Diffusive searches in high-dimensional spaces and apparent 'two-state' behaviour in protein folding



2006 J. Phys.: Condens. Matter 18 1861

(http://iopscience.iop.org/0953-8984/18/6/003)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **18** (2006) 1861–1868 doi:10.1088/0953-8984/18/6/003

# Diffusive searches in high-dimensional spaces and apparent 'two-state' behaviour in protein folding

**T C B McLeish**

Department of Physics and Astronomy and Astbury Centre for Molecular Biology, University of Leeds, Leeds LS2 9JT, UK

Received 1 August 2005, in final form 30 November 2005
Published 23 January 2006
Online at stacks.iop.org/JPhysCM/18/1861

**Abstract**

We extend a simple model for protein folding as a high-dimensional diffusive search. By solving a steady-state diffusion equation on a hypersphere centred on an absorbing 'native state' we find the general property that the kinetics of such a search will always be nearly single exponential. This explains the common observation of such simple 'two-state' folding kinetics in models that contain considerable intermediate structure. It also suggests that the experimental signature of single-exponential folding kinetics does not imply a simple two-state structure to the folding space.

## 1. Introduction

The current conceptual map of protein folding kinetics is dominated by the coexistence of several apparently distinct approaches. They may be categorized loosely into 'energy landscape' (Bryngelson *et al* 1995, Onuchic *et al* 1995), 'diffusion–collision' (Karplus and Weaver 1976, 1994), and 'topomer search' (Makarov and Plaxco 2003) models. Each of these has its own way of visualizing how the collapse of a random coil to a native globule can ever be accomplished in observable timescales, a problem pointed out long ago (Karplus 1997). Each has advantages and drawbacks, but all appeal to low-dimensional projections of the folding dynamics. For example, the 'folding-funnel' picture of the energy landscape has the advantage of visualizing both guided folding and the emergence of on-pathway and off-pathway intermediate states (Dinner *et al* 2000). Yet it is hard to escape from the deceptive simplicity of low-dimensional projections of folding funnels that appear necessarily in all graphical portrayals of it. In practice of course, the dimensionality of the folding space is enormous. Even small (∼100-residue) proteins have a configurational space dimensionality of several hundred (think of the bond angles along the polypeptide main chain alone). In such high-dimensional spaces, qualitatively new features may arise, such as energetically flat domains that nonetheless are extremely difficult to escape from and so behave as kinetic traps. A second feature is the potential for high cooperativity of structure in several simultaneous dimensions. This corresponds to the existence of narrow gullies in the hypersurface that are hard to find.

  

The 'diffusion–collision' approach, on the other hand, is supported by strong experimental evidence that folding rates are controlled by the rate of diffusion of pieces of the open coil in their search for favourable contacts, rather than a driven collapse along some continuous energy surface (Jacob *et al* 1999, Plaxco and Baker 1998, Goldberg and Baldwin 1995). Pre-formed units of secondary structure diffuse hydrodynamically and merge. Larger proteins may do this in an increasingly hierarchical way. The importance of diffusive searches is unsurprising, since under biological conditions, all candidates for energetic interactions, including electrostatics, are *locally* screened to a few ångströms: much smaller than the dimensions over which sections of protein must move to find their native configurations. Put another way, the vast majority of the space covered by the energy landscape must actually be *flat* (on a scale of $k_{\mathrm{B}}T$) rather than funnelled. The conclusion again is that thinking of protein folding as a Brownian search, rather than a low-dimensional activated, or funnelled process, might be more fruitful.

The dimensional reduction of all these approaches is both natural and necessary, since data from kinetic experiments project out many degrees of freedom in an analogous way, but there is a danger in overlooking aspects of folding that rely essentially on the presence of many degrees of freedom. Some other recent studies have pointed out how descriptions of the folding space that retain its necessarily high-dimensional structure may lead to new insights. Crippen and Ohkubo (1998) projected the conformation space of a series of model polypeptides onto a 12-dimensional manifold of thenceforth exactly enumerable states, by maintaining units of pre-formed secondary structure. They found multiple folding pathways that showed no resemblance to a funnelled landscape at all, and folding rates that did not correlate with measures of roughness. Ozkan *et al* (2003) also explored a high-dimensional 'exact enumeration' of states and their transitions in model proteins, in this case employing a 2D lattice Go model. These authors find very common 'two-state' behaviour (that is, near single-exponential kinetics of folding from a population of unfolded states), in spite of the richness of transition routes and states. McLeish (2005) showed, by analytic treatments of diffusion in a hypersphere, that a general property of diffusive searches in high-dimensional spaces is that of exponentially long search times, unless the space supports a series of lower-dimensional subspaces that energetically restrict the search for the native state, proposing one candidate structure for the subspaces in the case of three-helix bundles.

