Biophysical Journal Volume 106 April 2014 1729–1740

The Role of High-Dimensional Diffusive Search, Stabilization, and Frustration in Protein Folding

Supreecha Rimratchada,[†] Tom C. B. McLeish,[‡*] Sheena E. Radford,[§] and Emanuele Paci[†‡*]

[†]School of Physics and Astronomy, University of Leeds, Leeds, United Kingdom; [‡]Department of Physics and Biophysical Sciences Institute, Durham University, Durham, United Kingdom; and [§]Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, United Kingdom

ABSTRACT Proteins are polymeric molecules with many degrees of conformational freedom whose internal energetic interactions are typically screened to small distances. Therefore, in the high-dimensional conformation space of a protein, the energy landscape is locally relatively flat, in contrast to low-dimensional representations, where, because of the induced entropic contribution to the full free energy, it appears funnel-like. Proteins explore the conformation space by searching these flat subspaces to find a narrow energetic alley that we call a hypergutter and then explore the next, lower-dimensional, subspace. Such a framework provides an effective representation of the energy landscape and folding kinetics that does justice to the essential characteristic of high-dimensionality of the search-space. It also illuminates the important role of nonnative interactions in defining folding pathways. This principle is here illustrated using a coarse-grained model of a family of three-helix bundle proteins whose conformations, once secondary structure has formed, can be defined by six rotational degrees of freedom. Two folding mechanisms are possible, one of which involves an intermediate. The stabilization of intermediate subspaces (or states in low-dimensional projection) in protein folding can either speed up or slow down the folding rate depending on the amount of native and nonnative contacts made in those subspaces. The folding rate increases due to reduced-dimension pathways arising from the mere presence of intermediate states, but decreases if the contacts in the intermediate are very stable and introduce sizeable topological or energetic frustration that needs to be overcome. Remarkably, the hypergutter framework, although depending on just a few physically meaningful parameters, can reproduce all the types of experimentally observed curvature in chevron plots for realizations of this fold.

INTRODUCTION

Protein folding is a dynamical process occurring in a vast conformation space as a result of a protein's many degrees of freedom. Mere random search cannot yield the native state within physiologically relevant timescales, often referred to as Levinthal's paradox (1,2). Although from a physical point of view the solution of the paradox is obvious: conformations are not energetically equivalent, i.e., the energy surface is not flat and the search is a biased one, different models and theories have been proposed, which illustrate in different ways how the paradox is overcome in practice. The diffusion-collision model (3,4) reduces the search space to those of preformed microdomains searching for the correct native contacts. The nucleation-condensation model (5) focuses on the cooperative nature of native contact formation. Those contacts produce a folding nucleus that subsequently leads to the condensation of the fully formed native structure. In the nucleation-condensation model a preformed secondary structure is not assumed: secondary and tertiary structures can form concurrently. The diffusion-collision model, where the local cooperativity of secondary structure formation is more prominent, can be thought of as a limiting case of the nucleation-condensation model. In the landscape theory (6) the folding process takes place on a funnel-like energy surface. A protein need not follow a specified pathway to fold: each pathway and metastable state on the landscape could be occupied with well-defined probability. This statistical description of protein folding offers an opportunity to explain and qualitatively unify various scenarios of folding mechanisms (7,8).

Taken at face value, it is hard to see how the continuous, and apparently long-range, energetic correlations implicit in a folding funnel arise. A continuous energy dependence on the funnel height with any spatial ordinate implies a continuous spatial energy gradient, or force. This in turn invokes interactions that extend across substantial regions of the protein's spatial extend when unfolded. The forces involved in the folding process are themselves short range in nature (van der Waals and screened electrostatic). There is no candidate for the origins of long-range forces that steer a protein toward the native state along a funnel induced by pairwise interresidue forces alone. It is not as frequently restated as it should be that the shape of the folding funnel actually emerges by projecting the trajectory in the high dimensional conformation space onto a few (typically just one or two) degrees of freedom. Although the picture of very low-dimensional folding funnels grasps the overall structure of folding dynamics, the extreme dimensional projection that

Submitted October 11, 2013, and accepted for publication January 31, 2014.

*Correspondence: t.c.b.mcleish@durham.ac.uk or e.paci@leeds.ac.uk

Editor: David Eliezer.

0006-3495/14/04/1729/12 $2.00

http://dx.doi.org/10.1016/j.bpj.2014.01.051

results in a folding funnel tends to obscure the consequence of the very short range of the true intramolecular interactions: that much of the folding trajectory actually occupies subspaces that are energetically flat.

An alternative representation was previously proposed (9) that captures an explicit accounting of a protein folding process via diffusional search on largely flat multidimensional conformation subspaces to find a sequence of multidimensional energetic wells, dropping sequentially onto the next metastable subspace of states. These energetic wells employed during the folding process emerge from the short-range intrachain favorable interactions, typically nonnative, between partially formed sections of the protein. Such a dimensional reduction process repeats itself until the native state is reached. Hence, the search time is drastically reduced by the presence of intermediate subspaces of decreasing dimensionality (see Fig. 1).

The overall folding time is determined typically by the dimension of the largest intermediate search space (this is because the search time depends exponentially on the dimension of the search space). By taking care not to project out too many conformational degrees of freedom in this way, it is possible to relate folding times to the hypergutter structure of the intermediate-dimensional folding pathway. The funnel picture emerges from the projection along the progress of this process constituted by the concatenation of the intermediate search subspaces. The hypergutter model is therefore consistent with an interpretation of the folding funnel model in which the vertical axis corresponds to the free energy of the protein given the macrostates specified by its horizontal axes. The dimensional reduction on successive searches generates an entropic reduction corresponding to the narrowing of the funnel. It is an essential consequence of this approach, however, that the funnel surfaces should be understood as discrete steps, corresponding to the successive restriction to smaller search subspaces. Moreover, it is within this projection that intermediate subspaces appear as intermediate states. Nonnative interactions have an important role in stabilizing the intermediate subspaces—when they do they are able to accelerate folding, and constitute a way of generating the highly cooperative folding pathways observed (10). An enhancement of folding rate from limited nonnative interactions has been noted previously in theoretical and simulation studies (11). This work is consistent with those findings, but also goes some way toward explaining the origin of the effect, which is the restriction of the search-space for native contacts that nonnative contacts create. Nevertheless, nonnative interactions may also lead to topological or energetic frustration that could slow down the folding process.

In this work, we extend the hypergutter model (9) through a detailed illustration of its application, showing how closed-form expressions for both folding rates and chevron plots may be obtained once the short-range interactions have been parameterized. We also show that frustration induced by nonnative interactions plays a crucial role in the folding mechanism. Depending on the relative strengths of dimensional reduction and frustration, which work to affect the folding kinetics in opposing ways, different classes of experimental behavior emerge. We find that this simple physical model can reproduce all of the types of chevron plots seen experimentally while their shape is controlled by physically meaningful parameters.

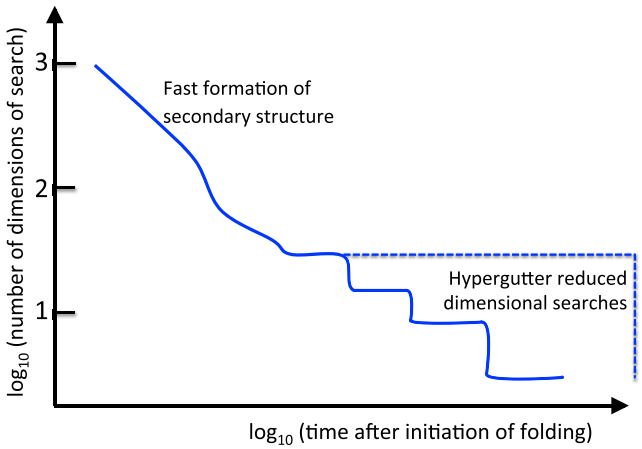


FIGURE 1 Schematic diagram of the folding process, and the emergent funnel in the space of remaining nonnative degrees of freedom (*dimensions of search*). After most of the fine-grained degrees of freedom form into secondary structure, the remaining ~10 degrees of freedom may either suffer an extremely long simultaneous search (*dashed line*), or fold exponentially faster via an energetically stabilized series of lower-dimensional subspaces (*solid line*), whose search times correlate with their dimension. To see this figure in color, go online.

