OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

Is Protein Folding Sub-Diffusive?

Sergei V. Krivov*

Institute of Molecular and Cellular Biology, University of Leeds, Leeds, United Kingdom

Abstract

Protein folding dynamics is often described as diffusion on a free energy surface considered as a function of one or few reaction coordinates. However, a growing number of experiments and models show that, when projected onto a reaction coordinate, protein dynamics is sub-diffusive. This raises the question as to whether the conventionally used diffusive description of the dynamics is adequate. Here, we numerically construct the optimum reaction coordinate for a long equilibrium folding trajectory of a Go model of a λ-repressor protein. The trajectory projected onto this coordinate exhibits diffusive dynamics, while the dynamics of the same trajectory projected onto a sub-optimal reaction coordinate is sub-diffusive. We show that the higher the (cut-based) free energy profile for the putative reaction coordinate, the more diffusive the dynamics become when projected on this coordinate. The results suggest that whether the projected dynamics is diffusive or sub-diffusive depends on the chosen reaction coordinate. Protein folding can be described as diffusion on the free energy surface as function of the optimum reaction coordinate. And conversely, the conventional reaction coordinates, even though they might be based on physical intuition, are often sub-optimal and, hence, show sub-diffusive dynamics.

Citation: Krivov SV (2010) Is Protein Folding Sub-Diffusive? PLoS Comput Biol 6(9): e1000921. doi:10.1371/journal.pcbi.1000921

Editor: Jin Wang, State University of New York at Stony Brook, United States of America

Received January 31, 2010; **Accepted** August 6, 2010; **Published** September 16, 2010

Funding: This work was supported by a RCUK fellowship. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The author has declared that no competing interests exist.

* E-mail: s.krivov@leeds.ac.uk

Introduction

A free energy surface (FES) projected onto one or a small number of coordinates is often used to describe the equilibrium and kinetic properties of complex systems with a very large number (100 to 1,000 or more) of degrees of freedom. Studies of protein folding are an important case where this type of projected surface has been introduced and coordinates such as the number of native contacts and radius of gyration have been used [1–3]. Protein folding then is described as diffusion on the projected free energy surface. Diffusive dynamics is characterized by means square displacement linearly growing with time, $\langle \Delta x^2(t) \rangle = 2Dt$, where D is the diffusion coefficient. For a single reaction coordinate diffusive dynamics is completely specified by the free energy profile (FEP), i.e. the free energy as a function of the coordinate and coordinate-dependent diffusion coefficient, which conveniently can be computed from conventional and cut based free energy profiles [4]. Construction of a "good" reaction coordinate (i.e. the one that preserves systems dynamics) is challenging. In many cases, the standard progress variables (e.g. number of native contacts, radius of gyration, root mean square distance from the native structure) are not good reaction coordinates, because they do not preserve the barriers on the FES and thus may mask the inherent complexity of the latter [5]. A number of methods to construct good reaction coordinates have been suggested [4,6–9].

Employing the Mori-Zwanzig formalism [10,11] one can derive generalized Langevin equations, which describe system dynamics projected on the reaction coordinates. The generalized Langevin equation contains a memory kernel, which leads to non-Markovian dynamics and subdiffusion. Subdiffusion is characterized by the mean square displacement growing slower than that for diffusion, $\langle \Delta x^2(t) \rangle \sim t^{2\alpha}$ with exponent $2\alpha < 1$. To completely specify dynamics in this case one has to compute the memory kernel, which is not trivial, since it requires the solution of a multidimensional partial differential equation [12]. Long-term memory in correlation functions and anomalous diffusion in proteins was observed experimentally and theoretically [13–23]. This raises the question whether the folding dynamics of proteins can be described as simple diffusion on the projected free energy surface, as is often done, or if one has to use more sophisticated descriptions, e.g. generalized Langevin equations [24,25], fractional Fokker-Plank equations [26] or multiscale state space networks [19]. Here we show that if the reaction coordinate is properly optimized, then the dynamics projected onto this coordinate is diffusive, while the same dynamics projected onto a sub-optimal coordinate is sub-diffusive.

Results/Discussion

The equilibrium folding dynamics of the C_α Go model [27] of the N-terminal domain of phage λ-repressor protein is analyzed [28]. Structure-based Go models containing attractive native interactions and repulsive nonnative interactions correspond to perfectly funneled energy landscapes with energetic frustration completely absent [3,29]. A trajectory of 4×10^6 frames (saved with $\Delta t =$ 7.5 ps) was obtained by simulating with Langevin molecular dynamics at T = 323 K and contains about 100 folding-unfolding events. The saving interval of 7.5 ps is used below as the unit of time. Note that the timescales in the simulation do not correspond directly to the timescales of the folding dynamics of the real protein because the coarse-grained model of the protein

Author Summary

To understand dynamics of complex systems with many degrees of freedom, one often projects it onto one or several collective variables. Protein folding, the complex, concerted motion of a protein chain towards a unique three-dimensional structure, is one example of where such reduction of complexity is useful. It is usually assumed that the projected dynamics is diffusive. However, many experiments and simulations have shown that the projected dynamics is sub-diffusive, i.e., the mean square displacement grows slower than linear with time. It means that the dynamics has a memory; that the free energy surface together with diffusion coefficient do not properly define the dynamics; and that such projections cannot be used to accurately describe dynamics. Here, we show that if one carefully constructs the reaction coordinate by optimizing (maximizing) its free energy profile, one can use a simple (memory-less) diffusive description. Loosely speaking, when the complex dynamics is projected onto a simple coordinate, all the complexity of the original dynamics goes into the memory of the projected dynamics. If the dynamics is projected onto the (complex) optimum reaction coordinate, all the complexity of the original dynamics is in the reaction coordinate, and the projected dynamics is simple.

without explicit representation of the solvent is employed. Relation between the folding timescales of coarse-grained models of proteins and that of real proteins is discussed in [30]. The protein has complex FES with five basins: denatured, native, native′, intermediate and intermediate′ and two symmetrical folding pathways [28].