In the following we develop this simple approach of retaining high dimensionality, and a diffusive search, as a paradigm for protein folding. Keeping our energetic structure simple renders our method entirely complementary to low-dimensional simplifications that retain very complex energetic structure. We will show that a rather general property of high-dimensional searches for 'native' targets is that the kinetics are nearly single exponential, even if the overall structure of the space is very much richer than the two-state one-dimensional landscapes usually invoked as models for this behaviour. In the next section we review the structure of the high-dimensional search, and results for the mean folding time. Then we consider the form of the eigenvalue spectrum for the search kinetics, and find a large gap between the first and all higher eigenvalues. Furthermore, we find that the weight of the first eigenvalue dominates in models of folding from the denatured state, so that the single-exponential behaviour corresponding to the lowest eigenvalue dominates observed measures of the folding of a population.

## 2. High-dimensional diffusive subspaces and searches

We start with a very simple and abstract model for protein folding, but one that explicitly retains a very large number of degrees of freedom. The total search space is modelled as the interior of a hypersphere of dimension $d$ and radius $R$, and the native (target) state as a small sphere of radius $R_N$ at the origin of the space. The entire configuration of the protein

corresponds to a single-point particle executing a random walk in the hypersphere. The ratio of $R$ to $R_N$ describes the typical localization on folding in the values of a degree of motional freedom. If the degree of freedom is spatial the appropriate scales are the radius of gyration of a denatured protein and the fluctuation volume of a residue in the native state. If it is angular, then they are the angle of libration of a bond fluctuating in one local minimum as a fraction of $2\pi$. In either case the appropriate order of magnitude estimate is $(\frac{R}{R_N}) \simeq 10$–$100$. Bicout and Szabo (2000) introduced this very general framework for discussing flat and funnelled landscapes, but then restricted themselves to three-dimensional spaces, a simplification that we shall avoid, following McLeish (2005). To explore the timescales of the search for the target space (on which the diffuser will 'stick') we write down the time-dependent diffusion equation for a particle, restricting ourselves to the case of a flat potential at first. The most convenient function to use is the probability density $P(r, t)$, that the system is a radial distance $r$ from the centre of the hypersphere at time $t$, which obeys

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

supplemented by the absorbing boundary condition $P(R_N, t) = 0$, signifying the stability of the native state. The timescale for the search steps is set by the effective diffusion constant $D$. The mean passage time from the unfolded ensemble to the native state can be calculated by introducing a uniform current $J$ of diffusers (representing a population of folding proteins) on the boundary of the hypersphere at $r = R$, as the other boundary condition, and finding the consequent steady state solution to (1). The mean time to pass from $R$ to $R_N$ over the ensemble of systems is then just the total number of diffusers at steady state normalized by the current, leading to

$$\begin{aligned}
\tau_1 &= \frac{4R^2}{\pi D} \qquad d = 1 \\
\tau_2 &= \frac{R^2}{2D}\left[\ln\left(\frac{R}{R_N}\right) - \frac{1}{2}\right] + \frac{R_N^2}{D} \qquad d = 2 \\
\tau_3 &= \frac{R^2}{D}\frac{1}{d(d-2)}\left(\frac{R}{R_N}\right)^{(d-2)} \qquad d \geqslant 3.
\end{aligned} \tag{2}$$

This expression indicates how very much *qualitatively* longer the mean search time is in *high* dimensions ($d > 2$), than the low-dimensional estimation of the characteristic time $\tau \simeq R^2/D$ that applies in $d = 1$ and 2. This fundamental time is scaled up by the denatured system size (measured in units of the target size $R_N$) to the power of the number of effective dimensions greater than two.