HIGH-DIMENSIONAL SEARCH, HYPERGUTTER, AND FRUSTRATION

A typical illustration of a free energy surface for folding is that of a one- or two-dimensional funnel. Such a picture is appealingly simple but of little practical use, and is sometimes misleading. Proteins fold in a space with as many dimensions as the protein's degrees of freedom. One way to reduce the dimensionality of a search space is by preforming secondary structure (there are others, such as molten-globule formation). Local and fast secondary structure formation greatly reduces the accessible conformation space. Nevertheless, the search for the native tertiary contacts still occurs in a high-dimensional space (of order 10 dimensions) where the dimensionality is determined by the mutual translational and rotational degrees of freedom of those helices and strands. We argue that if the dimensionality is kept of this order, most of the energy landscape is flat as a result of this tertiary contact search, which is largely nonenergetic except when interacting atoms are in close proximity, for example, within the range of van der Waals or screened electrostatic interactions. Of course desolvation, which involves

an entropic effect, also plays a crucial role in the short-range intrachain interactions (10). But at the level of the hypergutter model all of this local physics can be accounted for in a set of coarse-grained effective energy parameters for intrachain interactions (see below). The hypergutter model does not assume the secondary structures to be fully formed but only that the dimensions involved in secondary structure formation are explored much faster than diffusion occurs on dimensions relevant to tertiary contacts (12). The only requirement for the hypergutter coarse-graining to apply is that secondary structure elements can be treated as objects with spatial degrees of freedom that are well defined (see an example below). This is true even when there are strong internal fluctuations in their degree of secondary structure order. During the search in a subspace other degrees of freedom are not dynamically frozen but simply not relevant to the search in such subspace.

When a target space of lower dimension is reached (i.e., all required favorable contacts are simultaneously satisfied) the protein drops into the next metastable subspace or into the native state if all native contacts are formed. This target space at each stage of folding is relatively small in comparison with the search space (for example two subunits of a partially folded protein reduce mutual translational degrees of freedom by at least two orders of magnitude at the point they come into contact and stay together). This picture is also compatible with the flatness of the vast majority of conformation space, which arises from the highly cooperative nature of the folding process (10). In a one-dimensional free-energy representation of a search-space this looks like a drop from a cliff at the end of a plateau; in a two-dimension space it looks like a drop into a gutter at the edge of a horizontal surface. For the high-dimensional energy surface this becomes a high-dimensional gutter or hypergutter (9,13). The funneled landscape emerges from the concatenation of these searches and drops (see Fig. 1).

In lower-dimensional subspaces, short-range and heterogeneous interactions could introduce energetic frustration (where the energy landscape has no obvious ground state) and hinder the diffusive search (14). The energetic barriers are on a smaller scale than the free-energy barriers between subspaces, but nonetheless must be overcome in making search steps within the subspaces. Topological frustration can also arise from partially native or nonnative topology that needs to be broken before the search for native topology. Hence, frustration is a factor that can increase the folding time within more compact subspaces. An appropriate coarse-graining of this level of frustration results in a lower value of the effective diffusion constant for local conformational search within compact subspaces. We will therefore adopt a parameter (called f below), which serves to account for reduction in diffusion constant due to many small barriers or topological frustration.

To illustrate each of these effects we chose a three-helix protein as a model because of its simple topology.

THREE-HELIX BUNDLE AS A MODEL PROTEIN

As mentioned earlier the diffusive search process involves formation of nonlocal contacts. For a three-helix-bundle these nonlocal contacts occur between helices when brought into close proximity. One way to derive the appropriate hypergutter model for such a topology is to concentrate on helices as mutually diffusing subunits, assuming that their secondary structure forms early in folding. This coarse-graining is equivalent to a coarse-grained model of the protein as three rods (Fig. 2). Each rod represents a helix and has three angular degrees of freedom denoted by (φ, θ, ψ). φ is the rotational angle of conformation about the long axis along the rod. θ and ψ are respectively altitudinal and azimuthal angles of the direction of its long axis in space. By subtracting the three degrees of orientational freedom of the folded protein as a whole, we find the total number of degrees of freedom of this protein is six. Hence, even projecting the fast-folding secondary-structure degrees of freedom onto the helices themselves, the protein still folds in a nontrivial search space of six-dimensions.

Because the folding time for a diffusive search is exponential in the dimensionality of its space, there remain to this family of proteins different folding strategies (ways of partitioning the remaining six dimensions) with widely differing folding times. In our case we assume that the three-helix protein can fold either via a two-state fully simultaneous six-dimensional search route, or via a sequential, two-step process. The first step is a four-dimensional search and a collapse to a state with helices I and III in contact. The second stage is then a two-dimensional search for the native state by helices I and III on (θ_1, θ_3) space (Fig. 3). The latter route is motivated by the observation that two helices possessing hydrophobic stripes (15) can satisfy those interactions by making contact in any configuration in which the stripes are in contact—therefore constituting a two-dimensional subspace of mutual orientation and translation. In traditional language this two-dimensional subspace would be termed an intermediate state (though it is really a large subset of states, stabilized by nonnative

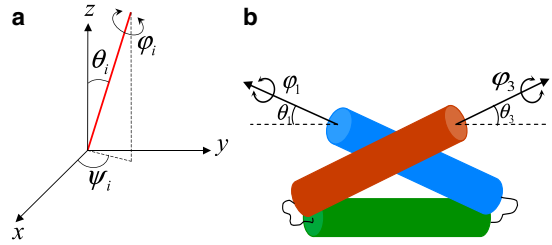


FIGURE 2 The model protein is a three-helix bundle coarse-grained into three rods. The rotational degrees of freedom of interest here are shown in (*a*) for a single rod in red. The three-rod protein consists of six internal degrees of freedom, three each from helix I (*blue*) and III (*red*). Rotating helix II (*green*) does not change the overall protein conformation. (*b*) Shows θ_1 and θ_3 in the final search, whereas φ_1 and φ_3 are not taking part. To see this figure in color, go online.

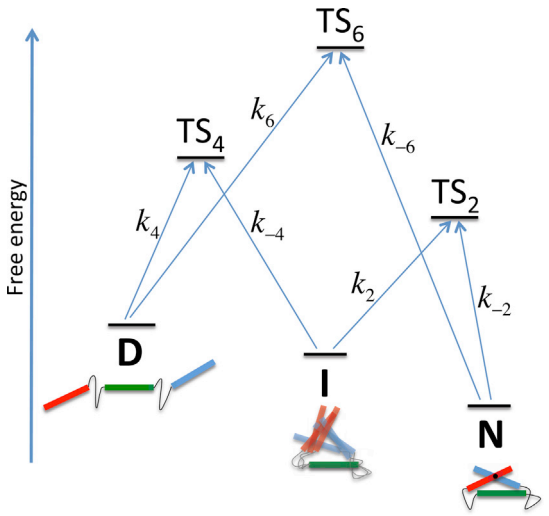


FIGURE 3 Free energy diagram of the three-helix bundle in hypergutter model. *D*, *I*, and *N* are denatured, intermediate, and native states, respectively. TS_d is the transition state after *d*-dimensional search. Arrows indicate transitions with rate k_d for the folding process. To see this figure in color, go online.

interactions). The two folding routes determined by the mutual configurational search of the helices cover a large number of pathways in the full conformation space. This simple example is useful to stress the importance of, and the relationship between diffusive search, stabilizing energy and frustration. It might be objected that we have not included paths in which helix I and II, or II and III first meet. However, providing these configurations are not stabilized by nonnative interactions in the same way, these random moves within the initial search space are in fact counted in our folding rate calculations; they just do not require any special labeling. A different scenario would arise if these off-pathway configurations were stabilized, as such partially misfolded states would then need to unform for folding to ensue.

