Solving the Low Dimensional Smoluchowski Equation with a Singular Value Basis Set

Gregory Scott,[1] Martin Gruebele[1,2,3]

[1]Department of Chemistry, University of Illinois, Urbana, Illinois 61801, USA

[2]Department of Physics, University of Illinois, Urbana, Illinois 61801, USA

[3]Center for Biophysics and Computational Biology, University of Illinois, Urbana, Illinois 61801, USA

Abstract: Reaction kinetics on free energy surfaces with small activation barriers can be computed directly with the Smoluchowski equation. The procedure is computationally expensive even in a few dimensions. We present a propagation method that considerably reduces computational time for a particular class of problems: when the free energy surface suddenly switches by a small amount, and the probability distribution relaxes to a new equilibrium value. This case describes relaxation experiments. To achieve efficient solution, we expand the density matrix in a basis set obtained by singular value decomposition of equilibrium density matrices. Grid size during propagation is reduced from $(100–1000)^N$ to $(2–4)^N$ in N dimensions. Although the scaling with N is not improved, the smaller basis set nonetheless yields a significant speed up for low-dimensional calculations. To demonstrate the practicality of our method, we couple Smoluchowsi dynamics with a genetic algorithm to search for free energy surfaces compatible with the multiprobe thermodynamics and temperature jump experiment reported for the protein α_3D.

Key words: Fokker-Planck equation; singular value decomposition; genetic algorithm; protein folding; free energy surface

Introduction

Activation barriers are often so large that chemical reaction kinetics can be computed within the framework of master equations and transition state theory.[1,2] When the reaction barriers are comparable to k_BT, this approximation is no longer valid. The populations of reactants, intermediates, and products cannot be assigned neatly to "states,"and diffusion processes contribute directly to the observed signal. Methods are required that bridge from the macroscopic (state) to the microscopic (atom-by-atom) view, such as stochastic dynamics on low-dimensional free energy surfaces.[3]

For example, one can propagate individual trajectories by Langevin dynamics if the reaction occurs in the Kramers over-damped limit,[4] as is the case for most reactions in dissipative environments.[5] Langevin dynamics are most effective for comparison with short single molecule trajectories.[6] Comparison with ensemble relaxation experiments requires the averaging of many trajectories to obtain the time-dependent probability density *p(x,t)* along the reaction coordinate *x*.

Instead, one could solve directly the Fokker-Planck equation for the time-evolving probability density $p(\mathbf{x},\mathbf{p},t)$, or solve the Smoluchowski equation

$$\frac{\partial \rho}{\partial t} = \hat{O}(\mathbf{x})\rho = \frac{\partial}{\partial \mathbf{x}}\left\{ D(\mathbf{x}) e^{-\beta G(\mathbf{x})} \frac{\partial}{\partial \mathbf{x}} [e^{+\beta G(\mathbf{x})} \rho] \right\} \qquad (1)$$

for $p(\mathbf{x},t)$ if only the position distribution is of interest. Here G(**x**) is the free energy of reaction, D(**x**) is the diffusion tensor, $\beta = 1/k_B T$ is the inverse temperature, and the bold **x** indicates the appropriate summation over partial derivatives along *N* reaction coordinates $\mathbf{x} = \{x_1 \cdots x_N\}$. The equilibrium solution

$$\rho = \rho_0 e^{-G(\mathbf{x})/k_B T}, \int_{\text{all space}} d\mathbf{x}\rho = 1 \qquad (2)$$

is obtained readily. By comparison, relaxation of a nonequilibrium probability can be time consuming to compute even in a few dimensions for processes, such as ligand binding,[7] molecular force transduction,[8] relaxation, and alignment of nanostructures,[9] or protein folding,[10] where the dynamics cannot always be reduced to a single reaction coordinate. Even finite element methods require rather larger grid sizes,[7] making such calculations expensive if one needs to evaluate $p(\mathbf{x},t)$ very often, as for instance when fitting experimental data with an optimization algorithm.