The central result of (2) depends on two key physical assumptions: (1) the dimensionality of the space corresponds to the number of degrees of freedom for protein folding—of the order of a hundred or more (this is realistic), and (2) the stability of the folded state is highly cooperative and requires all residues to attain native-like configuration before their attractive interactions become effective (this is not realistic). With these assumptions alone, the model of high-dimensional diffusion we have described is inevitable, and the timescales unreasonably long. The exponentially large search times arise transparently from the factor $\left(\frac{R}{R_N}\right)^{(d-2)}$ in equation (2) for $d > 2$. This is of course a restatement of Levinthal's paradox (Karplus 1997), but a helpful one, in that the two necessary assumptions for the paradox to arise are clearly seen. The first just gives the large dimensionality of hypersphere, the second the flat diffusive landscape across the entire folding space.

Put this way, there are two ways of circumventing the problem. One may drop the assumption of local forces and allow the protein to 'fall' towards the single native state down

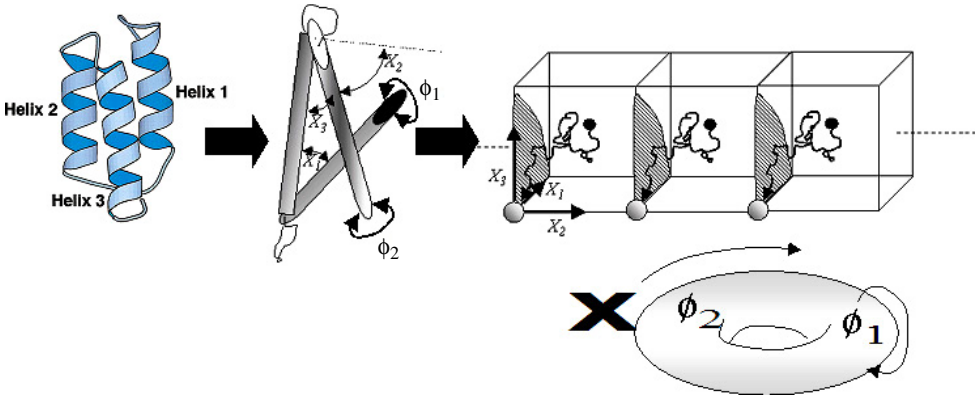


**Figure 1.** The three-helix bundle (BdpA on the left) is coarse grained to a system of three rods. The angles constituting the diffusive subspaces are labelled $X_i$ ($i = 1, 2, 3$) for the relative orientation of the rods and $\phi_i$ ($i = 1, 2$) for their relative rotation. The five-dimensional folding space then looks like the periodic cubic lattice on the right (only the $X_2$ direction is shown periodic, for clarity), at every point of which lies a torus described by $\phi_1$ and $\phi_2$. The attractive gutter is the 2D space spanned by $X_1$ and $X_3$ once ($\phi_1$, $\phi_2$, $X_2$)-diffusion has brought the third helix into contact with the first along their attractive sides. But for small angles $X_2$ there is a large topological barrier between the 'correct' and 'incorrect' sides of attachment of the third helix onto the bundle formed by the other two, and identical with the rapid diffusional subspace of $X_1$ and $X_3$. By stabilizing the ($X_1$, $X_3$) plane in this way, a five-dimensional search is split into three- and two-dimensional searches, whose total mean search time is much faster.

(This figure is in colour only in the electronic version)

a 'funnel' created by forces whose ranges permeate the entire volume. As we have remarked above, however, candidates for such long-range forces do not present themselves. Without recourse to a continuous funnel-shaped landscape, there is only one other possibility: *all diffusive searches take place in low-dimensional subspaces* of the full configurational space. The structure of this approach was investigated in McLeish (2005). The subspaces may be stabilized by native or by tailored non-native interactions that constitute the attractive hypersurfaces (or 'hypergutters') of the diffusive subspaces themselves. Since the diffusion is always maintained in some low-dimensional subspace of the full folding space, $\tau \sim R^2/D$ for each subspace providing that it is of $d < 3$, so that the folding time may be reduced to a minimum of $\tau_{\min} \simeq d^2R^2/D$ rather than the exponentially larger $\left(R^2/R_N^2\right)^d$. This clearly reduces the folding time enormously, signifying that only a tiny fraction of possible states is visited in the search (Dinner *et al* 2000). More likely scenarios for real proteins lie between the extreme limits of impossibly long searches in the full configurational space and this maximally fast concatenation. Some of the intermediate search spaces are likely to be of moderate dimension ($d \lesssim 10$), so creating folding times that lie in the (very wide) experimental range of $10^{-6} - 10^{1}$ s. McLeish (2005) treated the example of a three-helix bundle that contains a diffusive hyper-toroidal subspace with $d = 5$ (see figure 1). Since both experiment and simulation typically measure the time dependence of folding as well as the mean folding time, it is of interest to explore the form of the kinetics by which a distribution of diffusers spread initially throughout a high-dimensional search space collapses onto an absorbing native state.

