

# Water pulls the strings in hydrophobic polymer collapse

Gerhard Hummer\*

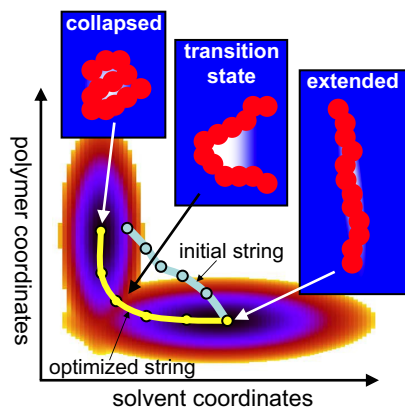
Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520

The molecular mechanisms of chemical reactions and conformational transitions are captured by their transition states and reaction coordinates. Transition states can be thought of as being intermediate between the reactant and product states, typically at elevated potential energy and correspondingly low population yet visited on transition paths between the two metastable states. Reaction coordinates are measures of the progress of a reaction.

To define transition states and reaction coordinates, one would like to invoke simple structural criteria, such as the length of a certain chemical bond or the value of a torsional angle. Such geometric approaches work well for low-dimensional reactions in gas phase. From maps of the energy surface, transition states can be identified as low-energy saddles connecting the reactant and product minima. Reaction coordinates then parameterize the paths connecting the minima along the valleys and over the low-lying saddles.

For reactions in solution, one may try to retain such simple geometric descriptions. The gas-phase energy surface would then be replaced by the free-energy surface (or potential of mean force) in the relevant solute degrees of freedom. The role of the solvent would be reduced to generating an effective friction along the valleys and saddles on that surface, leading us from transition state theory toward Kramers' theory of diffusive barrier crossing. However, electron transfer, arguably one of the simplest reactions in solution, shows the hazards present on this route. In the highly successful Marcus theory of nonadiabatic electron transfer (1), it is precisely the solvent medium that plays the dominant role. More specifically, solvent motions govern the fluctuations of the energy gap between the reactant and product states and thereby determine the reaction rate (2, 3).

Whenever solvent coordinates contribute to the reaction, simple intuition is prone to fail and finding appropriate transition states and reaction coordinates becomes extremely challenging. In effect, among all of the motions in the high-dimensional space of the solvent coordinates, one needs to identify those that are coupled to the reaction. Moreover, this coupling has to be causal (i.e., the motions in the solvent are essential



**Fig. 1.** Schematic free-energy surface of hydrophobic polymer collapse as a function of polymer and solvent degrees of freedom. At the transition state, water (blue) recedes from a kink near the middle of the polymer (red), forming a small “vapor bubble” (white). In the string method (6, 14), an initial guess for the reaction path (light blue), consisting of a sequence of configurations (circles), is iteratively optimized to find the string representing the dominant reaction path (yellow).

for the reaction to proceed and are not just a response to the progress of the reaction). Remarkably, even in molecular simulations of reactions, in which one can in principle follow every degree of freedom, it is far from trivial to identify the relevant solvent motions, simply because they are hidden in a sea of thermal noise (4, 5).

In a recent issue of PNAS, Miller *et al.* (6) study the collapse of a hydrophobic polymer in water from an extended, fully hydrated state to a compact state that largely excludes water (Fig. 1). They show that this seemingly simple molecular transition, driven by the hydrophobic effect, is closely connected to fluctuations in the local water density. Specifically, the hydrophobic collapse is found to be triggered by a transient local dewetting event in the elbow of a slightly kinked, but otherwise extended, polymer. A fluctuation in the local water density thus turns out to be the major rate-limiting event.

Hydrophobic polymer collapse has long served as a paradigm of the molecular self-assembly processes in biology, in particular the folding of proteins (7–11). Previous studies have focused on the thermodynamics of polymer collapse, finding that the collapse is driven by water-mediated hydrophobic interactions. In essence, water is a poor solvent

for nonpolar solutes such that a compact collapsed state of the hydrophobic polymer is favored over an extended, fully solvent-exposed state (12).

