Supporting Information

Krivov and Karplus 10.1073/pnas.0800228105

SI Text

Depedence of F_C on dt for the Model System

shows how the F_C changes with the value of $dt = 2^0 \ldots 2^{20}$ for the model system in Fig. 1. At small dt, consecutive F_C values are displaced by a constant distance of $\Delta F = kT \ln(\sqrt{2}) = 0.25\ln(2) \approx 0.17$ as expected for diffusive motion. As dt increases, ΔF for the local minima becomes notably larger than 0.17. This can be understood by assuming that the average distance that the system jumps during dt is no longer determined by diffusive motion but, rather, by the size of the basin; this limit is valid for times $dt \sim \Delta x^2/D \approx 1/0.005$. For steps with constant length, F_C is proportional to $\sim 1/dt$; that is, $\Delta F = kT \ln(2) = 0.5\ln(2) \approx 0.34$. At the same time, when the step becomes larger than the size of the transition-state region, multiple recrossing events are no longer registered and the dynamics of the transition between basins is in the "ballistic" regime so that F_C is constant, which remains valid for times less than mean transition time between basins. On still-longer time scales, when the two basins essentially join into one, the whole picture repeats; that is, initially one sees an indication of diffusive motion ($\Delta F \approx 0.17$) over a smooth "averaged" profile (1) and on still-longer time scales the size of the jumps are limited by the size of system ($\Delta F \approx 0.34$).

Relation to the Rhee and Pande Approach

Recently, Rhee and Pande (2) presented an approach to transform the F_H along p_{fold} (commitor) as the reaction coordinate to a coordinate for which the diffusion coefficient is constant (i.e., the "natural coordinate," as defined above). Because their method and the present approach must give the same answer for p_{fold} as the reaction coordinate, the F_H along p_{fold} uniquely defines the F_C. To demonstrate this point, we use the notation used in ref. 2. Consider a one-dimensional system performing diffusive motion with a diffusion coefficient $D(x)$ on potential $U(x)$. For the profile expressed in terms of the original coordinate x, we have $Z_H(x) = e^{-\beta U(x)}$ and $Z_C(x) = Z_H(x)(D(x)dt/\pi)^{1/2} = (D(x)dt/\pi)^{1/2}e^{-\beta U(x)}$, as shown above. The p_{fold} reaction coordinate for a one-dimensional system as a function of x is (2)

$$p_{fold}(x) = \frac{\int_a^x 1/D(x)e^{\beta U(x)}}{\int_a^b 1/D(x)e^{\beta U(x)}}$$

The Z_H as a function of p_{fold} as the reaction coordinate is equal to

$$Z_H(p_{fold}) = Z_H(x(p_{fold}))/(dp_{fold}/dx) = cD(x(p_{fold}))e^{-2\beta U(x(p_{fold}))},$$

where $c = \int_a^b 1/D(x)e^{\beta U(x)}$ is a constant. For the F_C, one obtains

$$Z_C(p_{fold}) = Z_C(x(p_{fold})) = (D(x(p_{fold}))dt/\pi)^{1/2}e^{-\beta U(x(p_{fold}))}$$
$$= (dt/\pi/c)^{1/2}\sqrt{Z_H(p_{fold})};$$

that is, the Z_C is equal to the square root of Z_H multiplied by a constant, if they are both computed along p_{fold}. The equation for the natural coordinate $dy/dp_{fold} = Z_H(p_{fold})/Z_C(p_{fold}) \sim \sqrt{Z_H(p_{fold})} \sim Z_C(p_{fold})$, which corresponds to equation 19 of ref. 2. Although the method of Rhee and Pande leaves the

diffusion constant undetermined (they suggest determining it by examining the correlation function of the dynamics (3, 4) projected on the transformed coordinate (where the diffusion coefficient is constant), the approach presented here allows one to determine the profile together with the diffusion coefficient. Moreover, the method can be applied to any reaction coordinate, whereas that of Rhee and Pande is limited to the use of p_{fold} as the reaction coordinate, an important case, for which the F_H completely specifies the F_C.

Related Formulations

A number of papers (5–7) concerned with reactions have introduced alternative forms of the free energy expressed as a function of a chosen coordinate, the so-called reaction coordinate. That these alternatives are meaningful for a "reduced" free energy contrasts such a quantity with the free energy difference between two thermodynamically stable states; the latter is unique and independent of the choice of the measurement path between the two states. Such reduced path-dependent free energies were first introduced very early in the development of transition-state theory, which is based on defining an optimal dividing surface between reactants and products. In particular, in reference to transition-state theory (5), two free energies have been considered, one identical to $F_H(q)$ and the other with "dynamical" aspects related to but different from $F_C(q)$.

Most germane to the present development is an E and Vanden-Eijnden paper (6). One of the free energies of ref. 6 [$F(q)$] is the same as used by us (F_H), but the other [$G(q)$] of ref. 6 and the F_C of our paper are not. The more standard "free energy" $F_H(q)$, where q is the reaction coordinate, is defined in terms of the relative probabilities of the values of q. The other free energy, $F_C(q)$, is defined as the number of transitions through a given point q (surface in the multidimensional case) during the time interval dt. The relation between F_C and F_H is $\exp(-\beta F_C(q)) = \sqrt{D(q)dt/\pi} \exp(-\beta F_H(q))$, where $D(q)$ is a coordinate-dependent diffusion coefficient. The "corresponding" quantity of that in ref. 6, termed $G(q)$, is defined via an integral (equation 47 in ref. 6), which for the one-dimensional case is $\exp(-\beta G(q)) = |dq/dx| \exp(-\beta F(q))$ (equation 48 in ref. 6) with $|dq/dx|$ a Jacobian factor that makes $G(q)$ gauge-invariant, as is $F_C(q)$. Implicitly, E and Vanden-Eijnden assume a constant diffusion coefficient, as can be seen by comparing the equations; the value of the constant diffusion coefficient is not determined by their approach. The $F_C(q)$ takes into account the coordinate-dependent diffusion coefficient and, with a knowledge of F_H, makes possible the evaluation of that diffusion coefficient, thus completely determining the kinetics. This is the essential difference between the two developments. Moreover, this is why the development in our paper is more fundamental than what is presented by E and Vanden-Eijnden.