Optimum one-dimensional reaction coordinates are constructed by numerically optimizing the mean first passage time to the native basin for a sufficiently broadly chosen functional form of a reaction coordinate (see Methods). Two different functional forms of reaction coordinates are considered. For each coordinate we show the cut based free energy profile (FEP) F_C together with the exponent $\alpha(x)$; the latter is used to distinguish between diffusive and sub-diffusive dynamics (see Methods). The coordinate dependent exponent $\alpha(x)$ describes how the mean absolute displacement grows with time, $<|\Delta x(t)|> \sim t^{\alpha}$ and can be determined from the distance between $F_C(x)$ computed at two different sampling intervals Δt (see Methods); the smaller is the distance, the higher is the exponent $\alpha(x)$. $\alpha(x)$ is equal to $1/2$ for diffusive and is less than $1/2$ for sub-diffusive dynamics. Each coordinate is transformed to the natural coordinate (see Methods), so that the diffusion coefficient is constant and is equal to one and diffusive dynamics is completely specified by the FEP $F_C(x)$.

The first coordinate (X_1) generalizes the number of native contacts coordinate NNC as: $X_1 = \sum_{i<j} a_{ij}\Theta(\Delta_{ij} - r_{ij})$, where a_{ij} is either 1 or -1, r_{ij} is the distance between atoms i and j and Δ_{ij} is the distance threshold, when contact between the atoms is considered to be formed; Θ is the Heaviside step function, whose value is zero for a negative argument and one for a positive argument. Figure 1a shows $F_C(x)$ and $\alpha(x)$ for the (sub-optimal) reaction coordinate, just initialized to the NNC, i.e., $X_1 = \sum_{ij\in \mathrm{enc}} \Theta(\Delta - r_{ij})$ with sum over pairs of atoms (ij) in the set of native contacts. The value of $\Delta = 12$ gives the highest barrier for the transition state for the simple variants of NNC, where the distance threshold is the same for all the native contacts. The relatively large value (inter-atom distances between C_α atoms in the native contacts are within 4 to 12 Å) may be explained by the fact that the optimal reaction coordinate should better distinguish between the denatured and native basins (rather than indicate a formed native contact), which happens around the transition state and sufficiently far from the native structure. On the FEP one can notice three basins: denatured $0 < x < 17$, native $x > 29$; the third basin $17 < x < 29$ consists of a number of overlapping free energy basins. The exponent $\alpha \approx 0.25$ shows that the dynamics is sub-diffusive. To confirm this Figure 2 shows the mean square displacement (MSD) as a function of time ($\langle \Delta x^2(\Delta t) \rangle$) averaged over pieces of the trajectory that start from the transition state (TS) ($x \approx 17$). The MSD grows approximately as $\sim t^{0.5}$ (the mean absolute displacement as $<|\Delta x(t)|> \sim t^{0.25}$), indicating sub-diffusive dynamics. The number of folding events computed with Kramer's equation (Eq. 4) is 1200, i.e. an order of magnitude more than the actual number of 100 events. It means that the reaction coordinate is "bad" and the computed folding free energy barrier is lower than the correct one. Limited structural information can be exploited by making the distance threshold proportional to the native distance $\Delta_{ij} = \mu r^0_{ij}$ for each native contact (ij), so that $X_1 = \sum_{ij\in \mathrm{enc}} \Theta(\mu r^0_{ij} - r_{ij})$. However, it does not improve the reaction coordinate since the highest barrier for the transition state, obtained at $\mu = 1.6$ (see Figure S1 in Text S1), is similar to that obtained with the constant threshold (Figure 1a).

Figure 1b shows $F_C(x)$ and $\alpha(x)$ for the optimized reaction coordinate $X_1 = \sum_{i<j} a_{ij}\Theta(\Delta_{ij} - r_{ij})$. The FEP is more informa-

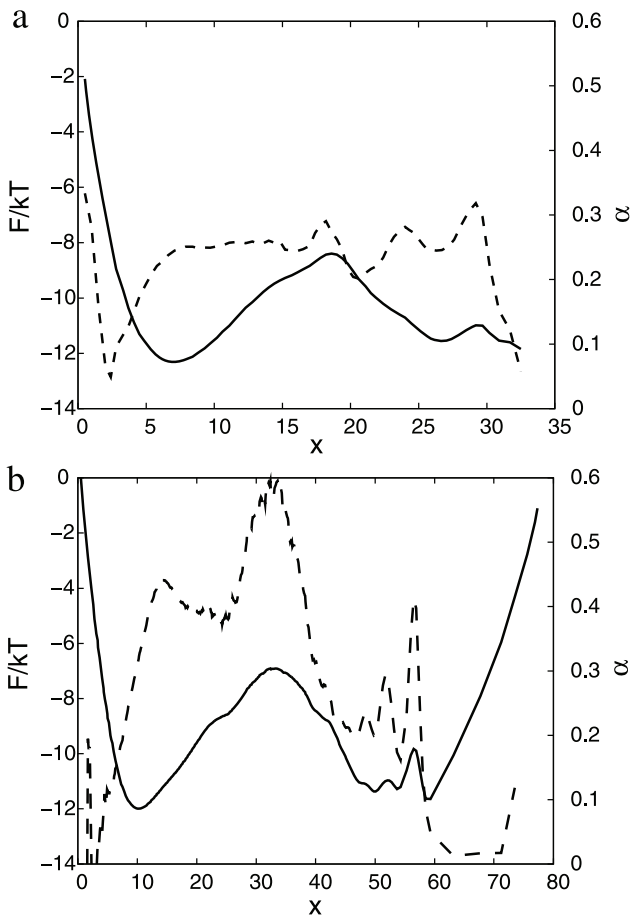

Figure 1. Optimization of X_1 reaction coordinate. F_C (solid line) and $\alpha(x)$ (dashed line) for X_1 as a reaction coordinate; (a) X_1 initialized to NNC, (b) optimized X_1. Reaction coordinates are transformed to the natural reaction coordinate.
doi:10.1371/journal.pcbi.1000921.g001