This kinetic framework can be represented as two parallel pathways. The barriers that arise from a diffusive search with the folding rate of k_d in *d*-dimensional space are essentially entropic. The unfolding barriers are characterized by the stability of the favorable contacts formed after the end of a search. To our knowledge, this high-level representation is not, of course, novel. The advantage of deriving it from the hypergutter search model is that the parameters of the one-dimensional landscapes can be calculated in terms of the physical parameters of energetic interaction and frustration, and the search times at each stage can be estimated from the dimensionality of their corresponding subspaces.

NONNATIVE CONTACTS AND FRUSTRATION

In the hypergutter model the whole conformation space can be partitioned into many subspaces. If contacts stabilizing each intermediate subspace are solely native the protein will fold step by step. On the other hand favorable nonnative contacts in any intermediate subspace, although enhancing folding through dimensional reduction, may at the same time introduce an extra barrier for reaching the next subspace in addition to the diffusive search barrier. Such nonnative contacts need to be eliminated before the final folding process can proceed. Compact intermediates either with or without favorable nonnative contacts may additionally create energetic frustration that will in turn modify the diffusion coefficient (typically reducing it) in that subspace. In some cases, native contacts made in intermediate states can also prevent the reorganization of the intermediate. Hence, those contacts also need to be broken. For the sake of simplicity, we call all of those contacts nonnative. The hypergutter model accommodates these effects by introducing an extra barrier surrounding the target subspace of any diffusive search in an intermediate subspace that contains nonnative contacts. In our parallel-pathway kinetics this extra barrier is expressed as fraction (f) of intermediate state stability (ε_4). We allow a small energetic barrier of height $f \cdot \varepsilon_4$ to inhibit the search in the two-dimensional subspace. This could be interpreted as the effective fraction of contacts that need to be broken before the protein reaches the folded conformation.

Three scenarios are here considered, i.e., when i), $f = 0$; ii), $0 < f < 1$; and iii), $f = 1$. Case (i) represents a form of hierarchical folding where energy decreases monotonically at each folding step. An intermediate search-subspace or hypergutter can speed up the folding process by reducing the rate-determining barrier from a six-dimensional to a four-dimensional one. In case (ii) the benefit from a reduced barrier can still overwhelm the nonnative interactions that need to be broken. However, if the intermediate state becomes too stable there will be more frustration, i.e., a too sticky intermediate state can slow down the folding. Topological frustration is also possible if contacts stabilizing the intermediate prevent the search for the native state. This is more likely to occur in a three-helix bundle with longer helices (16). Finally, in case (iii) where $f = 1$ all of the contacts stabilizing the intermediate state must be broken to reach the native state. This means that all native contacts must be satisfied at once, i.e., a total cooperative folding, as in the six-dimensional route to reach the native state. Such a totally misfolded intermediate is normally considered to be off-pathway (17). It is effectively off-pathway also in a sense that it is acting as a nonproductive kinetic trap.

METHODS

Our goal is to derive closed expressions for the folding rates under different conditions of energetic stability of the intermediate subspaces. To do this we coarse-grain the many interresidue and solvent-mediated interactions into an effective energy, ε_d, for each *d*-dimensional subspace of configurations. We also simplify the search geometry of each subspace to the diffusion of a point-particle within a hypersphere with a dimensionless (in terms of target size) radius of the search-space $\sigma = (R/R_N)$ set equal to the

dimensionless angular search space of the angular variables of the helices. The target of the search is at the origin of the space. This is not as unreasonable as it might appear at first glance: the overwhelmingly dominant controlling parameter on folding rates is the number of dimensions that need to be searched within (this parameter is exponentiated in the folding rate—see Eq. 4 below) rather than their topology (e.g., *d*-sphere, *d*-torus, which determines only the prefactor to the search rate expression) (9). Furthermore, the typical dimension of all coarse-grained search spaces is of the order of the protein globule size, and the target spaces of the dimensions of a single residue, so they are not in any case distributed widely.

Diffusion in *d*-dimensional space

We model the protein folding process as diffusion in the space of conformations specified by a given set of degrees of freedom. Each point in this space represents a conformation of the protein. Generalizing an idea introduced by Bicout and Szabo (18) we assume that diffusion takes place in a *d*-dimensional sphere of radius R and the native state is represented by a sphere with same center and smaller radius R_N. The degrees of freedom in this hyperspherical search problem may map onto either translational or rotational degrees of freedom in a coarse-grained model of a protein. For example, we will use the result for a search in four-dimensional space to estimate the search time in the (φ_1, φ_3, ψ_1, ψ_3) space of helices I and III in our example. The diffusion equation in a space of arbitrary dimension d is

$$\frac{\partial P(r,t)}{\partial t} = D\frac{1}{r^{d-1}}\frac{\partial}{\partial r}r^{d-1}\frac{\partial P(r,t)}{\partial r}, \qquad (1)$$

where $P(r,t)$ is the probability density of finding the diffuser at distance r from the origin at time t in this space and D is the diffusion coefficient. As is standard in diffusion theory this construction represents either the flux of (noninteracting) proteins or a single protein diffusing in the hyperspace. An absorbing boundary condition is applied on the (small) native hypersphere at radius $R_N = R/\sigma$, i.e., $P(R_N,t) = 0$. By solving the previous equation in the steady state (i.e., $\partial P/\partial t = 0$) with this boundary condition the population distribution and the flux, and from those two the mean search time can be found.

Flux-over-population calculation

In this section, we recapitulate how solution of the high-dimensional search problem of the Diffusion in *d*-dimensional space section yields mean search times. By definition, the intrinsic rate of a given process for many independent particles is the flux for the process divided by the number of particles. Here, the protein folding process is modeled as a flux of proteins diffusing in a *d*-dimensional space from denatured to native state. The search dimension includes all relevant degrees of freedom for a particular search step. For simplicity, each search space is defined to be spherically symmetric. The outer (hyperradius R) and inner (hyperradius R_N) spheres represent denatured and native states, respectively. Hence, one can write the folding rate as

$$k_f = \frac{j_f}{N}, \qquad (2)$$

where flux and population number are defined, respectively, as

$$j_f = \left(-D\frac{\partial P}{\partial r}\bigg|_{r=R}\right)S \quad N = \int_R^{R_N} dr^d P(r). \qquad (3)$$

The hypersurface area for the hypersphere in dimension d and of radius R is $S = s_d R^{d-1}$, where s_d is a prefactor depending on d. The integral of population density over hypervolume gives the total number of proteins in the folding population. The incoming flux of the steady-state population enters the hypersurface at R and gets absorbed at R_N. The integral is over the population density at steady-state or time-independent $P(r)$. For this method to work $P(r)$ must be given or obtained from, for example, diffusion or Smoluchowski equations. As a result one can write the diffusive rate k_d in d-dimensional space in terms of $\sigma = R/R_N$, which characterizes the (dimensionless) size of the search space. Equivalently, $d \cdot k_B \ln\sigma$ counts the change in configurational entropy arising from a variable on the completion of a search to its target within its diffusional subspace. The search-rate has a simple relationship with its search volume as explained in the Results section below. σ is also related to the volume entropy in that space and, in the case of the application to protein folding, would be naturally controlled by the denaturant concentration as developed in the Parameterization of denaturant dependence [*D*] section.

Kinetic frameworks

We calculate folding rates for the complete hypergutter by combining subrates for each subspace. The most significant discriminator of the search rates is the dimension of the corresponding subspace. From the kinetic scheme of connected partial folding and unfolding pathways as illustrated in Fig. 4, the overall folding and unfolding rates can be derived and written in terms of underlying rate constants by means of the flux-over-population method.