We present an efficient method that propagates the Smoluchowski equation in a few dimensions. The method is designed to simulate relaxation processes, in which the free energy surface G is suddenly switched, and the probability distribution evolves to a new equilibrium. It relies on singular value decomposition[11,12] to produce an orthonormal basis set for the probability density p. For each propagation in time, the basis set transformation needs to be carried out only once. The time propagation itself occurs with a master equation-like propagator matrix grid typically 100–500 times smaller per dimension than a finite element grid, leading to savings of $>10^4$ in time for two dimensions. The reduction in basis set size makes the method suitable for combination with optimization algorithms that require large numbers of propagations.

We illustrate the method by fitting data from fast relaxation experiments. The synthetic three-helix bundle α_3D[13] has been observed to unfold by fast temperature jumps with infrared[14] and fluorescence detection.[10] Upon such a jump, the protein free energy $G(\mathbf{x})$ shifts slightly, and protein population relaxes rapidly to the new equilibrium. α_3D is a “downhill folder,”meaning its free energy barriers are $\sim k_B T$, thus master equations do not provide an accurate description of the dynamics. A genetic algorithm coupled with our singular value Smoluchowski propagation

of *p* optimizes free energy surfaces by comparison with thermodynamic and kinetic experimental data. The optimization requires that the probability distribution be propagated in time up to 10^6 times. We confirm that a 1-D free energy surface cannot account for the observed dynamics except in a very trivial model, whereas a 2-D free energy surface provides an adequate fit that matches the intuition derived from the experiments.

Method

Consider a free energy G(**x**,*d*) dependent on a perturbation parameter d. d could be the temperature, an applied force, or some other external variable. The perturbation is switched on at time t = 0, so the surface switches from G(**x**,0) to G(**x**,*d*). The idea is illustrated in Figure 1: the probability density starts out at equilibrium on the surface G(**x**,0), and will evolve after the jump to a new equilibrium on the surface G(**x**,*d*).

To construct a set of basis functions for propagating p in time, consider the set of density operators that solve eq. (1) at equilibrium for different values of *d*,

$$p_{\mathrm{eq}}(\mathbf{x},\mathrm{d}) = p_0(d)e^{-\beta G(\mathbf{x},d)} \qquad (3)$$

An optimal basis can be constructed by using singular value decomposition of this set. Let Gi = G(**x**,di) be one of *n* free energy surfaces where $d_i = d(i\text{-}1)/(n\text{-}1)$. *i* = 1 corresponds to the surface before the perturbation is turned on, i = n corresponds to the free energy surface after the perturbation is fully turned on. To each G_i belongs a $p_{eq}(\mathbf{x},d_i)$, as shown in Figure 1. After discretizing the *n* different p_{eq} (**x**,d_i), onto a suitable sampled coordinate grid $\mathbf{x}_j$, $j = 1\cdots m/N$ we can group the n vectors into a *m X n* matrix p_{eq}.[t] We then singular value decompose p_{eq} as

$$\rho_{\mathrm{eq}} = \rho^{SVD} w\, a^{\dagger} \qquad (4)$$

The matrix p^{SVD} has *n* orthonormal basis vectors p^{SVD} i=1...*n*($x_j^{=1\cdots}$m) as columns. The *n X n*

matrix *w* contains singular values to judge the importance of the basis vectors in p^{SVD}. The *n X n* matrix a[1] contains the orthonormal expansion coefficients of p_{eq} in terms of the basis vectors p^{SVD}.

The key savings is that *n « m* because the density operator tends to be much smoother than the coordinate grid required to converge integration of eq. (1). This is particularly true in relaxation experiments, where the surface G(**x**) is generally perturbed only by a small amount *d*. Furthermore, the matrix *w* provides an objective means for a cutoff to reduce basis set size. As *d* → 0, all but the first two singular values w_i rapidly approach zero. When fitting data with a signal-to-noise range SR, one needs to keep only singular values $w_i > w_{max}/SR$.