## 3. Eigenvalue spectrum in high-dimensional diffusion

We continue to treat the case of the hypersphere introduced by Bicout and Szabo (2000) for which the mean search times have already been calculated in (2). We will take a standard

approach and calculate the eigenfunctions and eigenvalues of the diffusion operator on this space with appropriate boundary conditions. Then by choosing a superposition that represents a good model of an ensemble of denatured, unfolded states at $t = 0$ we will find the kinetics of adsorption of the population onto the absorbing folded state. By separation of variables in the $d$-dimensional diffusion equation we write (setting $D = 1$ for clarity of presentation)

$$\frac{1}{r^{d-1}}\frac{\partial}{\partial r}r^{d-1}\frac{\partial\phi_n(r)}{\partial r} = k_n^2\phi_n(r) \tag{3}$$

for the eigenfunctions $\phi_n(r)$, and their associated eigenvalues $k_n$, taking their time-dependences to be $\mathrm{e}^{-k_n^2 t}$. The appropriate boundary conditions for the problem are (1)

$$\phi_n(R_N) = 0 \tag{4}$$

indicating absorption at the (inner) native state boundary, and (2)

$$\left.\frac{\partial\phi_n(r)}{\partial r}\right|_{r=R} = 0 \tag{5a}$$

indicating zero outflow from the configuration space at the outer boundary of the hypersphere, which models the ensemble of completely unfolded states. We may make the auxiliary substitutions

$$\phi_n(r) = r^{(2-d)/2}f(r) \tag{6}$$

$$x = k_n r \tag{7}$$

$$y(x) = f(r) \tag{8}$$

to find for the auxiliary function $y(x)$

$$x^2y''(r) + xy'(r) + \left[x^2 - \left(\frac{d-2}{2}\right)^2\right]y(x) = 0. \tag{9}$$

This is Bessel's equation of order $\nu = (d-2)/2$. So the general non-normalized solution of (3) is

$$\phi_n(r) = r^{(d-2)/2}\left[J_{(\frac{d-2}{2})}(k_n r) + C_Y Y_{(\frac{d-2}{2})}(k_n r)\right] \tag{10}$$

where the constant $C_Y$ gives the degree to which the Bessel function of the second kind is admixed to that of the first kind for each eigenvalue. Our interest in large dimensions assists us analytically since simple asymptotic forms for the Bessel functions exist for large values of their order. Specifically, for large values of $\nu$, we may use the following expressions:

$$\begin{aligned} J_\nu(x) &\simeq \frac{1}{\sqrt{2\pi\nu}}\left(\frac{ex}{2\nu}\right)^\nu \\ Y_\nu(x) &\simeq -\sqrt{\frac{2}{\pi\nu}}\left(\frac{ex}{2\nu}\right)^{-\nu} \end{aligned} \tag{11}$$

to find, from (4), that the constant measuring the admixture of the Bessel function of the second kind in the solution, $C_Y$, satisfies

$$C_Y \simeq (k_n R_N)^{2\nu}\frac{\sqrt{4\pi\nu}}{\Gamma(2\nu+1)}. \tag{12}$$

Keeping the assumptions of large $d$, we now look for a first eigenvalue $k_1$. The key insight here is that the corresponding first eigenfunction possesses no nodes in the range $R_N < r < R$: it passes from the zero at $R_N$ to the zero gradient at $R$ by the opposite asymptotic properties of $J_\nu(x)$ and $Y_\nu(x)$, which suppresses exponentially the effect on the curvature of the functions

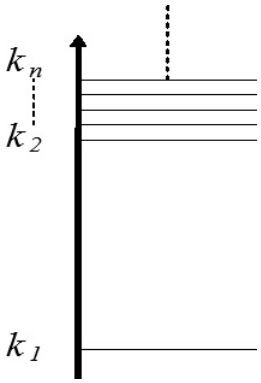


**Figure 2.** The form of the eigenvalue spectrum for a high-dimensional diffusive search. The first eigenvalue $k_1$ is of order $\left(\frac{R_N}{R}\right)^{(d-2)} \frac{D}{R^2}$. All higher eigenvalues are of the same order as the low-dimensional search rate $\frac{D}{R^2}$.