This method is compatible with Cleland's method (19) and Waley's flux method (20). Both of them involve the sum of inverse conductance (rate constants) or the resistance (reaction times) along a steady current (flux or reaction velocity). Applying the method to protein folding requires only a careful choice of definition of folding rate, amounting to assigning the intermediate population with the native state (on the one hand) or with the denatured state (on the other). We take the latter choice, defining the folding rate as the net flux into the native state from the ensemble of other states under dynamic equilibrium. In the calculation of the folding rate this definition is equivalent to placing the absorbing boundary just outside the native state, not immediately outside the denatured state—the folding rate is calculated from the flux across this absorbing boundary. This is a natural choice from the point of view of experiment, because most probes reporting folding in time-dependent studies detect and report the native, but not intermediate, states

Parameterization of denaturant dependence [*D*]

The hypergutter model is defined by three model parameters: σ, the dimensionless size of search space, ε_d, the stabilizing energy and f, the

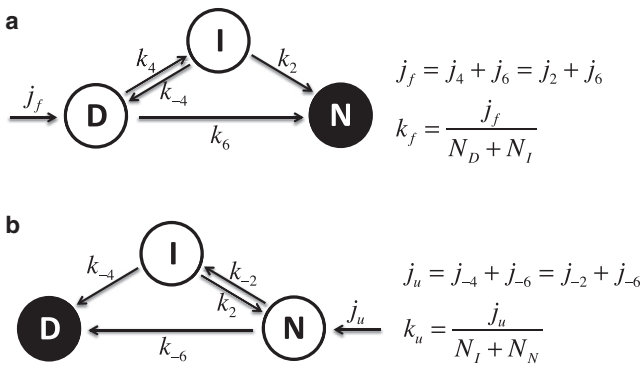


FIGURE 4 Folding (j_f) and unfolding (j_u) proteins flux through the six-dimensional and (four+two)-dimensional routes. The kinetics is solved in the steady state where native (N) and denatured (D) states are absorbing boundaries (*zero population*), indicated by solid circles, for folding and unfolding routes, respectively. N_D and N_N are the denatured and native populations. N^f_I and N^u_I are the intermediate populations in the folding and unfolding reactions, respectively.

fraction of energetically favorable contacts in the intermediate subspaces, which need to be broken before final folding, generating the introduction of an enthalpic component to transition states, as underlined in (e.g., reference (10)). ε_d and the logarithm of σ (actually $d \cdot k_B \ln\sigma$, as above) can be considered as the energy and entropy, respectively, of their corresponding subspace. If we are to compute predicted chevron plots from the model, we need to link these parameters to denaturant concentration. Here, we take an approach inspired by Tanford's empirical relation (21), which is a consequence of the observation in protein denaturation experiments that the change in free energy is linear in denaturant concentration: we assume that the conformational entropy and energy depend linearly on the denaturant concentration for each subspace independently. This is a generalization of Tanford's parameterization for an entire folding process onto the folding subspaces of the hypergutter model. Denaturant has the effect of increasing the entropy but reducing the energy. Hence, we assume the relationships

$$\ln\sigma = \ln\sigma_0 + m_\sigma[D],\ \varepsilon_d = \varepsilon_{0(d)} - m_{\varepsilon(d)}[D],$$

where d is the dimensionality of the search space preceding each ε_d barrier and $[D]$ is the denaturant concentration. The zero subscript indicates the parameter in the absence of denaturant. σ and ε_d each have an associated m value, which, within the Tanford's parameterization, measures the response of such parameters to denaturant concentration. In the specific model of the three-helix bundle we examine here, we assume that the concentration of denaturant alters the strength, but not the fraction f, of nonnative contacts. We note also that, in the intermediate state in which the nonnative contacts are present, the structure is a highly fluctuating one, and that, although the number of nonnative contacts is approximately constant, the actual nonnative residues in contact at any moment sample the entire (two-dimensional) space of touching configurations of helices I and III. The chevron plots can be drawn by assigning values to the parameters σ_0, $\varepsilon_{0(d)}$, m_σ, $m_{\varepsilon(d)}$, f, and the prefactor $2D/R^2_N$.

RESULTS

Diffusion rates in various dimensions of search-space

By using the flux-over-population method the search rates (from denatured to native hyperspheres) in various dimensions can be calculated. For $d \geq 3$ a general form exists:

$$k_{d\geq 3} = \frac{d(d-2)D}{R_N^2}\sigma^{-2}\left(\sigma^{d-2} - \sigma^{-2} - \frac{d}{2}\left(1-\sigma^{-2}\right)\right)^{-1}. \quad (4)$$

The derivation is given in Appendix A in the Supporting Material. In high dimension and large σ limit the diffusion rate scale with σ as $k_d \propto \sigma^{-d}$ and illustrates the point that the search rate in a folding subspace is dependent on the relative sizes of the search and target spaces, and that the most sensitive determinant is the dimensionality of the search space.

The effect of the model parameters (f, σ, ε_d) on the rates and populations

From the kinetic framework introduced in the Kinetic frameworks section, the overall folding and unfolding rates can be calculated from the flux-over-population method. The detailed calculations are shown in Appendix B. The results are given below for the overall folding and unfolding rates in terms of the rates in subspaces of dimensions two, four, and six:

$$k_f = \frac{k_2^* k_4 + k_6\left(k_2^* + k_{-4}\right)}{k_2^* + k_4 + k_{-4}},\ k_u = \frac{k_{-2}k_{-4} + k_{-6}\left(k_2^* + k_{-4}\right)}{k_2^* + k_{-2} + k_{-4}}, \quad (5)$$

where $k_2^* = k_2 e^{-f\varepsilon_4}$ is the forward rate for achieving the native state from the intermediate. We include the possibility that this is slowed (relative to the search from the denatured state) by the requirement that a fraction, f, of the nonnative interactions responsible for ε_4 need to be broken on final folding. Similarly, it is also of interest to write the ratio of folding or unfolding flux along each pathway to the total population. These are the fractional rates, $k_{\varphi(route)}$, where the route is either the six- or (four+two)-dimensional one; $k_{-\varphi(route)}$ is the fraction of unfolding flux. These fractional rates take the following forms:

$$\begin{aligned} k_{\varphi(4+2)} &= \frac{k_2^* k_4}{k_2^* + k_4 + k_{-4}} & k_{-\varphi(4+2)} &= \frac{k_{-2}k_{-4}}{k_2^* + k_{-2} + k_{-4}} \\ k_{\varphi(6)} &= \frac{k_6\left(k_2^* + k_{-4}\right)}{k_2^* + k_4 + k_{-4}} & k_{-\varphi(6)} &= \frac{k_{-6}\left(k_2^* + k_{-4}\right)}{k_2^* + k_{-2} + k_{-4}} \end{aligned}. \quad (6)$$

The detailed balance condition that applies at steady state provides the constraint

$$K_d \equiv \frac{k_d}{k_{-d}} = \sigma^{-d} e^{\varepsilon_d}, \quad (7)$$

where k_{-d} is the unfolding rate (reversing the search) in a d-dimensional subspace. Another necessary constraint is the pathway-independence of the change in state variables (such as energy and entropy) under equilibrium conditions. This implies the equality of energy-change on each folding pathway, $\varepsilon_6 = \varepsilon_4 + \varepsilon_2 - f \cdot \varepsilon_4$. The entropy changed is also obviously conserved, i.e., $\ln(\sigma^{-6}) = \ln(\sigma^{-4}) + \ln(\sigma^{-2})$.