F_C and F_H are computed directly from the time series of the reaction coordinate evaluated with a given time scale. This is done irrespective of the precise formulation of the dynamics in the original space [what makes it more general than $G(q)$], which is valid only for Langevin dynamics. In particular, we have applied F_C to analyze Hamiltonian dynamics of a dipeptide in the diffusive regime, whereas $G(q)$ cannot be used for that. E and Vanden-Eijnden state: "The main remaining theoretical difficulty is associated with pure Hamiltonian dynamics. This is the case of most practical interest. We hope that some of the notions reviewed here can at least serve as a starting point for developing

approximations for transition pathways and transition rates in that case."

One could consider $G(q)$ as a limiting case of $F_C(q)$, for a constant diffusion coefficient and Langevin dynamics, even though it does not allow one to determine the diffusive constant. In this case, $D(q) = D(x)(dq/dx)^2 = D(dq/dx)^2$ and $\exp(-\beta F_C(q)) = \sqrt{D(q)dt/\pi}\exp(-\beta F_H(q)) = \sqrt{Ddt/\pi} \times |dq/dx| \exp(-\beta F_H(q))$ so that $G(q) = F_C(q) + \log(Ddt/\pi)/2\beta$ [i.e., in this limiting case, $G(q)$ and $F_C(q)$ differ just by the constant].

Transformation to the natural coordinate we introduce is a more general concept compared with the transformations described by equations 81 and 88 in ref. 6, because it will "undo" the possible coordinate dependence of the diffusion coefficient.

Apart from the fact that we consider general dynamics and introduce a coordinate-dependent diffusion coefficient, there is another major difference between the two formulations for the "variational" reaction coordinate. In section 11 of ref. 6 it is assumed that p_{fold} (equation 96) is the correct reaction coordinate and that one then only has to find the functional whose minimum has equation 96 as the solution. Thus, in the variational formulation of E and Vanden-Eijnden one tries to approximate p_{fold} but not necessarily the true reaction coordinate. In our variational principle, we make no assumptions about the nature of the reaction coordinate. It should be noted, also, that E and Vanden-Eijnden state: "Another nice property of the solution of (96) offers possible ways of generalizing the concept to other type of dynamics or to use more than one reaction coordinate." The set of other dynamics seems to be rather restricted, because equation 96 implicitly assumes that the diffusion coefficient is constant, as does equation 99 (6). For the case of Hamiltonian dynamics, the diffusion coefficient is not known *a priori* and should be determined on the basis of dynamics, as we do.

We also would like to add something to the debate in the mentioned articles about which of F_C ($G(q)$) or F_H ($F(q)$) is the "true" free energy. F_C is invariant with respect to nonlinear transformations, which allows one to correctly locate the transition state and partition the FES into basins. However, one has to use F_H to correctly describe equilibrium probabilities and define the kinetics as diffusion on the (F_H) profile with the transformed diffusion coefficient. This is why the natural coordinate, for which the two profiles are identical, is useful.

1. Zwanzig R (1988) Diffusion in a rough potential. *Proc Natl Acad Sci USA* 85:2029–2030.
2. Rhee YM, Pande VS (2005) One-dimensional reaction coordinate and the corresponding potential of mean force from commitment probability distribution. *J Phys Chem B* 109:6780–6786.
3. Socci ND, Onuchic JN, Wolynes PG (1996) Diffusive dynamics of the reaction coordinate for protein folding funnels. *J Chem Phys* 104:5860–5868.
4. 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.
5. Schenter GK, Garrett BC, Truhlar DG (2003) Generalized transition state theory in terms of the potential of mean force. *J Chem Phys* 119:5828–5833.
6. E W, Vanden-Eijnden E (2004) Metastability conformational dynamics, and transition pathways in complex systems. *Multiscale Modelling and Simulation*, eds Attinger S, Koumoutsakos P (Springer), pp 277–312.
7. Vanden-Eijnden E, Tal FA (2005) Transition state theory: Variational formulation, dynamical corrections, and error estimates. *J Chem Phys* 123:184103.
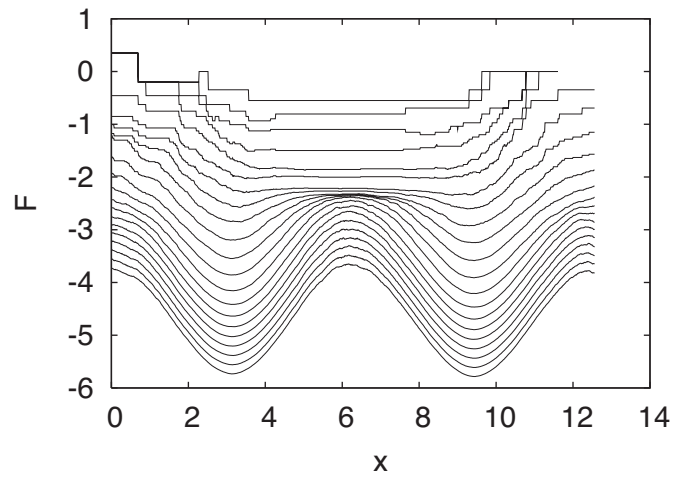
Fig. S1. $F_C(x)$ at various quench time intervals $dt = 2^0 \dots 2^{20}$ for the model system.