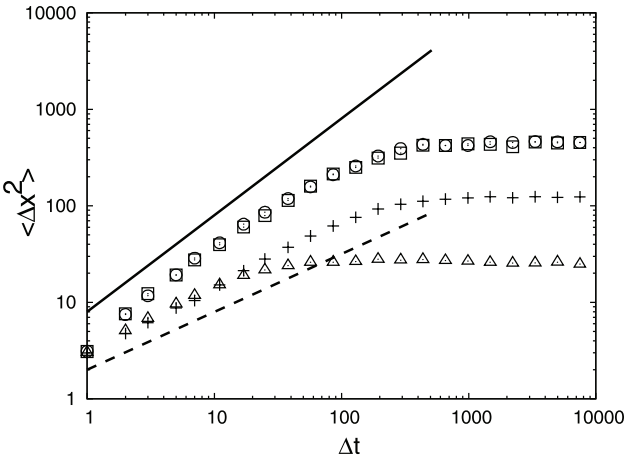

Figure 2. MSD for pieces of the trajectory starting from the corresponding transition states. Pluses are for unoptimized X_1, squares are for optimized X_1, triangles are for unoptimized X_2, circles are for optimized X_2. The solid line shows diffusive ($\Delta x^2 \sim t$) and the dashed line sub-diffusive ($\Delta x^2 \sim t^{0.5}$) MSD to guide the eye.
doi:10.1371/journal.pcbi.1000921.g002

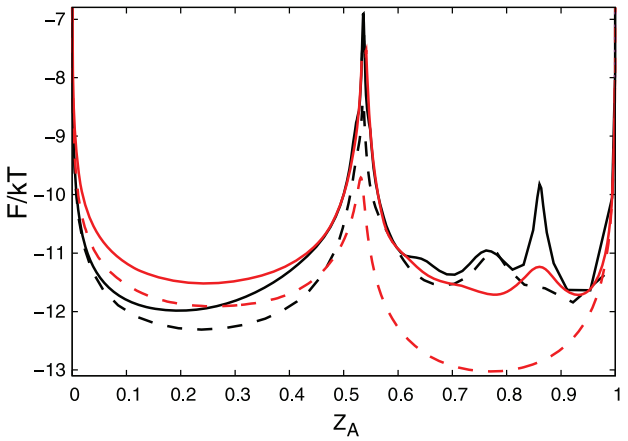

Figure 3. The reaction coordinates comparison. Black and red lines show the free energy profiles along the X_1 and X_2 coordinates, respectively. Solid and dashed lines show optimized and non-optimized coordinates, respectively.
doi:10.1371/journal.pcbi.1000921.g003

tive now: one can distinguish the three basins, that were overlapping on Figure 1a. The free energy of the transition state ($x \approx 33$) of the optimized reaction coordinate is higher than that for the sub-optimal one (Figure 1a). The relative position of the transition state for the optimum coordinate is shifted to the left compared to the NNC coordinate which may give a misleading impression that the transition states occupy different regions of the configuration space. The optimum and NNC reaction coordinates have different coordinate dependent diffusion coefficients. When the coordinates are transformed to the natural coordinate with diffusion coefficient equal to unity the same regions of the configuration space may occupy different positions. Figure 3 shows FEPs along the Z_A reaction coordinate, which is invariant to coordinate transformation, and can be used to compare different coordinates. $Z_A(x) = \int_0^x Z_H(x)dx / \int_0^\infty Z_H(x)dx$ measures the relative partition function of the coordinate segment between 0 and x. The transition states on Figure 1 correspond to those on Figure 3, since the cut free energy profiles are invariant under coordinate transformation [4]. The transition states on Figure 3 are located at the same position, i.e., they occupy the same region of the configuration space. F_C for the optimum coordinates are uniformly higher than that for the corresponding sub-optimum ones. Z_A coordinate, however, is of limited use to correctly represent the dynamics since the diffusion coefficient is not constant, which leads to such artifacts as sharply peaked transition states.

The scaling exponent α for the optimized reaction coordinate (Figure 1b) is no longer a constant. It is a bit higher than 0.5 at the TS region ($x \approx 33$) and a bit lower than 0.5 in the denatured state and at the second barrier ($x \approx 57$), indicating diffusive dynamics. After the TS $\alpha(x)$ is around 0.25 indicating sub-diffusive dynamics. Values of α higher than 0.5 (superdiffsion) are an artifact due to over-fitting of the trajectory by the reaction coordinate. The estimated number of folding events for the optimized reaction coordinate is 168, which is quite close to the actual number. Figure 2 shows MSD for the pieces of the trajectory starting from the TS ($x \approx 33$). The MSD grows linearly with time, confirming diffusive dynamics. The reaction coordinate can be optimized in another region, e.g. by maximizing the mfpt to go from the TS ($x \approx 33$) to the native structure ($x \approx 60$). In that case dynamics in the region around the second barrier ($x \approx 57$) becomes diffusive, while that at the TS is back to sub-diffusive. Optimization of the reaction coordinate inside the native basin has increased the exponent $\alpha(x)$ in the basin from 0 to 0.3, indicating that the dynamics in the basin is still sub-diffusive. This can be due to a relatively large value of the sampling interval (Δt) of 7.5 ps, at which MSD between two subsequent snapshots is close to an equilibrium value inside the native basin. Moreover, sub-diffusive dynamics inside the basins have relatively small influence on folding dynamics, which is determined mainly by diffusive dynamics at the transitions state regions. The Text S1 shows an all-atom structure based model of the lambda repressor protein where the optimum reaction coordinate is constructed so that the dynamics is diffusive for the whole coordinate, not just around the transition state.