Apart from the reaction rates, the population fractions can also be calculated analytically in our model. These fractions can be monitored or titrated against the model parameters. We now consider the full equilibrium picture (i.e., not only the steady-state picture as depicted in the calculation of overall folding and unfolding rates) because the intermediates can be populated during both folding and unfolding transitions. The general condition for thermodynamic equilibrium imposed here is that of zero-net-flux, i.e., the folding and unfolding fluxes are equal in magnitude and in opposite directions. Such conditions are also imposed on each reaction pathway to satisfy detailed balance. Using the flux-over-population method as in the calculation of the rates,

including the equilibrium condition, the fractions of denatured, intermediate, and native states are (see Appendix C):

$$F_D = \frac{1}{1+K_4+K_6},\ F_I = \frac{1}{1+K_2^*+1/K_4},$$
$$F_N = \frac{1}{1+1/K_2^*+1/K_6}. \quad (8)$$

We want to explore the parameter space of the three-helix hypergutter model, in terms of its effect on transition rates and the stabilization of native, intermediate, and denatured states. However, there are constraints on the variation of the model parameters. First, the change of state variables (energy and entropy) from denatured to native states must be equal irrespective of the pathway. Second, forward and backward reactions from one state to another share the same transition state: forward and backward barriers cannot be varied independently. These constraints are introduced into all subsequent calculations. The energetic stability of the native state is simply described by the parameter ε_6. Stabilization of the native state would simply reduce the unfolding rate and has no effect on the folding rate. This can be achieved by increasing ε_6 alone and will not affect other states, a rather trivial case (see Fig. D1 in Appendix D). Therefore, we will explore the physics of energetic stability, search space, and frustration by stabilizing the intermediate and denatured states only, choosing a convenient value of $\varepsilon_6 = 15\ k_BT$ in all cases. Fig. 5 illustrates in diagrammatic form the effects of changing the physical parameters.

Energetic stability (ε_4): stabilization of the intermediate state

Here, we are interested in the existence and the stability of the intermediate state/subspace and how these affect the transition rates and the population of all states. Stabilization of the intermediate state also affects the stability of the transition state (TS_2) between the intermediate and the native states due to the constraints mentioned previously. In this case ε_2 decreases following the stabilization of the intermediate state, i.e., an increase in ε_4 at constant total energy (ε_6). Fig. 5 *a* shows that, at $f = 0$, the folding rate increases with ε_4 from its initial value k_6 until it reaches its maximum value k_4. For $f = 1$, the overall folding rate is k_6 at low ε_4 and decreases for larger ε_4. For intermediate f (0.5) the rate reaches a maximum for some value of the intermediate stability ε_4 (close to 10 k_BT with the current parameterization of the model). This crossover effect illustrates the balance between the increase in folding rate due to the presence of intermediates (and thus alternatives to the full search) and the reduction in rate due to the frustration induced by nonnative interactions stabilizing the intermediate.

Similarly, the unfolding rate for $f < 1$ initially increases, as shown in Fig. 5 *b*, when the intermediate is being stabilized. The rate-determining barrier shifts from ε_6 to ε_2 and eventually to ε_4 for higher ε_4. For $f = 1$ the unfolding rate never exceeds k_{-6}. As the intermediate state is stabilized,

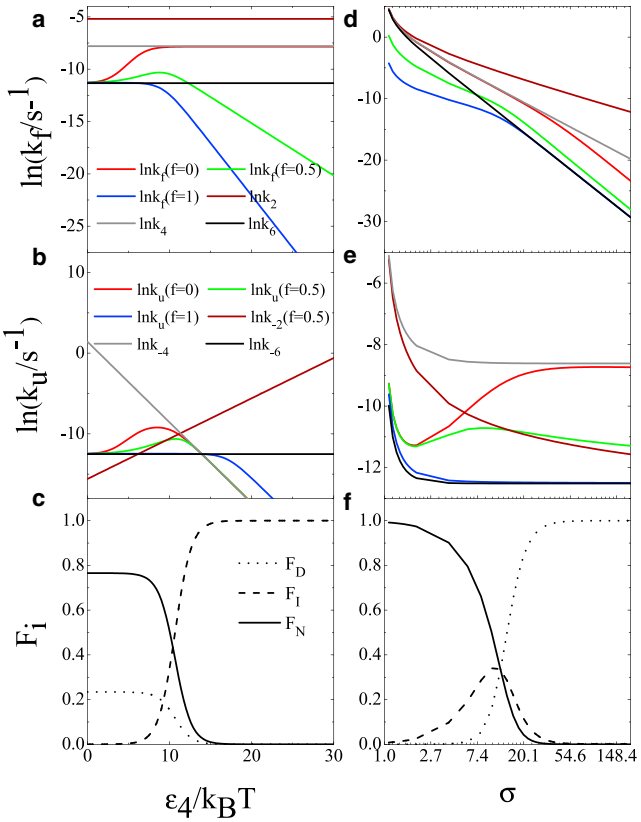


FIGURE 5 Folding/unfolding rates and fractions of equilibrium populations as a function of model parameters. $\sigma = 10$ in (*a*), (*b*), and (*c*). $\varepsilon_4 = 10\ k_BT$ in (*d*), (*e*), and (*f*). $\varepsilon_6 = 15\ k_BT$ in all cases. Each pair of (*a*,*d*), (*b*,*e*), and (*c*,*f*) has the same color and line coding. Note that σ is shown on logarithmic scale. To see this figure in color, go online.

the fractions of native and denatured states are reduced accordingly as shown in Fig. 5 *c*. Interestingly, the folding and unfolding rates already keep increasing below 5 k_BT where the intermediate species is not significantly populated. Therefore, this hidden high-energy intermediate, though undetected, can also help speed up the folding process. Note that the population distribution between states only depends on the relative free energy difference between them. Therefore, parameter f plays no role in the population plot, but affects only the kinetics.

Size of search space (σ): stabilization of the denatured state

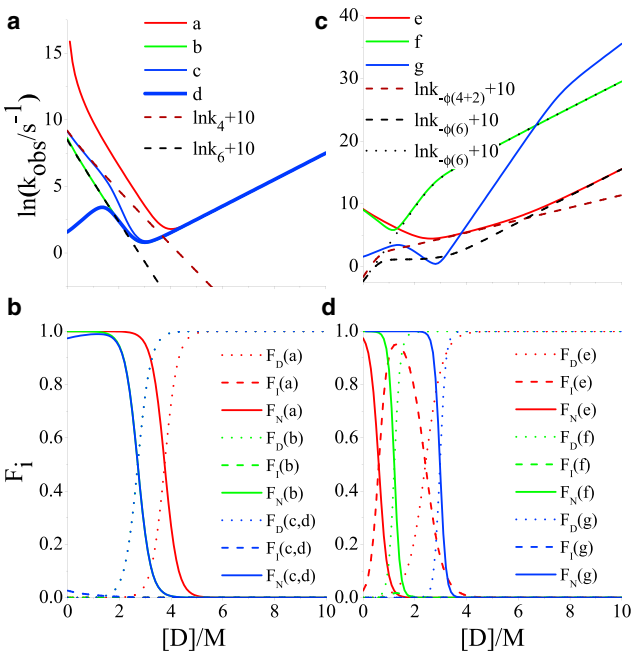
An increase of σ stabilizes the denatured state but also affects the stability of both the intermediate state and TS_4 due to the constraints mentioned previously. These effects can be shown by manipulating the barriers in Fig. 3 subject to the constraints. First, the native state free energy and the unfolding barriers are independent of σ. By increasing σ the intermediate state must be stabilized without affecting TS_2 because ε_2 is kept unchanged. As a result TS_4 will be stabilized because ε_4 is kept unchanged. Finally, the denatured state is also stabilized because k_4 and k_6 directly depend on σ. When the intermediate contains only native contacts,

i.e., $f = 0$, the expansion of the search space will hinder the folding rate by increasing the stability of both the denatured and intermediate states. As a consequence of this stabilization, for the larger search space, the folding flux will eventually go through the six-dimensional route irrespective of the value of f, as when ε_4 is kept unchanged the free energy of the intermediate two-dimensional space becomes inadequate to stabilize it; this effect is illustrated in Fig. 5 *d*. When the denatured and intermediate states are stabilized relative to the native state the unfolding rate inevitably increases at least initially as shown in Fig. 5 *e*. However, a very stable intermediate may slow down unfolding by acting as a kinetic trap. As a result the unfolding flux will behave like the folding flux by shifting to the six-dimensional route in the limit of very large search space. The initial stabilization of the intermediate also has the effect of trapping intermediate species. When the denatured state becomes more stable than the intermediate state in the large search space the latter species can no longer be trapped and eventually dies out as evident in Fig. 5 *f*.

