
[13] H.R. Ma, M. Gruebele, J. Comput. Chem. 27 (2006) 125.
[14] S.J. Hagen, Proteins-Struct. Funct. Bioinform. 68 (2007) 205.
[15] M. Knott, H.S. Chan, Proteins-Struct. Funct. Bioinform. 65 (2006) 373.
[16] W.Y. Yang, M. Gruebele, Biophys. J. 87 (2004) 596.
[17] M.M. Garcia-Mira, M. Sadqi, N. Fischer, J.M. Sanchez-Ruiz, V. Munoz, Science 298 (2002) 2191.
[18] A.N. Naganathan, U. Doshi, V. Munoz, J. Am. Chem. Soc. 129 (2007) 5673.
[19] H.R. Ma, M. Gruebele, Proc. Natl. Acad. Sci. USA 102 (2005) 2283.
[20] F. Liu, D.G. Du, A.A. Fuller, J.E. Davoren, P. Wipf, J.W. Kelly, M. Gruebele, Proc. Natl. Acad. Sci. USA 105 (2008) 2369.
[21] S.J. Hagen, J. Hofrichter, A. Szabo, W.A. Eaton, Proc. Natl. Acad. Sci. USA 93 (1996) 11615.
[22] D.T. Leeson, F. Gai, H.M. Rodriguez, L.M. Gregoret, R.B. Dyer, Proc. Natl. Acad. Sci. USA 97 (2000) 2527.
[23] H.S. Chung, Z. Ganim, K.C. Jones, A. Tokmakoff, Proc. Natl. Acad. Sci. USA 104 (2007) 14237.
[24] F. Liu, M. Gruebele, J. Mol. Biol. 370 (2007) 574.
[25] Y. Zhu, D.O.V. Alonso, K. Maki, C.Y. Huang, S.J. Lahr, V. Daggett, H. Roder, W.F. DeGrado, F. Gai, Proc. Natl. Acad. Sci. USA 100 (2003) 15486.
[26] T. Wang, Y.J. Zhu, F. Gai, J. Phys. Chem. B 108 (2004) 3694.
[27] Y. Xu, P. Purkayastha, F. Gai, J. Am. Chem. Soc. 128 (2006) 15836.
[28] J. Kubelka, W.A. Eaton, J. Hofrichter, J. Mol. Biol. 329 (2003) 625.
[29] J. Kubelka, T.K. Chiu, D.R. Davies, W.A. Eaton, J. Hofrichter, J. Mol. Biol. 359 (2006) 546.
[30] M.J. Parker, S. Marqusee, J. Mol. Biol. 293 (1999) 1195.
[31] M. Sadqi, D. Fushman, V. Munoz, Nature 442 (2006) 317.
[32] W.Y. Yang, J.W. Pitera, W.C. Swope, M. Gruebele, J. Mol. Biol. 336 (2004) 241.
[33] M. Gruebele, Proteins-Struct. Funct. Bioinform. 70 (2008) 1099.
[34] S.S. Cho, P. Weinkam, P.G. Wolynes, Proc. Natl. Acad. Sci. USA 105 (2008) 118.
[35] H. Nguyen, M. Jager, A. Moretto, M. Gruebele, J.W. Kelly, Proc. Natl. Acad. Sci. USA 100 (2003) 3948.
[36] C. Dumont, Y. Matsumura, S.J. Kim, J.S. Li, E. Kondrashkina, H. Kihara, M. Gruebele, Protein Sci. 15 (2006) 2596.
[37] D. Nettels, I.B. Gopich, A. Hoffmann, B. Schuler, Proc. Nat. Acad. Sci. USA 104 (2007) 2655.
[38] T.V. Pogorelov, Z. Luthey-Schulten, Biophys. J. 87 (2004) 207.
[39] G. I. Nettels D, Hoffmann, A, Schuler B, PNAS 104, 2007, 5.
[40] B. Schuler, E.A. Lipman, W.A. Eaton, Nature 419 (2002) 743.
[41] G.S. Hammond, J. Am. Chem. Soc. 77 (1955) 334.
[42] D.M. Vu, E.S. Peterson, R.B. Dyer, J. Am. Chem. Soc. 126 (2004) 6546.
[43] T.L. Religa, C.M. Johnson, D.M. Vu, S.H. Brewer, R.B. Dyer, A.R. Fersht, Proc. Natl. Acad. Sci. USA 104 (2007) 9272.
[44] M.R. Bunagan, X. Yang, J.G. Saven, F. Gai, J. Phys. Chem. B 110 (2006) 3759.

Feng Liu was born on December 19th, 1975. He grew up in China and attended Shan Dong University where he obtained a Bachelor's degree in physics. Then he went to Beijing University and graduated with a Master of Engineering degree in 2002. After obtaining a Master's degree in Physics at University of Pittsburgh in 2004, he enrolled in the Center of Biophysics and Computational Biology at the University of Illinois, studying protein folding in Gruebele group. He has published 11 papers, including two on downhill protein folding in J. Mol. Biol. and PNAS.

Martin Gruebele was born in Stuttgart in 1964, and lived in Austria and Spain before moving to Berkeley, where he obtained his BS and PhD at the University of California. He went on to do femtochemistry in the lab of Ahmed Zewail, and moved to the University of Illinois in 1992, where he is currently William and Janet Lycan Professor of Chemistry, Physics, and Biophysics and Computational Biology. He is a Fellow of the American Physical Society and of the Biophysical Society, as well as recipient of the Coblentz and Wilhelm Bessel Awards, among others. His research focuses on protein and RNA folding, vibrational energy flow in molecules and its control, and single molecule absorption spectroscopy detected by scanning tunneling microscopy.