The second coordinate is a linear combination of all interatom distances $X_2 = \sum_{i<j} \alpha_{ij} r_{ij}$, where r_{ij} is the distance between atoms i and j. It was initialized to be the distance between atoms C_{13} and C_{46} (Figure 4a). The end to end distance (the distance between C_1 and C_{80} atoms), often employed in single molecule experiments, does not separate the denatured and native basins; the free energy profile along the distance is barrier-less. Figures 4 and 2 show $F_C(x)$, $\alpha(x)$ and MSD for X_2 and lead to the similar conclusions. For the sub-optimal X_2 dynamics is sub-diffusive at the TS ($x \approx 21$), with the exponent $\alpha(x)$ steeply decreasing to zero just after the TS. The exponent $\alpha(x) = 0$ means that the MSD has reached the equilibrium value (at this time scale and in this region of the reaction coordinate). The estimated number of folding events is about 8700. The optimized reaction coordinate (panel b) has a higher folding barrier and shows that the dynamics is diffusive at the TS ($x \approx 40$) and estimates the number of folding events as 154.

Figure 5 shows $F_C(x)$ for different values of the sampling interval $\Delta t = 1,2,4,...,1024$ for the optimal and sub-optimal coordinates X_1. The constant distance between the profiles at fixed x and different (small) Δt means that α is independent of Δt and that $\langle |\Delta x(\Delta t)| \rangle \sim \Delta t^{\alpha}$ (see also Figure 6).

The partition function of the cut based free energy profiles $Z_C(x)$ ($F_C(x) = -kT \ln Z_C(x)$) at point x is defined as the number of transitions through the point [4] (see Methods). For the sufficiently large sampling intervals $\Delta t > t^b$, when the system

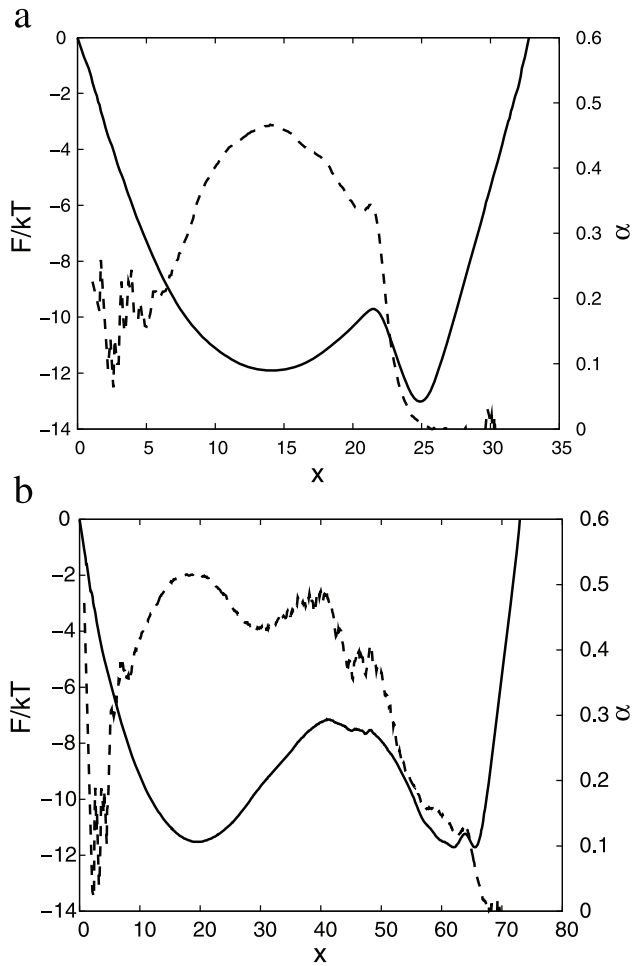

Figure 4. Optimization of X_2 reaction coordinate. F_C (solid line) and $\alpha(x)$ (dashed line) for X_2 as a reaction coordinate; (a) X_2 initialized to $r_{C_{13}\,C_{46}}$, (b) optimized X_2. Reaction coordinates are transformed to the natural reaction coordinate.
doi:10.1371/journal.pcbi.1000921.g004

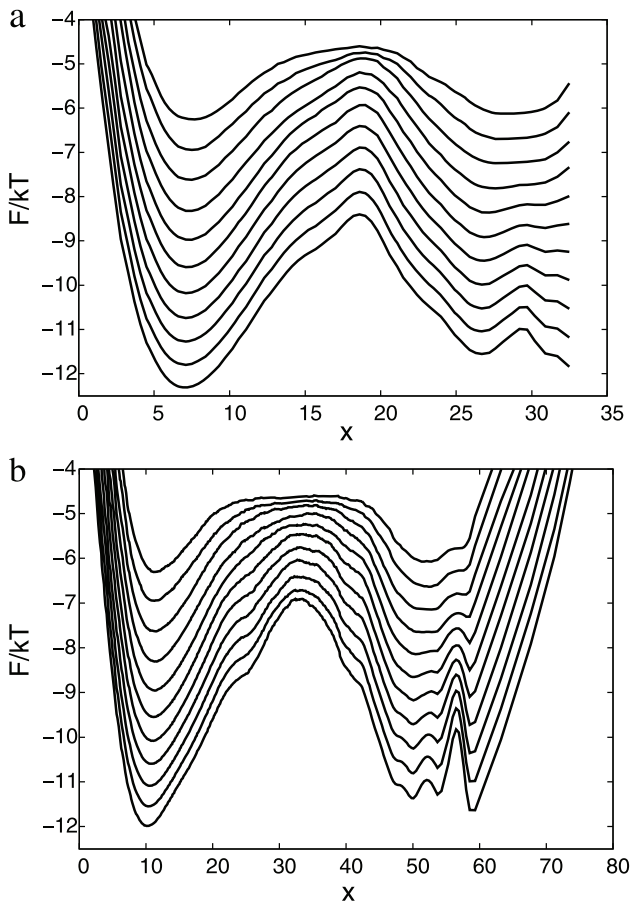

Figure 5. F_C computed at different sampling intervals for X_1 as a reaction coordinate. The sampling intervals are $\Delta t = 1,2,4...,1024$; (a) the NNC reaction coordinate, (b) the optimum reaction coordinate. Higher free energy barrier for the optimum reaction coordinate implies lesser space between the profiles and larger α compared to the NNC.
doi:10.1371/journal.pcbi.1000921.g005