On the curvature of chevron plots

In addition to its simplicity, an important feature of the hypergutter model is that it qualitatively reproduces any experimentally observed shape of chevron plots (see Fig. 6, *a* and *c*), by simply varying the model parameters: of importance, this is a consequence of coarse-grained protein physics, not direct curve fitting. The present model is compatible with the experimental probe generally used to obtain the chevron plots: tryptophan fluorescence quenching to probe the refolding/unfolding reactions mainly on the level of tertiary structure formation. Hence, the presumed diffusive search of helices, which determines the net folding/unfolding rates, is directly applicable to the fluorescence probe. Experimentally, chevron plots are found with or without curvature of either sign. From a theoretical perspective, the diffusion-collision model has also been used to predict the shape of the chevron plot of the domain of protein A (22). The observed diversity within chevron plots can be explained in terms of intermediates on the folding landscape (23). Here, we investigate the nature of chevron plot curvature on the folding and unfolding branch separately with increasing kinetic complexity from one-state, two-state, three-state, and a form of four-state kinetics where both on-pathway and off-pathway or totally misfolded intermediates are present. We choose to define the sign of curvature so that upward or downward curvature in a region of the chevron plot has a positive or negative second derivative of $\ln(k_{obs})$ with respect to $[D]$.

An upward curvature in the chevron plot (Fig. 6 *a*, *curve* a) is unusual. In the hypergutter model a protein folds downhill if the conformational search and the frustration become negligible, i.e., σ approaches 1 and $f = 0$, respectively; the upward curvature in the folding branch of the chevron plot



Curves	σ_0	f	$\varepsilon_{0(4)}$	$m_{\varepsilon(4)}$	$m_{\varepsilon(6)}$	Kinetic scenarios
a	1	0	0	3	1	Barrierless downhill
b	2	0	0	3	1	Two-state
c	2	0	10	3	1	On-pathway three-state
d	2	1	10	3	1	"Off-pathway" three-state
e	2	0	10	1	6	Parallel pathway shift
f	2	0	3	1	6	High-energy intermediate
g	2	1	10	3	6	"Four-state"

FIGURE 6 Chevron plots (*a*,*c*) and the fractions of equilibrium populations (*b*,*d*). $\varepsilon_{0(6)} = 15\ k_BT$; $m_\sigma = 0.5\ k_BT{\cdot}M^{-1}$; $\ln(2D/R^2_N) = 10\ \ln(s^{-1})$ in all cases except curve g where $\varepsilon_{0(6)} = 30\ k_BT$. To see this figure in color, go online.

arises from the singularity in the rate expressions when σ approaches 1 and the dominant folding process is the organization of side chains for tertiary contacts rather than the diffusive search to find those contacts. This also implies that barrierless proteins are generally small (small search volume) and minimally frustrated ($f = 0$).

By introducing the diffusive search ($\sigma = 2$) into the downhill scenario a V-shaped plot without curvature (Fig. 6 *a*, *curve* b) is obtained; this represents the typical two-state kinetics where the intermediate is absent ($\varepsilon_4 = 0$). With the presence of an intermediate stabilized solely by native contacts ($f = 0$), downward curvature on the folding branch becomes apparent. Curve c in Fig. 6 *a* illustrates this scenario for $\varepsilon_4 = 10\ k_BT$. In this case, forming an intermediate state can speed up the folding process by shifting the rate-determining barrier from the larger six-dimensional to the smaller four-dimensional one. By introducing nonnative contacts (increasing f) in the intermediate the curvature will become more downward.

In the limit of an intermediate stabilized solely by nonnative interactions ($f = 1$) the curvature is maximally downward as shown in Fig. 6 *a* curve d. This scenario corresponds to presence of a totally misfolded intermediate

where all contacts must be broken before the native state is reached. In the last two cases of $f = 0$ and $f = 1$ the intermediate state is only marginally populated at low denaturant concentration as shown in Fig. 6 *b*.

In fact, by making the native state more stable the intermediate will totally disappear. In general, the population of folding species depends on the relative free energy of the species involved. Therefore, the curvatures in the chevron plots, which depend on the folding barriers, can still be observed without detectable intermediate species.

In those previous scenarios we assume that the native subspace is smaller, i.e., the native structure is more compact than the intermediate one. Accordingly, its response to denaturation is also weaker ($m_{\varepsilon(6)} < m_{\varepsilon(4)}$). However, if we create the opposite scenario, downward curvature on the unfolding branch can be observed as shown in Fig. 6 *c*. By making $m_{\varepsilon(6)} > m_{\varepsilon(4)}$ curve c in Fig. 6 *a* can be transformed into curve e in Fig. 6 *c*. The unfolding flux is initially on the (four+two)-dimensional route but shifts to the six-dimensional route at higher denaturant concentration. Initially the route with intermediate has a smaller rate-determining barrier ($\varepsilon_4 = 10\ k_BT$) than the other route ($\varepsilon_6 = 15\ k_BT$). However, the latter barrier decreases faster than the former one. Therefore, at higher denaturant concentration it is more favorable to unfold on the six-dimensional route. For a relatively higher energy intermediate ($\varepsilon_4 = 3\ k_BT$) the downward curvature on the unfolding branch can be found as shown in Fig. 6 *c* curve f. The combination of downward curvature on both the folding and unfolding branch is also possible as shown in Fig. 6 *c* curve g. This happens when a protein folds with an off-pathway (or totally misfolded intermediate in our model) but unfolds with an on-pathway intermediate. Curve g can be produced from curve d by reducing the compactness of the native state ($m_{\varepsilon(6)}$) relative to that of the intermediate state ($m_{\varepsilon(4)}$). In this case, we also need to stabilize the native state further ($\varepsilon_6 = 30\ k_BT$) to see curvature in both branches. Unsurprisingly, the population of intermediate species is more robust against denaturation in the case of curve e where the intermediate is more stable than that of curve f; this effect is clearly shown in Fig. 6 *d*. In addition, the intermediate is not significantly populated in the case of curve g even though curvature is present in the chevron plot. As argued earlier, this is the result of a stable native state relative to the intermediate state.

Finally, the folding branch in Fig. 6 *c* curve e and f can be turned upward by merely taking $\sigma = 1$. To the best of our knowledge none of these chevron plots has been observed experimentally. The remaining combination of curvatures, i.e., downward folding branch and upward unfolding branch cannot be obtained with any combination of the parameters. The downward curvature on the folding branch means an intermediate less compact than the native structure but an upward curvature on the unfolding branch requires the opposite. In addition, a monotonic stabilization of D, I, and N produces the expected chevron plots as shown in Fig. D2 of Appendix D. Therefore, from a theoretical perspective we have shown that the curvature on the chevron plots arises naturally and logically from pathway shift, frustration, and the relative response to the denaturant without any ad hoc exponential terms in the rate or nonlinear terms in the logarithm of the rate (24).

Curvature in chevron plots has been observed by simulating a native-centric Gō model (25) and arises from the presence of intermediates with nonnative topologies. A different native-centric HP model also produced a curved chevron plot (26). Both models assume that the increase in the denaturant concentration is equivalent to the reduction in the mean interaction energy (ε). Chevron plots can be produced from simulation data in a more systematic way through the so-called molecular transfer model (27). Gō and HP models only take into account the energetic part of the folding free energy to be parameterized in terms of denaturant concentration. The molecular transfer model relies on experimental data on the solvent-accessible surface area to evaluate the free energy change as a function of denaturant concentration. Although curved chevron plots have been previously obtained using atomistic models, with the hypergutter model one can do so only assuming that the (entropic) diffusive searches and (energetic) stabilizing contacts are linear functions of denaturant concentration.