themselves from the size of the space. We find from substitution into the second boundary condition (5*a*) at the outer boundary, and expansion in powers of $k_1$, that

$$k_1^2 \simeq d\,(d-2)\left(\frac{R_N}{R}\right)^{(d-2)} \frac{D}{R^2} \tag{13}$$

as the *only* self-consistently small value (replacing the diffusion constant $D$ to achieve correct dimensions of the result). We note immediately that this lowest eigenvalue has the same structure as the inverse mean search time for high dimensions found in (2). This suggests that the weight accorded to higher eigenfunctions is small. All higher eigenvalues have much higher values: we observe that, by orthogonality, the higher eigenfunctions must use the turning points of the Bessel function $J_\nu(x)$ itself to satisfy the boundary conditions, and therefore acquire values

$$k_n^2 \simeq \frac{D}{R^2}, \qquad n > 1. \tag{14}$$

The spectrum of the high-dimensional problem is therefore highly differentiated into a ground state of very much lower frequency (decay time $\tau = 1/(Dk_1)$) than all other states. This structure is illustrated in figure 2.

Furthermore, the amplitude of the decay from an initial distribution of states covering the 'shell' of unfolded states is dominated by this ground state. For $\nu \gg 1$, only the ground state has a significant amplitude at the surface of the hypersphere—just the region where most of the hypervolume resides. Formally, the amplitudes $a_n$ of all higher eigenfunctions are less than that of the first by $(R/R_N)^{-d/2}$, by the projection formula

$$a_n = \int P(\mathbf{r}, 0)\phi_n(\mathbf{r})\,\mathrm{d}^d\mathbf{r}. \tag{15}$$

Finally, the folding kinetics, described by the time dependence of the integral over the surviving diffusing population is nearly single exponential:

$$\begin{aligned} I(t) &= \int P(\mathbf{r}, t)\,\mathrm{d}^d\mathbf{r} \\ &= \sum_n a_n \mathrm{e}^{-k_n^2 t} \\ &\simeq \mathrm{e}^{-k_1^2 t} \end{aligned}$$

with corrections to this dominant term of the order of $(R/R_N)^{-d/2}$.

## 4. Discussion

We have found that single-exponential kinetics emerge from diffusive searches generally, not only in the case of one-dimensional diffusion over barriers (the usual 'two-state' model) but also in the case of very high-dimensional spaces (with a small 'target' or 'native' absorbing space) that support rather complex structures. Although we have pursued a calculation in the highly symmetric case of a hypersphere, the result is expected to be more general. The diffusion operator in the presence of a wide class of boundary shapes is still self-adjoint, so supports the same eigenfunction decomposition that we have used above. In particular, the key feature that the first eigenfunction requires no nodes in the interior of the hyperspace, while all higher eigenfunctions do, is expected to survive to more general high-dimensional geometries. Since this is the essential reason for the specially small value of the first eigenvalue, the single-exponential kinetics are expected to be robust to perturbations of the boundary.

The spectrum of eigenvalues of figure 2 resembles, of course, another spectrum that arises in the protein folding problem: that of the configuration-state free energies themselves. Protein function requires that the native state be separated from nearby, but non-functional, states by an energy gap large compared with the thermal energy $k_{\mathrm{B}}T$. It seems that a similar structure arises in the 'energy-like' spectrum of rates that correspond to the eigenvalues of the folding operator. The two structures are however quite distinct.

It is worth discussing the relation of this model to other conceptual frameworks for protein folding. Other views emerge from different projections of the high-dimensional space. In the case of the unrealistically symmetric and potential-free hypersphere, for example, projection onto the radial co-ordinate renormalizes the diffusion equation into a one-dimensional version but this time with a barrier highly proportional to the logarithm of the search time in the full space. This 'entropic barrier' is exactly that of the 'two-state' model for this case. The diffusion–collision model, on the other hand, may appear in some choices of dimensional reduction: those that involve spatial searches for the right juxtaposition of pre-folded or globular subdomains of the protein. In these cases the model appears from the high-dimensional search when all non-diffusive degrees of freedom are projected out first.