"flies" ballistically over the TS barrier, i.e. no recrossing events are detected, the Z_C at the TS is equal to the total number of folding events (100 here). This value denoted as Z_C^b (F_C^b) is, evidently, the same for the optimal and sub-optimal coordinates. The optimum reaction coordinate has higher $F_C(\Delta t = 1)$ at the TS compared to the sub-optimal coordinate. Hence, α (at the TS) estimated as (see Methods)

$$\alpha = 1 + \frac{F_C(\Delta t = 1) - F_C^b}{kT \ln t^b} \qquad (1)$$

is higher for the optimum reaction coordinate than it is for the sub-optimal one. In other words, an inadequacy of a sub-optimal reaction coordinate (low F_C) which leads to faster kinetics is corrected by making the dynamics sub-diffusive (slower).

We assume here that α is roughly a constant for the sampling intervals between $\Delta t = 1$ and $\Delta t = t_b$ (i.e., $\Delta x \sim \Delta t^{\alpha}$), which is validated by Figure 5. However the assumption evidently breaks down for the very small time scales, when the system follows Newtons equations of motion with $\Delta x \sim v\Delta t$ meaning $\alpha = 1$. Thus, the sampling interval Δt should be chosen sufficiently large so that the dynamics is in the (sub)diffusive regime.

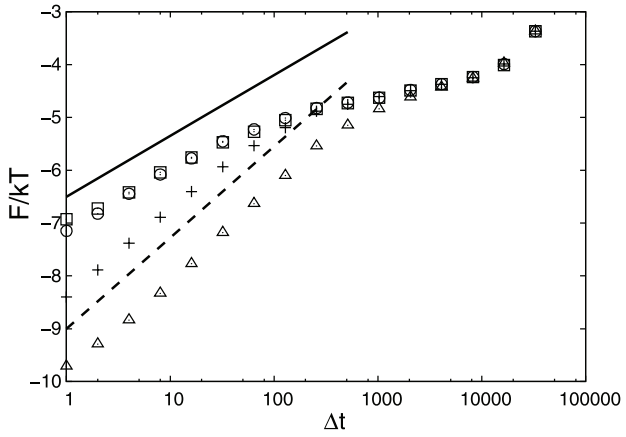

Figure 6. Scaling of F_C at the transition state with the sampling interval. The sampling intervals are $\Delta t = 2^0,...,2^{16}$. F_C of the TS are shown by symbols; notation as in Figure 2. The solid line shows the diffusive slope ($0.5 \ln \Delta t$) and the dashed line shows the sub-diffusive slope ($0.25 \ln \Delta t$) to guide the eye.
doi:10.1371/journal.pcbi.1000921.g006

Figure 6 shows F_C computed at the TS as a function of Δt for different reaction coordinates. Initially, F_C curves have a constant slope, which is close to diffusive for the optimized reaction coordinates and to sub-diffusive for the sub-optimal reaction coordinates. The slope changes when F_C approaches the limiting value of F_C^b. The latter is not strictly constant, though its dependence on Δt is rather weak. As Δt increases further ($\Delta t > 10000$, the mean life time in the basins), the probability of the system to visit another basin undetected (between successive sampling events) increases as well and Z_C (the number of detected transitions) decreases. F_C for different reaction coordinates fall on the same curve at sufficiently large Δt, i.e. $F_C^b(\Delta t)$ are the same when local differences between the coordinates become negligible.

However, Figure 6 shows that the ballistic time (t^b) for different reaction coordinates is slightly different, while in deriving Eq. 1 it was assumed to be constant. To take this into account we proceed as follows. The curve $F_C(\Delta t)$ (Figure 6) is approximated by two straight lines as $Z_C(\Delta t) \sim \Delta t^{\alpha-1}$ for Z_C less than the limiting value of Z_C^b ($\Delta t < t^b$) and constant $Z_C(\Delta t) = Z_C^b$ ($\Delta t > t^b$), where t^b is the time when dynamics becomes ballistic. Define $\Delta x(\Delta t) = \Delta x t^{\alpha}$, and $Z_C(\Delta t) = Z_C t^{\alpha-1}$; where Δx and Z_C denote, respectively, $\Delta x(\Delta t = 1)$ and $Z_C(\Delta t = 1)$. The diffusion coefficient is set to unity by transforming the reaction coordinate to the natural coordinate, which means that $\Delta x = 2/\sqrt{\pi}$. The time t^b can be estimated as time when mean absolute displacement is about the barrier width (w), i.e. $2/\sqrt{\pi}(t^b)^{\alpha} = w$. At this time $Z_C(\Delta t^b) = (t^b)^{\alpha-1} Z_C = Z_C^b$. Eliminating t^b from the two equations, one finds

$$\alpha = \frac{\ln(w\sqrt{\pi}/2)}{\ln(w\sqrt{\pi}/2) + \ln Z_C^b - \ln Z_C}. \qquad (2)$$

Taking $w = 13$ and $Z_C^b = 100$, one obtains α equal to 0.32, 0.39 and 0.49 for F_C/kT equal to -9.72, -8.34 and -7.13, respectively, in reasonable agreement with Figure 6. The ballistic times are 1924, 487 and 144, respectively. From Eq. 2 it follows that the higher is the free energy barrier the higher is the exponent α and the closer is the dynamics to the diffusive one.