DISCUSSION AND CONCLUSIONS

By thinking of the protein folding process as a partitioning of a high-dimensional space into several small-dimensional ones, one can also incorporate other folding models into the hypergutter picture presented here; for example, our approach can be seen as a generalization of the diffusion-collision model to incorporate nonspatial coordinates as in our three-helix bundle model. The hypergutter also offers an explanation of how the diffusive search can take place within a compact intermediate. Nonetheless, we do not assume fully preformed microdomains, but only require that the secondary structure such as helices forms faster than the tertiary structure, and at least to some intermediate degree of order. They can also be coupled to each other and form concurrently. However, a sufficient amount of secondary structure is needed to perform the diffusive search. In this way, the notion of a nucleation-condensation model is also part of our picture.

The high-dimensional picture of protein folding also offers an explanation for the origin of barriers in protein folding kinetics. Folding barriers are largely entropic as a result of the diffusive search in high-dimensional space but could also be energetic, e.g., when the intermediates are stabilized by nonnative contacts. Those contacts need to be broken before the native structure can be achieved. Such nonnative interactions can potentially create energetic or topological frustration. Energetic frustration arises from

the inhomogeneous interactions between secondary structure elements in the intermediate subspace. Topological frustration arises from the nonnative contacts that may lead to misfolding of the protein into a topologically different metastable state. On the other hand, in our model, the unfolding barrier is solely energetic (containing no conformational entropy) as a result of favorable native or nonnative contacts made in each folding step.

We also demonstrated that our kinetic framework with a coarse-grained molecular model reproduces the variety of chevron plots with various curvatures and in different combinations. The main features of folding kinetics can, in principle, be explained in terms of the folding flux along folding pathways. The presence of an intermediate shifts the folding flux to the one with the lower rate-determining barrier of a four- rather than six-dimensional search. The intermediate states can be stabilized by both native and nonnative interactions and nonnative interactions can accelerate protein folding by forming intermediates, which partition the search space. Thus, the rate-determining folding barrier is smaller on this alternative pathway. This effect is different from the accelerated rate arising from the presence of relatively high-energy intermediates according to previous theoretical studies (28). To the best of our knowledge only curves b, c, and f in Fig. 6 have been experimentally observed for three-helix proteins. Most small three-helix proteins such as the B domain of protein A (29,30) fold via two-state kinetics, with a V-shaped chevron plot (curve b). Such folding kinetics corresponds to the full six-dimensional search in our model. However, computer simulation suggests alternative folding pathways with various intermediate states (31,32): for example, helix I and II or II and III could first form a hairpin followed respectively by the docking of helix III and I onto those motifs.

The folding of a four-helix immunity protein Im7, which folds via an on-pathway intermediate, could be considered within this class of three-helix folding because the small helix III only docks onto the rest of helices the very last stage. The typical rollover seen for Im7 and mutant Im9 is plausibly a consequence of $f = 0$ or $0 < f < 1$ (33,34). An on-pathway intermediate is also detected in three-helix bundles such as engrailed homeodomain and mutated c-Myb (35). For $f = 0$ one can potentially see the shift from six- to (four+two)-dimensional search in the Fig. 6 *a* curve c. On the other hand, a relatively high energy intermediate produces downward curvature on the unfolding branch as shown in Fig. 6 *c* curve f. This type of curvature has been observed for spectrin, a three-helix bundle protein (36). In our calculations in Fig. 6 *c* curves e, f, and g, $m_{\varepsilon(6)}$ is $>m_{\varepsilon(4)}$, i.e., the intermediate is more compact than the native structure. This is unusual but not impossible for our three-helix protein. For example, in the intermediate helix I and III might be in contact such that those interresidue contacts are better buried than those native ones. More recently the folding rates of spectrin domains R16 and R17 were increased by introducing nucleation sites along the core residues that guide the docking of helix A and C (37). The curvature is also present on the folding branch of mutated R17. Our model offers an explanation for both the increase in folding rate and the presence of curvature. They are, respectively, the result of shifting the rate-determining barrier and the transition state from full conformational search to a partial search with an intermediate state.

Curves a, d, e, and g have not been identified for the folding of three-helix proteins. The upward curvature on the folding branch (curve a) is unusual but has been recently observed in a natural downhill folder PTB1:4W (38), an α-β protein. The apparent upward curvature could be the result of having the broad midpoint of the two-state chevron plot close to the zero denaturant concentration similar to those plots found in some mutants of BBL, a small three-helix protein (39). As mentioned earlier, for small three-helix proteins, the downhill scenario could be achieved through the minimization of search space and the absence of nonnative interactions. For $f = 1$ the rollover effect is maximal according to our model (curve d) and is similar to the case of off-pathway kinetics as found in protein S6 (40), an α-β protein. For three-helix proteins such as the B domain of protein A or spectrin domain one could map this scenario onto the misdocking between helix I and helix III that need to be swapped before the protein can fold to the native state. For a relatively more stable intermediate, which is more compact than the native structure, upward curvature can be observed (curve e). Titin I27, a β-sandwich protein, exhibits this form of curvature (41). In that study the explanation was also in a form of parallel-pathway kinetics. In our model the unfolding flux shifts from (four+two)-dimensional route to the six-dimensional route at high denaturant concentration because the former contains the highly compact intermediate.

Finally, the downward curvature on both branches (curve g) can also be obtained from our model. This kind of chevron plot is found in apoflavodoxin (42), a mixed α-β protein. It folds via an off-pathway intermediate but unfolds via an on-pathway intermediate. At first sight the folding and unfolding fluxes of our three-helix model protein are both on the six-dimensional route (fractional rate plots not shown here); this is merely because the (four+two)-dimensional route contains the relatively stable and unfavorable off-pathway or the totally misfolded ($f = 1$) intermediate. We note that, although the class of chevron plots is very rich, there are, even in the case of three-helix bundles, different structures of nested folding spaces and their stabilization energies lead to similar plots. When the class of proteins is broadened, this many-to-one mapping will be enhanced further. This is, however, not a weakness of the model, but supports the long-established need for multiple experimental measures of the folding process. Indeed, titrating against salt, pH, and temperature provide independent measurements of the space, providing that the

stabilization energies can be parameterized accordingly. We note additionally that mutation studies are another route to explore hypergutter models, especially where these indicate folding-rate effects from nonnative interactions.

We have chosen to keep the effective diffusion constants uniform in value across the subspaces in which they apply. This might be thought questionable in light of results (43,44) showing that in low-dimension projection the diffusion constant needs to become coordinate dependent in some cases. This is especially true of very low-dimensional representations (one or two). Because we have retained a relatively high-dimensional representation (of order 10) despite the dimensional reduction that must be present, we are less susceptible to this form of coordinate dependence of D. Our diffusion constants are much closer to real physical diffusion constants than the artificial stochastic processes projected onto spaces such as that defined by the number of native contacts, or radius of gyration. There is every reason to suppose that our Ds are in fact constant over the subspaces in which they operate, as these represent sets of configurations with the same degree of compactness. Furthermore, the slower dynamics in collapsed states is physical, emerging from multiple energetic interactions between side groups of the protein in close proximity. We do not account for this effect explicitly, but note that it contributes to folding rates in the same way as the parameter f, which exponentially reduces D (an intermediate search can be slowed down either by its intrinsic dynamics or by a barrier surrounding its target state).

In conclusion, this simple and analytically solvable model, based on few assumptions, can incorporate various scenarios of protein folding and provides quantitative predictions for folding kinetics once coarse-grained structural parameters are known. Furthermore, the physical understanding of model parameters could be employed to predict mutations that affect folding rates.