Expanding the density operator in terms of the singular value basis,

$$\rho(\mathbf{x},t)=\sum_{i=1}^{n}a_i(t)w_i\rho_i^{\mathrm{SVD}}(\mathbf{x}) \tag{5}$$

and inserting into eq. (1) we obtain

$$w_i\frac{\partial a_i}{\partial t}=\sum_{j=1}^{n}w_jG_{ij}a_j(t) \tag{6}$$

where the $n \times n$ propagator matrix G_{ij} is given by

$$G_{ij}=\int d\mathbf{x}\rho_i(\mathbf{x})\hat{O}(\mathbf{x})\rho_j(\mathbf{x}) \tag{7}$$

and the initial condition is given by

$$a_i(0)=(a^{\dagger})_{i1}. \tag{8}$$

The advantage of eq. (6) is that it reduces a large continuum propagation problem back to a very small master equation propagation. Instead of propagating state populations, the master equation propagates expansion coefficients of *p* in a small orthonormal basis. The matrix $G_{ij,}$ is expensive to calculate, but only needs to be computed once. The actual propagation over many

[1] The simplest grid would just be evenly spaced as shown in Figure 1. Importance-sampled grid or finite element grid are superior. Whichever way the coordinates are sampled and whatever the dimension N of the grid is, all the coordinates can simply be arrayed into a vector of length m into one of the columns of the matrix p_{eq}.

small time steps is inexpensive, and a back-calculation of $p(\mathbf{x},t)$ is only necessary at times t where a signal must be evaluated for comparison with data.

[Insert Figure 1]

Finally, there are practical considerations to optimize performance. To speed up the calculation, eq. (1) should be reduced to a form without exponentials, such as

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial \mathbf{x}} \left\{ \frac{\partial \beta G}{\partial \mathbf{x}} D(\mathbf{x}) \rho + D(\mathbf{x}) \frac{\partial}{\partial \mathbf{x}} \rho \right\}. \qquad (9)$$

The derivatives of G can be evaluated analytically if possible, together with the functions G(**x**,*d*) and D(**x**,*d*). In eq. (6), a large dynamic range w_j/w_i can cause stiffness problems for the differential equation solver. For typical $SR \leq 100$ encountered in experimental kinetics data, one should either select only $n' < n$ basis functions with $w_i/w_{max} > 0.01$ for the basis, or redo the singular value decomposition with a smaller n so $w_{min}/w_{max} > 0.01$. With that truncation of basis set size, we found that even a simple Runge-Kutta integrator is adequate. Mapping m^N position data from an evenly spaced grid sequentially into the row dimension of matrix p_{eq} performs well for 1–3 dimensions. For higher dimensions, a smarter mapping is needed. For example, the pyramidal algorithm[15] decomposes the free energy surface into wavelets and retains only those high frequency wavelet coefficients where rapid variation G(**x**,*d*) warrants it. Another alternative is to importance-sample the grid using –lnG(**x**) because the reduction to a master eq. (6) does not rely on any particular grid spacing or sampling.

Results and Discussion

Equation (6) couples the advantages of simple master equation propagation with the ability to calculate relaxation dynamics after switching an arbitrarily-shaped free energy surface at $t = 0$. Low barrier dynamics can be computed exactly for simple diffusion processes, without resorting

to transition state models. In effect, the master eq. (6) propagates orthogonal components of the density matrix instead of states.

[Insert Figures 2 and 3]

To demonstrate the utility of this approach, we applied it to a biophysical problem that requires calculation of many thermodynamic and kinetic data points to fit experimental data (Fig. 2). The Gai lab and we recently showed that fluorescence- and infrared-detected folding relaxation kinetics of the three-helix bundle protein α_3D have very different temperature dependences.[10] In that experiment, the protein solution was subject to a small temperature jump, and the protein population evolved on the free energy landscape towards a new equilibrium. The infrared-detected rate was nearly temperature-independent between 327 and 344 K, whereas the fluorescence-detected rate slowed down by more than a factor of 3 when the temperature was raised over the same range (Fig. 2A). When the protein thermally unfolded, infrared and fluorescence measurements yielded different unfolding curves in the 275–372 K range (Fig. 2B). No satisfactory 1-D fit was obtained by trial-and-error with Lange-vin dynamics and a diffusion coefficient fixed at 0.05 nm^2/ns,[10] the value for free diffusion of two small helices in solution.[16] We speculated that at least a 2-D surface would be required to fit the data.