The two optimized reaction coordinates, while having very different functional forms, show very similar behavior (at the TS regions), e.g. the width and the height of the TS barrier is the same ($x \approx 33$ on Figure 1b and $x \approx 40$ Figure 4b), the MSDs are identical (Figure 2) as well as $F_C(\Delta t)$ dependencies (Figure 6). This, likely, indicates that the two coordinates have converged to and closely approximate the true reaction coordinate (at the TS region). The residual difference between the estimated and the actual numbers of folding events which is due to limited statistics and insufficient flexibility of the chosen functional forms, is relatively small so it does not affect the results. The fact that the diffusive character of dynamics is determined by the height of free energy barrier F_C, rather than the chosen functional form of the coordinate indicates the robustness of the approach. It also means that the method of constructing the optimum reaction coordinate by optimizing its FEP (F_C) [4,6] has an advantage over the other approaches [7–9], in that it guarantees that the optimum reaction coordinate has dynamics closest to diffusive. Distribution of folding times is single exponential and identical for all four coordinates because folding events can be detected with high likelihood by any sufficiently good order parameter.

The analysis suggests that the higher is the free energy profile the closer is dynamics to diffusive. Evidently, the most optimal reaction coordinate is the one which has its free energy highest for every value of reaction coordinate. Consider invariant parametrization of reaction coordinate, namely the partition function of the configuration space from the initial value to the position x $Z_A(x) = \int_A^x Z_H(x)dx$. The optimum reaction coordinate is the one that attains $\max F_C(Z_A)$ or $\min Z_C(Z_A)$ for any Z_A, assuming that Z_C for different values of Z_A can be varied independently. This defines the optimum reaction coordinate introduced in [4], which has the largest mean first passage time. Conversely, diffusive dynamics on the constructed reaction coordinate can serve as an indication of optimality of the reaction coordinate.

To illustrate that the results presented are robust with respect to particular choice of the protein or the interaction potential, a protein with different secondary structure content (β-sheet) and an all-atom structure based model of the lambda repressor protein are analyzed in Text S1. The analysis confirms that the dynamics is sub-diffusive when projected onto a sub-optimum reaction coordinate and diffusive, when projected onto the optimum reaction coordinate.

Low free energy barrier *per se* does not mean that the dynamics is sub-diffusive, for example, a freely diffusing particle has flat free energy profile. Dynamics should be sub-diffusive, when the reaction coordinate is sub-optimal, i.e., the free energy barrier along the coordinate is much lower than the correct one. The latter is defined either as the highest barrier attained by the optimum reaction coordinates, or as a solution of the multidimensional minimum cut problem ($\min Z_C$), which locates the transition state [4].

The analysis above just considers the dynamics around the transition state, i.e., at the top of the free energy barrier. The conclusion that the higher the free energy profile the closer the dynamics to diffusive is likely to be valid in general, e.g., for the barrier-less folding proteins. The quantitative analysis exploits the fact that at the very large sampling intervals, when the system flies ballistically over the barrier, the two free energy profiles for optimal and sub-optimal reaction coordinates are very similar, because the two coordinates distinguish equally well between the basins. It can be extended to the following general qualitative argument. The two sufficiently good reaction coordinates likely differ significantly only at relatively small spatial scales with the large scale description of the dynamics being very similar. As the sampling interval Δt increases, the characteristic change of the reaction coordinates during the sampling interval ($\Delta x(\Delta t)$) increases as well. When ($\Delta x(\Delta t)$) is comparable to the large scale, so that the relative difference between the coordinate is negligible, the description of the dynamics by the two coordinates is similar and results in similar free energy profiles. Since the distance between the higher profile and the joint profile at large sampling intervals is smaller than that for the lower profile, the dynamics in former case is closer to diffusive compare to the later. It is assumed that $\Delta x(\Delta t) \sim \Delta t^{\alpha}$ is valid for the whole range of Δt from the small sampling intervals, when the dynamics start to manifests itself as (sub)diffusive to the large sampling intervals, where the profiles for the different reaction coordinates become very similar. This equation connects the dynamics and the free energy profiles at these different time scales.

The model of the protein employed in the analysis is relatively simple, thus allows for extensive simulation with large number of folding-unfolding events. More realistic simulation of protein folding would include explicit representation of solvent configuration degrees of freedom. The dynamics projected on the optimum reaction coordinate constructed by considering only protein degrees of freedom might be sub-diffusive because neglected solvent degrees of freedom could be important.

The analysis suggests that without specifying the reaction coordinate, the question why the dynamics is sub-diffusive is rather ill-posed. It is more appropriate to ask: is it possible, for a given

trajectory, to construct the optimum reaction coordinate, so that the projected dynamics is diffusive?

In conclusion, we have shown that dynamics projected onto a reaction coordinate can be diffusive or sub-diffusive depending on the coordinate employed for the projection. If one has a flexibility in choosing the reaction coordinate, e.g. when describing protein folding, dynamics can be made diffusive (or close to it) by optimizing the reaction coordinate (making F_C higher). When the coordinate describing the process is specified and can not be varied, for example, the donor-acceptor distance in the single molecule FRET or ET experiments [24,31] or the mean square displacement in the neutron scattering experiments [13], the dynamics is likely to be sub-diffusive [13,14,31]. However, this does not necessarily mean that the dynamics *per se* is sub-diffusive. A properly chosen reaction coordinate (too complex to realize in experiment) may show that dynamics of transition between free energy basins is diffusive. A relatively small deficiency of the putative reaction coordinate (difference in 1 kT in free energy (F_C) of the folding barrier) is sufficient to make the dynamics sub-diffusive. Hence, one should model protein dynamics as diffusion on a putative reaction coordinate [32,33] with care, because, it is very likely that the coordinate is sub-optimal, unless it has been specifically constructed (optimized) [5,28].