SUPPORTING MATERIAL

Two figures and Appendices A–D are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00264-1.

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) funded White Rose Doctoral Training Centre. S.R. acknowledges support from the Government of Thailand.

REFERENCES

1. Levinthal, C. 1968. Are there pathways for protein folding? *J. Chim. Phys.* 65:44–45.
2. Karplus, M. 1997. The Levinthal paradox: yesterday and today. *Fold. Des.* 2:S69–S75.
3. Karplus, M., and D. L. Weaver. 1976. Protein-folding dynamics. *Nature.* 260:404–406.
4. Karplus, M., and D. L. Weaver. 1979. Diffusion-collision model for protein folding. *Biopolymers.* 18:1421–1437.
5. Fersht, A. R. 2000. Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. *Proc. Natl. Acad. Sci. USA.* 97: 1525–1529.
6. Leopold, P. E., M. Montal, and J. N. Onuchic. 1992. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA.* 89:8721–8725.
7. Bryngelson, J. D., J. N. Onuchic, …, P. G. Wolynes. 1995. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins.* 21:167–195.
8. Onuchic, J. N., and P. G. Wolynes. 2004. Theory of protein folding. *Curr. Opin. Struct. Biol.* 14:70–75.
9. McLeish, T. C. B. 2005. Protein folding in high-dimensional spaces: hypergutters and the role of nonnative interactions. *Biophys. J.* 88: 172–183.
10. Chan, H. S., Z. Q. Zhang, …, Z. R. Liu. 2011. Cooperativity, local-nonlocal coupling, and nonnative interactions: principles of protein folding from coarse-grained models. *Annu. Rev. Phys. Chem.* 62: 301–326.
11. Clementi, C., and S. S. Plotkin. 2004. The effects of nonnative interactions on protein folding rates: theory and simulation. *Protein Sci.* 13:1750–1766.
12. Eaton, W. A., V. Munoz, …, J. Hofrichter. 1998. Kinetics and dynamics of loops, alpha-helices, beta-hairpins, and fast-folding proteins. *Acc. Chem. Res.* 31:745–753.
13. McLeish, T. C. B. 2006. Diffusive searches in high-dimensional spaces and apparent 'two-state' behaviour in protein folding. *J. Phys. Condens. Matter.* 18:1861–1868.
14. Soranno, A., B. Buchli, …, B. Schuler. 2012. Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proc. Natl. Acad. Sci. USA.* 109:17800–17806.
15. Gouda, H., H. Torigoe, …, I. Shimada. 1992. Three-dimensional solution structure of the B domain of staphylococcal protein A: comparisons of the solution and crystal structures. *Biochemistry.* 31: 9665–9672.
16. Wensley, B. G., S. Batey, …, J. Clarke. 2010. Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. *Nature.* 463:685–688.
17. Khorasanizadeh, S., I. D. Peters, and H. Roder. 1996. Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat. Struct. Biol.* 3:193–205.
18. Bicout, D. J., and A. Szabo. 2000. Entropic barriers, transition states, funnels, and exponential protein folding kinetics: a simple model. *Protein Sci.* 9:452–465.
19. Cleland, W. W. 1975. Partition analysis and the concept of net rate constants as tools in enzyme kinetics. *Biochemistry.* 14:3220–3224.
20. Waley, S. G. 1992. An easy method for deriving steady-state rate equations. *Biochem. J.* 286:357–359.
21. Tanford, C. 1968. Protein denaturation. *Adv. Protein Chem.* 23: 121–282.
22. Myers, J. K., and T. G. Oas. 2001. Preorganized secondary structure as an important determinant of fast protein folding. *Nat. Struct. Biol.* 8:552–558.
23. Brockwell, D. J., and S. E. Radford. 2007. Intermediates: ubiquitous species on folding energy landscapes? *Curr. Opin. Struct. Biol.* 17:30–37.
24. Sánchez, I. E., and T. Kiefhaber. 2003. Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* 325:367–376.
25. Kaya, H., and H. S. Chan. 2003. Origins of chevron rollovers in non-two-state protein folding kinetics. *Phys. Rev. Lett.* 90:258104.
26. Chan, H. S., and K. A. Dill. 1998. Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. *Proteins.* 30: 2–33.

27. Liu, Z. X., G. Reddy, …, D. Thirumalai. 2011. Collapse kinetics and chevron plots from simulations of denaturant-dependent folding of globular proteins. *Proc. Natl. Acad. Sci. USA.* 108:7787–7792.

28. Wagner, C., and T. Kiefhaber. 1999. Intermediates can accelerate protein folding. *Proc. Natl. Acad. Sci. USA.* 96:6716–6721.

29. Bai, Y., A. Karimi, …, P. E. Wright. 1997. Absence of a stable intermediate on the folding pathway of protein A. *Protein Sci.* 6:1449–1457.

30. Dimitriadis, G., A. Drysdale, …, 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. USA.* 101:3809–3814.

31. Sato, S., T. L. Religa, …, A. R. Fersht. 2004. Testing protein-folding simulations by experiment: B domain of protein A. *Proc. Natl. Acad. Sci. USA.* 101:6952–6956.

32. Itoh, K., and M. Sasai. 2006. Flexibly varying folding mechanism of a nearly symmetrical protein: B domain of protein A. *Proc. Natl. Acad. Sci. USA.* 103:7298–7303.

33. Capaldi, A. P., C. Kleanthous, and S. E. Radford. 2002. Im7 folding mechanism: misfolding on a path to the native state. *Nat. Struct. Biol.* 9:209–216.

34. Morton, V. L., C. T. Friel, …, S. E. Radford. 2007. The effect of increasing the stability of non-native interactions on the folding landscape of the bacterial immunity protein Im9. *J. Mol. Biol.* 371: 554–568.

35. Gianni, S., N. R. Guydosh, …, A. R. Fersht. 2003. Unifying features in protein-folding mechanisms. *Proc. Natl. Acad. Sci. USA.* 100:13286–13291.

36. Scott, K. A., and J. Clarke. 2005. Spectrin R16: broad energy barrier or sequential transition states? *Protein Sci.* 14:1617–1629.

37. Wensley, B. G., L. G. Kwa, …, J. Clarke. 2012. Protein folding: adding a nucleus to guide helix docking reduces landscape roughness. *J. Mol. Biol.* 423:273–283.

38. Liu, F., C. Maynard, …, M. Gruebele. 2010. A natural missing link between activated and downhill protein folding scenarios. *Phys. Chem. Chem. Phys.* 12:3542–3549.

39. Neuweiler, H., T. D. Sharpe, …, A. R. Fersht. 2009. The folding mechanism of BBL: plasticity of transition-state structure observed within an ultrafast folding protein family. *J. Mol. Biol.* 390:1060–1073.

40. Otzen, D. E., O. Kristensen, …, M. Oliveberg. 1999. Structural changes in the transition state of protein folding: alternative interpretations of curved chevron plots. *Biochemistry.* 38:6499–6511.

41. Wright, C. F., K. Lindorff-Larsen, …, J. Clarke. 2003. Parallel protein-unfolding pathways revealed and mapped. *Nat. Struct. Biol.* 10: 658–662.

42. Bollen, Y. J. M., I. E. Sánchez, and C. P. M. van Mierlo. 2004. Formation of on- and off-pathway intermediates in the folding kinetics of *Azotobacter vinelandii* apoflavodoxin. *Biochemistry.* 43:10475–10489.

43. Krivov, S. V., and M. Karplus. 2008. Diffusive reaction dynamics on invariant free energy profiles. *Proc. Natl. Acad. Sci. USA.* 105: 13841–13846.

44. Best, R. B., and G. Hummer. 2010. Coordinate-dependent diffusion in protein folding. *Proc. Natl. Acad. Sci. USA.* 107:1088–1093.