Our goal here was to sample 1-D and 2-D model free energy surfaces and signal functions more exhaustively than was possible by Langevin dynamics. We combined our Smoluchowsi propagator with a genetic algorithm that evolved a family of up to 100 free energy surfaces. The genetic algorithm mutated and combined the free energy surface parameters for up to 3000 generations, selecting those surfaces that best reproduced the experimental data summarized in Figure 2. The experimental kinetics data contained 8 traces to be fitted (one for each rate

coefficient in Fig. 2A). Thus ~10^6 propagations of the probability function/density matrix in 1-D or 2-D were required during optimization.

[Insert Figure 4]

Each free energy surface was encoded as a sum of *k* Gaussians dimples of variable depth $A_i(d)$, variable anisotropic width $\sigma_i(d)$ and position $\mathbf{x_i}(d)$:

$$G(\mathbf{x}, d) = -\sum_{i=1}^{k} A_i(d) e^{-(\mathbf{x}-\mathbf{x}_i(d))^2/\sigma_i(d)^2} \qquad (10)$$

(The bold vectors in the exponent stand for a sum of squares over $\underline{N}$ coordinates.) Because only the relative well-depths and the barriers between wells are physically significant, we restricted the Gaussian wells to a minimum depth such that the normalized equilibrium density, given by eq. (2), approached zero at the edges of the sampling grid. In this study, we kept the diffusion coefficient coordinate-independent, but allowed its average value to vary.

To compute signals from $p(\mathbf{x},t)$ and $p_{eq}(\mathbf{x})$, the genetic algorithm also had to adjust signal functions $S_i(x)$ that describe how the infrared, thermal fluorescence, and kinetics fluorecence signals vary along the reaction coordinate. The signal functions $S_i(x)$ were chosen to be sigmoids with height *h*, width *r*, slope *m*, and switching at position x_o. In 1-D,

$$S_i(x) = \frac{h_i}{1 + e^{-(x-x_{i0})/\sigma_i}} + m_i x. \qquad (11)$$

We choose baseline sigmoids because they can represent both a gradual and a sudden shift in signal along the reaction coordinate. The signals $S_i(t)$ (Fig. 2A) or $S_i(T)$ (Fig. 2B) were obtained by integrating the time-evolving population distribution $p(\mathbf{x},t)$ or equilibrium population $p_{eq}(\mathbf{x},T)$ over the signal function as

$$S_i(t) = \int_{\mathbf{x}_{min}}^{\mathbf{x}_{max}} d\mathbf{x}\rho(\mathbf{x}, t)S_i(\mathbf{x}) \text{ or } S_i(T) = \int_{\mathbf{x}_{min}}^{\mathbf{x}_{max}} d\mathbf{x}\rho_{eq}(\mathbf{x}, T)S_i(\mathbf{x}) \quad (12)$$

In two dimensions, the sigmoid was directed along a vector **c** defined by $c_1x_1 + c_2x_2 = 0$, and a plane $m_1x_1 + m_2x_2$ was allowed to tilt with two slopes m_1 and m_2. The signal functions were truncated to $S \geq 0$ in both 1-D and 2-D simulations.