Methods

Free energy profiles

The conventional way to construct the FEP, given the projection of a trajectory onto a reaction coordinate (the time-series of the value of the reaction coordinate) $x(t)$, is to compute a histogram and estimate the partition function (probability density) as $Z_H(x)=N_x/\Delta X$, where N_x is the number of time-series points in bin x and ΔX is the size of the bin. The free energy can then be found as $F_H(x)/kT=-\ln(Z_H(x))$. The partition function of the cut based free energy profile [4] at point x is defined as the number of transitions through that point, i.e. $Z_C(x)=1/2\sum_t \Theta\{(x(t)-x)(x-x(t+\Delta t)\}$, where Δt is the sampling interval and $\Theta\{x\}$ is the Heaviside step function; $F_C(x)/kT=-\ln(Z_C(x))$. Assuming that the $F_H(x)$ is approximately constant on the distance of the mean absolute displacement $\langle|\Delta x(\Delta t)|\rangle$, one can derive the following expression

$$Z_C(x)=\langle|\Delta x(\Delta t)|\rangle Z_H(x)/2, \tag{3}$$

where $\langle|\Delta x(\Delta t)|\rangle=\langle|x(t+\Delta t)-x(t)|\rangle$ is the mean absolute displacement during sampling interval; for diffusive dynamics it gives $Z_C(x)=Z_H(x)\sqrt{D(x)\Delta t/\pi}$.

A reaction coordinate (x) with a variable diffusion coefficient can be transformed to coordinate (y), called the natural coordinate [4], so that the diffusion coefficient is constant and equal to unity, by numerically integrating $dy=\pi^{-1/2}Z_H(x)dx/Z_C(x)$; i.e. that $Z_C(y)=\pi^{-1/2}Z_H(y)$.

Other approaches have been suggested to characterize diffusive dynamics by computing the free energy profile together with the coordinate dependent diffusion coefficient [32,34,35]. It is not clear, however, if they can be used to characterize the sub-diffusive regime.

Reaction coordinate optimization

It is reasonable to assume that any "bad" choice of reaction coordinate, when different parts of the configuration space overlaps at projection onto this coordinate, will result in faster kinetics, i.e. in a smaller mean first passage time (mfpt). Clearly, the longest mfpt is obtained on the original FES or from a projection where no such overlapping occurs. Hence, we define the optimum reaction coordinate as the one that has the longest mfpt, which can be computed by Kramer's equation [4]

$$\begin{aligned}\langle t_{A\,B}\rangle &= \int_A^B dx \frac{e^{\beta F(x)}}{D(x)} \int_{-\infty}^{x} dy e^{-\beta F(y)} \\ &= \frac{\Delta t}{\pi}\int_A^B dx \frac{Z_H(x)}{Z_C^2(x)} \int_{-\infty}^{x} dy Z_H(y).\end{aligned} \tag{4}$$

The optimum reaction coordinates are constructed by numerically optimizing the mfpt functional for a sufficiently broadly chosen functional form of reaction coordinate. Starting with the initial set of parameters, which are sufficient to distinguish between the two free energy basins, the coordinate is iteratively improved by changing parameters and accepting the change if mfpt is increased. For the first reaction coordinate $X_1=\sum_{i<j} a_{ij}\Theta(\Delta_{ij}-r_{ij})$ we pick a random pair of atoms ij, scan the whole parameter space for the pair ($a_{ij}=\pm 1$ and $\Delta_{ij}=0,0.5,1.0,...,30$) and select the one that gives the highest mfpt. For the second reaction coordinate $X_2=\sum_{i<j}\alpha_{ij}r_{ij}$ we pick a random pair ij, scan the whole parameter space for the pair ($\alpha_{ij}=\pm 0.01(1+\xi/2)1.5^i$ for $i=1,2,...,20$ and ξ is a random number uniformly distributed between 0 and 1) and select the one that gives the highest mfpt. For the given values of parameters the mfpt is computed by first computing Z_H and Z_C and then numerically integrating Eq. 4. Alternatively one may minimize the number of transitions, the quantity related to mfpt as $N_{A\,B}=Z_A/\langle t_{A\,B}\rangle=\int_{-\infty}^{TS} Z_H(y)dy/\langle t_{AB}\rangle$, where Z_A is the partition function of basin A and $x=TS$ is the position of the transition state between basins A and B.

Subdiffusion

For subdiffusion, the mean absolute displacement no longer scales as $\Delta t^{1/2}$, but rather as $\langle|\Delta x(\Delta t)|\rangle\sim\Delta t^{\alpha}$. The exponent α (possibly coordinate dependent), can be determined by comparing $Z_C(x)$ at two different sampling intervals (see Eq. 3). For a trajectory with fixed length and varying sampling interval Δt (when $Z_H\sim\Delta t^{-1}$) it is equal to

$$\alpha(x)=1+\frac{\ln Z_C(x,\,\Delta t_1)-\ln Z_C(x,\,\Delta t_2)}{\ln\Delta t_1-\ln\Delta t_2}. \tag{5}$$

Since F_C is invariant with respect to nonlinear coordinate transformation, the scaling exponent α computed by Eq. 5 is also invariant, while α computed from $\langle|\Delta x(\Delta t)|\rangle$ or $\langle\Delta x^2(\Delta t)\rangle$ are not invariant and are computed here after the coordinate has been transformed to the natural reaction coordinate.

Supporting Information

Text S1 Supporting information for "Is Protein Folding Sub-Diffusive?".
Found at: doi:10.1371/journal.pcbi.1000921.s001 (0.08 MB PDF)

Acknowledgments

I am grateful to Emanuele Paci for providing trajectories of the Go model simulations.

Author Contributions

Conceived and designed the experiments: SVK. Performed the experiments: SVK. Analyzed the data: SVK. Contributed reagents/materials/analysis tools: SVK. Wrote the paper: SVK.