Of particular interest is the relationship of the hypergutter picture to the topomer search model. This is because the rate-determining diffusive searches will in general be completed only when a topological, as well as a spatial, constraint in the final native state is satisfied for the first time. It is also to be expected since this model also seeks to understand folding rates without recourse to their dependence on native interactions. Again, the three-helix bundle serves as a model example—restriction from the 3D helical angular space to the faster 2D space with helices 1 and 3 in contact occurs when the topological orientation of the helices is satisfied. Similar conclusions would emerge if the slow searches were between a helix and a $\beta$-sheet, or between two or more $\beta$-turns. Retaining all the diffusional degrees of freedom leads to a close relationship between the contact order $Q_D$ of the topomer search model and the number of diffusional dimensions $d$ of the space of hypergutters in the rate-determining diffusive search, for we can identify the exponent $Q_D$ in the rate expression in the topomer search model (Makarov and Plaxco 2003)

$$k_f \sim \gamma \langle K \rangle^{Q_D} \tag{16}$$

with that in the diffusive-search result (2) above to find, formally at the level of the exponent, that $Q_D = (d-2)$, where here $d$ is the dimension of the largest diffusive search. We note, however, that departures from the correlation of folding time with contact order might be expected when non-native interactions are tuned to speed up folding in the way we have outlined. This is because such a strategy can reduce the effective dimension of the search

without changing the topology of the final state. Strong outlying behaviour in the correlation of folding time with contact order may be connected with the potential variability in the efficiency of hypergutter stability.

Both experiments and numerical calculations have frequently found near-single-exponential kinetics when it is known that more complex dynamic structures support the folding process (e.g. Ozkan *et al* (2003)). In the light of our result this is not a mystery, but it does point to some caution when experiments themselves find it, for the diffusive folding search may pass though many transiently stabilized subspaces of low effective dimension, yet both the overall folding rate and the kinetic form will be dominated by the search space of highest dimension in the pathway. Such a space needs only to retain a moderate number of simultaneously diffusing degrees of freedom for the kinetics to be the single exponential by virtue of the dimensionality of the space alone. From this point of view, global measures of kinetics are poor measures of the structure of folding spaces.

## Acknowledgments

I would like to thank Ken Dill and Joan Shea for helpful discussions and comments on the manuscripts, and the Kavli Institute for Theoretical Physics for the hospitality of its 2002 programme 'Dynamics of complex and macromolecular fluids', where part of this work was done (preprint number NSF-ITP-02-55).

## References

Bicout D J and Szabo A 2000 Entropic barriers, transition states, funnels, and exponential protein folding kinetics: a simple model *Protein Sci.* **9** 452–65

Bryngelson J D, Onuchic J N, Socci N D and Wolynes P G 1995 Funnels, pathways and the energy landscape of protein folding *Proc. Struct. Funct. Genet.* **21** 167–95

Crippen G M and Ohkubo Y Z 1998 Statistical mechanics of protein folding by exhaustive enumeration *Proteins Struct. Funct. Genet.* **32** 425–37

Dinner A R, Šali A, Smith L J, Dobson C M and Karplus M 2000 Understanding protein folding via free-energy surfaces from theory and experiment *TIBS* **25** 331–9

Goldberg J M and Baldwin R L 1995 Diffusional barrier crossing in a two-state protein folding reaction *Proc. Natl Acad. Sci. USA* **96** 2019–24

Jacob M, Geeves M, Holtermann G and Schmidt F X 1999 Diffusional barrier crossing in a two-state protein folding reaction *Nat. Struct. Biol.* **6** 923–6

Karplus M 1997 The Levinthal paradox: yesterday and today *Fold Des.* **2** S69–76

Karplus M and Weaver D L 1976 Protein folding dynamics *Nature* **260** 404–6

Karplus M and Weaver D L 1994 The diffusion–collision model and experimental data *Protein Sci.* **3** 650–68

Makarov D E and Plaxco K W 2003 The topomer search model: a quantitative, first-principles description of two-state protein folding kinetics *Protein Sci.* **12** 17–26

McLeish T C B 2005 Protein folding in high-dimensional spaces: hypergutters and the role of nonnative interactions *Biophys. J.* **88** 172–83

Onuchic J N, Wolynes P G, LuthySchulten Z and Socci N D 1995 Toward and outline of the topography of a realistic protein folding funnel *Proc. Natl Acad. Sci. USA* **92** 3626–30

Ozkan S B, Bahar I and Dill K A 2003 Computing the transition state populations in simple protein models *Biopolymers* **68** 35–46

Plaxco K W and Baker D 1998 Limited internal friction in the rate-limiting step of a two-state proteinfolding reaction *Proc. Natl Acad. Sci. USA* **95** 13592–6