The central role of water in the thermodynamics of collapse leads one to ask whether water plays an equally important role in the kinetics of collapse, beyond providing the driving force. Indeed, ten Wolde and Chandler (13) showed some time ago that the formation of a “vapor bubble” was the key event at the transition state to collapse, a result obtained by using transition path sampling of polymer collapse in a coarse-grained, lattice-model solvent.

By studying polymer collapse in a detailed molecular model of water, Miller *et al.* (6) face multiple challenges: Collapse is a rare event; the transition time between extended and collapsed states is relatively long, reflecting the diffusive character of the transition; and there is no evident reaction coordinate describing the relevant solvent fluctuations. To overcome the resulting sampling issues, these authors use an elegant formalism that allows them to identify the dominant reaction paths. In the so-called “string method” of Maragliano *et al.* (14), the dynamics of the detailed molecular system is projected onto a space of chosen coarse coordinates, and one attempts to find the “string” that best represents the ensemble of reactive trajectories in this projected space. Such a string consists of a sequence of molecular conformations whose coordinates have been collectively optimized, primarily to follow valleys of low free energy in the coarse coordinates between the reactant and product states (Fig. 1). For polymer collapse, reactive trajectories connect extended conformations with compact ones. Following earlier work (13), Miller *et al.* project the solvent motions onto coarse-grained local water densities that jointly describe the fluctuating water density field. Overall, this procedure results in a representation with >100,000 coarse-collective coordinates in addition to the coordinates describing the polymer configuration.

Author contributions: G.H. wrote the paper.

The author declares no conflict of interest.

See companion article on page 14559 in issue 37 of volume 104.

\*E-mail: gerhard.hummer@nih.gov.

The choice of the coarse coordinates is critical. On one hand, these coordinates have to capture the kinetically relevant motions in the solvent. This aspect is closely connected with the assumption of Markovian dynamics in the coarse space (i.e., the lack of memory, so that the future time evolution depends only on the present state and not on the preceding history). If one leaves out coordinates that are not relaxing fast relative to the process of interest, the projected motion will suffer from strong memory effects. In general, increasing the dimensionality of the space will make the projected dynamics more Markovian. Ultimately, in full phase space, Langevin or Newtonian dynamics are initial-value problems and, thus, are strictly Markovian. On the other hand, the coordinates should be coarse enough so that the assumption of a single dominant pathway (or string) in the projected space (or, at least, of a single dominant saddle over which most of the reaction paths cross) is valid. Increasing the dimensionality makes it less likely that a single path or saddle dominates, because the resulting free-energy landscape will become more structured with many local minima and saddles.

If the coarse coordinates are relevant and sufficient, then a string representing the dominant reaction path can be found iteratively (6, 14). By using restrained simulations, similar to what is done in coarse molecular dynamics (15, 16), estimates are obtained for local gradients in free energy and various geometric terms. The resulting string is

represented by configurations in the space of coarse-collective coordinates. Along the string, one also obtains an estimate of the gradient in free energy and of the splitting or commitment probability (17). The resulting reaction pathway can thus be validated by running trajectories from the calculated transition state and comparing the fraction ending up on the two sides (here,

## The choice of the coarse coordinates is critical.

extended and compact) with the theoretical prediction of 50%. Indeed, Miller *et al.* (6) find a splitting of 30%/70%, which is remarkably close to the optimum considering the fact that the commitment probability changes very rapidly from 0% to 100% near the transition state (or stochastic separatrix) (5, 18, 19).

A relevant question for the string and other methods focused on the “most likely” paths (20–22) is how unique the resulting paths are. The number of paths in high-dimensional space is enormous. For a polymer in which each of the  $N$  bonds can be in, say, *cis* or *trans*, one would estimate  $\mathcal{O}(N!)$  distinct advancing paths between the *all-cis* and *all-trans* state in the polymer coordinates alone. With the number of paths being large, degeneracy in path space is likely, and it

seems possible that the most likely path, as defined by a certain integration measure (22), is not always a “typical” path. To address this question for the collapse of the hydrophobic polymer, Miller *et al.* (6) invoke the high symmetry of the resulting transition state structure. Starting from an asymmetric initial configuration, the string method produces a symmetric transition state with the bend at the center of the polymer. The authors thus argue that the string represents an important reaction path that accurately captures the average transition state configuration.