References

1. Dobson CM, Sali A, Karplus M (1998) Protein folding: A perspective from theory and experiment. Angew Chem Int Ed 37: 868–893.
2. Shea JE, Brooks CL (2001) From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. Annu Rev Phys Chem 52: 499–535.
3. Onuchic JN, Socci ND, Luthey-Schulten Z, Wolynes PG (1996) Protein folding funnels: the nature of the transition state ensemble. Fold Des 1: 441–450.
4. Krivov SV, Karplus M (2008) Diffusive reaction dynamics on invariant free energy profiles. Proc Natl Acad Sci USA 105: 13841–13846.
5. Krivov S, Karplus M (2004) Hidden complexity of free energy surfaces for peptide (protein) folding. Proc Natl Acad Sci USA 101: 14766–14770.
6. Krivov S, Karplus M (2006) One-Dimensional Free-Energy profiles of complex systems: Progress variables that preserve the barriers. J Phys Chem 110: 12689–12698.
7. Ma A, Dinner A (2005) Automatic method for identifying reaction coordinates in complex systems. J Phys Chem B 109: 6769–6779.
8. Best RB, Hummer G (2005) Reaction coordinates and rates from transition paths. Proc Natl Acad Sci USA 102: 6732–6737.
9. Maragliano L, Fischer A, Vanden-Eijnden E, Ciccotti G (2006) String method in collective variables: minimum free energy paths and isocommittor surfaces. J Chem Phys 125: 24106.
10. Mori H (1965) Transport, collective motion, and brownian motion. Progr Theor Phys 33: 423–455.
11. Zwanzig R (2001) Nonequilibrium Statistical Mechanics. New York: Oxford University Press.
12. Darve E, Solomon J, Kia A (2009) Computing generalized langevin equations and generalized FokkerPlanck equations. Proc Natl Acad Sci USA 106: 10884–10889.
13. Kneller GR (2005) Quasielastic neutron scattering and relaxation processes in proteins: analytical and simulation-based models. Phys Chem Chem Phys 7: 2641–2655.
14. Min W, Luo G, Cherayil BJ, Kou SC, Xie XS (2005) Observation of a Power-Law memory kernel for fluctuations within a single protein molecule. Phys Rev Lett 94: 198302.
15. Michalet X, Weiss S, Jager M (2006) Single-Molecule fluorescence studies of protein folding and conformational dynamics. Chem Rev 106: 1785–1813.
16. Luo G, Andricioaei I, Xie XS, Karplus M (2006) Dynamic distance disorder in proteins is caused by trapping. J Phys Chem 110: 9363–9367.
17. Matsunaga Y, Li C, Komatsuzaki T (2007) Anomalous diffusion in folding dynamics of minimalist protein landscape. Phys Rev Lett 99: 238103.
18. Senet P, Maisuradze GG, Foulie C, Delarue P, Scheraga HA (2008) How main-chains of proteins explore the free-energy landscape in native states. Proc Natl Acad Sci USA 105: 19708–19713.
19. Li C, Yang H, Komatsuzaki T (2008) Multiscale complex network of protein conformational fluctuations in single-molecule time series. Proc Natl Acad Sci USA. pp 536–541.
20. Neusius T, Daidone I, Sokolov IM, Smith JC (2008) Subdiffusion in peptides originates from the Fractal-Like structure of configuration space. Phys Rev Lett 100: 188103–4.
21. Granek R, Klafter J (2005) Fractons in proteins: Can they lead to anomalously decaying time autocorrelations? Phys Rev Lett 95: 098106.
22. Magdziarz M, Weron A, Burnecki K, Klafter J (2009) Fractional brownian motion versus the Continuous-Time random walk: A simple test for subdiffusive dynamics. Phys Rev Lett 103: 180602.
23. Sangha AK, Keyes T (2009) Proteins fold by subdiffusion of the order parameter. J Phys Chem 113: 15886–15894.
24. Kou SC, Xie XS (2004) Generalized langevin equation with fractional gaussian noise: Subdiffusion within a single protein molecule. Phys Rev Lett 93: 180603.
25. Lange OF, Grubmuller H (2006) Collective langevin dynamics of conformational motions in proteins. J Chem Phys 124: 214903–214918.
26. Metzler R, Klafter J (2000) Kramers' escape problem with anomalous kinetics: non-exponential decay of the survival probability. Chem Phys Lett 321: 238–242.
27. Karanicolas J, L IBC (2003) Improved Go-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. J Mol Biol 334: 309–325.
28. Allen LR, Krivov SV, Paci E (2009) Analysis of the free-energy surface of proteins from reversible folding simulations. PLOS Comp Biol 5: e1000428.
29. Bryngelson JD, Wolynes PG (1987) Spin glasses and the statistical mechanics of protein folding. Proc Natl Acad Sci USA 84: 7524–7528.
30. Kouza M, Li MS, O'Brien, Hu C, Thirumalai D (2006) Effect of finite size on cooperativity and rates of protein folding. J Phys Chem 110: 671–676.
31. Yang H, Luo G, Karnchanaphanurach P, Louie T, Rech I, et al. (2003) Protein conformational dynamics probed by Single-Molecule electron transfer. Science 302: 262–266.
32. Nettels D, Gopich IV, Hoffmann A, Schuler B (2007) Ultrafast dynamics of protein collapse from single-molecule photon statistics. Proc Natl Acad Sci USA 104: 2655–2660.
33. Mglich A, Joder K, Kiefhaber T (2006) End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. Proc Natl Acad Sci USA 103: 12394–12399.
34. Best RB, Hummer G (2006) Diffusive model of protein folding dynamics with Kramers turnover in rate. Phys Rev Lett 96: 228104.
35. Chahine J, Oliveira RJ, Leite VBP, Wang J (2007) Configuration-dependent diffusion can shift the kinetic transition state and barrier height of protein folding. Proc Natl Acad Sci USA 104: 14646–14651.