Within constraints to prevent physically unrealistic functions (e.g. $A_i < 0$), the genetic algorithm *pikaia*[17] evolved a family of free energy surfaces and signal functions to higher fitness to match the thermodynamic and kinetic signals summarized in Figure 2. In the genetic algorithm, a complete set of parameters such as m_i or A_i describing one free energy surface and its signal functions formed the 'genes' of one population member. Genes were subject to random mutation (change of value), or cross-over (exchange among population members). The fitness of population members in the genetic algorithm was determined by a weighted least squares comparison to the experimental data. Specifically, we maximized the fitness function

$$f = \frac{1}{\sum_i [(O_i - S_i)/\sigma_i]^2}$$

where O_i are the experimental data, S_i are the calculated signals, and σ_i are the relative uncertainties in the experimental data. Thermodynamic data points were fitted directly as shown in Figure 2A. Points from raw kinetic data traces such as Figure 2 in ref. 10 were fitted directly, and Figure 2B shows the resulting rate coefficients.

Figure 2 shows the calculated rates and thermodynamic traces of the fittest free energy surfaces from both the 1-D and 2-D simulations. At a first glance, the 1-D fit (dotted curves) appears to be slightly better than the 2-D fit (solid lines), but the 1-D fit was unsatisfactory from a physical point of view: the 1-D fit allows no interconversion of the folded and unfolded populations on the experimentally observed time scale of <10 μs.

Figure 3 illustrates the problem with the 1-D surface: the free energy barriers are up to 12 k_BT in height, requiring $k \approx (1\ s)^{-1}$ folding times, 10^5 times slower than the experimental rates in Figure 2.

The real experiment showed no evidence for a 1 s phase. The fast calculated phase that actually matches the experimental rates in Figure 2 came from diffusion of sub-populations that slightly shift within wells as the well positions and curvatures change. When we guarded against this solution by constraining the diffusion coefficient to be greater than 0.004 nm2/ns, the genetic algorithm could not find a 1-D solution that fit the data in Figure 2. $D \geq 0.004$ nm^2/ns yields folding speed limits in the $\leq$1 μs range, the diffusional contact time measured by protein and peptide dynamics experiments.[16,18–20] Our result here confirms the trial-and-error based conclusion in ref. 10, that physically reasonable diffusion coefficients cannot yield a 1-D solution.

In contrast, we were able to obtain a physically satisfactory 2-D free energy surface. Figure 4 shows the fittest 2-D free energy surface and equilibrium probability density at two temperatures. The fitted signals displayed in Figure 2 (solid lines) reproduce the experimental trends. The 2-D free energy surface supports a large population transfer between the native and unfolded states over low barriers. It has several shallow local minima at low temperature (313 K), through which α_3D can nearly fold downhill from U to N. The 2-D surface reproduced the experimental trends with a fitted diffusion coefficient of 0.004 nm^2/ns, 20 times closer to the range expected for contact formation in a helix bundle than the 1-D surface. This is still about 10 times less than the 0.05 nm^2/ns diffusion coefficient expected for freely diffusing helices. It is possible that a complete description of the α_3D folding dynamics will require either a rougher free energy surface (more local minima than shown in Figure 4), or additional reaction coordinates (more than the 2 in Figure 4).

The computations for Figure 4 required approximately 3 CPU hours on a 40 processor (3 GHz) Linux cluster. For this particular example, $n = 3$ basis functions were kept, compared to a grid with $m = 62{,}500$ total points in 2-D, reducing propagation time by about a factor of 10^4.

The same genetic algorithm optimization on the full grid would thus have been impractical with the computational resources utilized here.

The surface in Figure 4 is not a unique solution, but it is representative of the family of free energy surfaces compatible with the experimental data. One nice feature of the Smoluchowski-genetic algorithm approach is that the free energy and signal functions are easily refined further as additional experimental data become available. A direct comparison with low-dimensional free energy surfaces from Markov modeling of molecular dynamics simulations[21–23] is possible, but will require one additional major step. Our reaction coordinates in Figure 4 are really defined through the signal functions $S_i(\mathbf{x})$. Signal functions for the same variables would have to be computed from molecular dynamic simulation, so the two sets of free energy surfaces can be mapped onto one another.