How relevant are solvent fluctuations in the kinetics of collapse, folding, and binding of proteins beyond the evident role of water in mediating the molecular interactions (23)? Despite advances in single-molecule techniques (24, 25), direct experimental investigations of this question face formidable challenges, simply because any rate-determining solvent fluctuations are likely small and highly transient. However, by using string methods (14, 26), path optimization (20–22), transition path sampling (4, 27), reaction coordinate optimization (5, 28), direct simulation (29, 30), or simple models (31, 32), theoretical studies will be able to explore the relevance of local solvent motions in proteins and other molecular systems.

This work was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health.

- Marcus RA, Sutin N (1985) *Biochim Biophys Acta* 811:265–322.
- Kuharski RA, Bader JS, Chandler D, Sprik M, Klein ML, Impey RW (1988) *J Chem Phys* 89:3248–3257.
- Hwang JK, Warshel A (1987) *J Am Chem Soc* 109:715–720.
- Bolhuis PG, Chandler D, Dellago C, Geissler PL (2002) *Annu Rev Phys Chem* 53:291–318.
- Best RB, Hummer G (2005) *Proc Natl Acad Sci USA* 102:6732–6737.
- Miller TF, III, Vanden-Eijnden E, Chandler D (2007) *Proc Natl Acad Sci USA* 104:14559–14564.
- Flory PJ (1963) *Principles of Polymer Chemistry* (Cornell Univ Press, Ithaca, NY).
- Bryngelson JD, Wolynes PG (1990) *Biopolymers* 30:177–188.
- Camacho CJ, Thirumalai D (1993) *Proc Natl Acad Sci USA* 90:6369–6372.
- Chan HS, Dill KA (1993) *J Chem Phys* 99:2116–2127.
- Levy Y, Onuchic JN (2006) *Annu Rev Biophys Biomol Struct* 35:389–415.
- Athawale MV, Goel G, Ghosh T, Truskett TM, Garde S (2007) *Proc Natl Acad Sci USA* 104:733–738.
- ten Wolde PR, Chandler D (2002) *Proc Natl Acad Sci USA* 99:6539–6543.
- Maragliano L, Fischer A, Vanden-Eijnden E, Ciccoliti G (2006) *J Chem Phys* 125:024106.
- Hummer G, Kevrekidis IG (2003) *J Chem Phys* 118:10762–10773.
- Kevrekidis IG, Gear CW, Hummer G (2004) *AICHE J* 50:1346–1355.
- Onsager L (1938) *Phys Rev* 54:554–557.
- Hummer G (2004) *J Chem Phys* 120:516–523.
- Berezukovskii A, Szabo A (2004) *J Chem Phys* 122:014503.
- Elber R, Ghosh A, Cardenas A (2002) *Acc Chem Res* 35:396–403.
- Faccioli P, Sega M, Pederiva F, Orland H (2006) *Phys Rev Lett* 97:108101.
- Adib AB (2007) *J Phys Chem B*, in press.
- Dill KA (1990) *Biochemistry* 29:7133–7155.
- Merchant KA, Best RB, Louis JM, Gopich IV, Eaton WA (2007) *Proc Natl Acad Sci USA* 104:1528–1533.
- Nettels D, Gopich IV, Hoffmann A, Schuler B (2007) *Proc Natl Acad Sci USA* 104:2655–2660.
- Ren W, Vanden-Eijnden E, Maragakis P, Weinan E (2005) *J Chem Phys* 123:134109.
- Juraszek J, Bolhuis PG (2006) *Proc Natl Acad Sci USA* 103:15859–15864.
- Ma A, Dinner AR (2005) *J Phys Chem B* 109:6769–6779.
- Rhee YM, Sorin EJ, Jayachandran G, Lindahl E, Pande VS (2004) *Proc Natl Acad Sci USA* 101:6456–6461.
- Liu P, Huang XH, Zhou RH, Berne BJ (2005) *Nature* 437:159–162.
- Cheung MS, Garcia AE, Onuchic JN (2002) *Proc Natl Acad Sci USA* 99:685–690.
- Hummer G, Garde S, Garcia AE, Paulaitis ME, Pratt LR (1998) *Proc Natl Acad Sci USA* 95:1552–1555.