Conclusions

Singular value decomposition of equilibrium density matrices provides a robust orthonormal basis set for propagating the non-equilibrium density matrix with the Smoluchowski equation. A large number of spatial grid points is reduced back to a small master equation propagation that can be integrated stably. The number of basis functions and dynamic range of singular value coefficients w_{max}/w_{min} can be adjusted to match the experimental signal-to-noise ratio. Simulation of relaxation experiments with small perturbations (e.g. temperature jumps) is about 100X faster per degree of freedom than grid or finite element methods. A genetic algorithm exploration of free energy surfaces and signal functions confirmed that the folding/unfolding kinetics and thermal melts of the designed three-helix bundle α_3D require at least a 2-D free energy surface to be modeled with a realistic diffusion coefficient and large population transfer from the native to the unfolded well.

Acknowledgment

MG and GS would also like to thank the Eiszner family.

References

1. Berne, B. J. In Activated Barrier Crossing: Applications in Physics, Chemistry and Biology; Hänggi, P.; Fleming, G. R., Eds.; World Scientific: Singapore, 1993; pp. 82–119.
2. Chandler, D. Modern Statistical Mechanics, Oxford University Press: Oxford, 1989.
3. Adelman, S. A.; Brooks, C. L. J Phys Chem 1982, 86, 1511.
4. Guo, Z. Y.; Thirumalai, D. Biopolymers 1995, 36, 83.
5. Hänggi, P.; Talkner, P.; Borovec, M. Rev Mod Phys 1990, 62, 251.
6. Liu, F.; Gruebele, M. J Chem Phys 2009, 131, 195101.
7. Song, Y. H.; Zhang, Y. J.; Shen, T. Y.; Bajaj, C. L.; McCammon, A.; Baker, N. A. Biophys J 2004, 86, 2017.
8. Yang, Q. Z.; Huang, Z.; Kucharski, T. J.; Khvostichenko, D.; Chen, J.; Boulatov, R. Nat Nanotechnol 2009, 4, 302.
9. Shaver, J.; Parra-Vasquez, A. N. G.; Hansel, S.; Portugall, O.; Mielke, C. H.; Von Ortenberg, M.; Hauge, R. H.; Pasquali, M.; Kono, J. ACS Nano 2009, 3, 131.
10. Liu, F.; Dumont, C.; Zhu, Y. J.; Degrado, W. F.; Gai, F.; Gruebele, M. J Chem Phys 2009, 130,
11. Golub, G. H.; van Loan, C. F. Matrix Computations; The Johns Hopkins University Press: Baltimore, 1996.
12. Press, W. H.; Flannery, B.P.; Teukolsky, S. A.; Vetterling, W. T. Numerical Recipes in Fortran; Cambridge University Press: New York, 1992.
13. Bryson, J. W.; Desjarlais, J. R.; Handel, T. M.; Degrado, W. F. Protein Sci 1998, 7, 1404.
14. Zhu, Y.; Alonso, D. O. V.; Maki, K.; Huang, C.-Y.; Lahr, S. J.; Daggett, V.; Roder, H.; Degrado, W. F.; Gai, F. Proc Natl Acad Sci USA 2003, 100, 15486.
15. Chui, C. K. An Introduction to Wavelets; Academic Press: New York, 1992.
16. Yang, W. Y.; Gruebele, M. Nature 2003, 423, 193.
17. Charbonneau, P. Astrophys J 1995, 101, 309.
18. Hagen, S. J.; Hofrichter, J.; Szabo, A.; Eaton, W. A. Proc Natl Acad Sci USA 1996, 93, 11615.
19. Lapidus, L. J.; Eaton, W. A.; Hofrichter, J. Proc Natl Acad Sci USA 2000, 97, 7220.
20. Bieri, O.; Wirz, J.; Hellrung, B.; Schutkowski, M.; Drewello, M.; Kiefhaber, T. Proc Natl Acad Sci USA 1999, 96, 9597.
21. Hubner, I. A.; Deeds, E. J.; Shakhnovich, E. I. Proc Nat Acad Sci USA 2006, 103, 17747.
22. Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C. J Chem Phys 2007, 126,
23. Noé, F.; Fischer, S. Curr Opin Struct Biol 2008, 154.

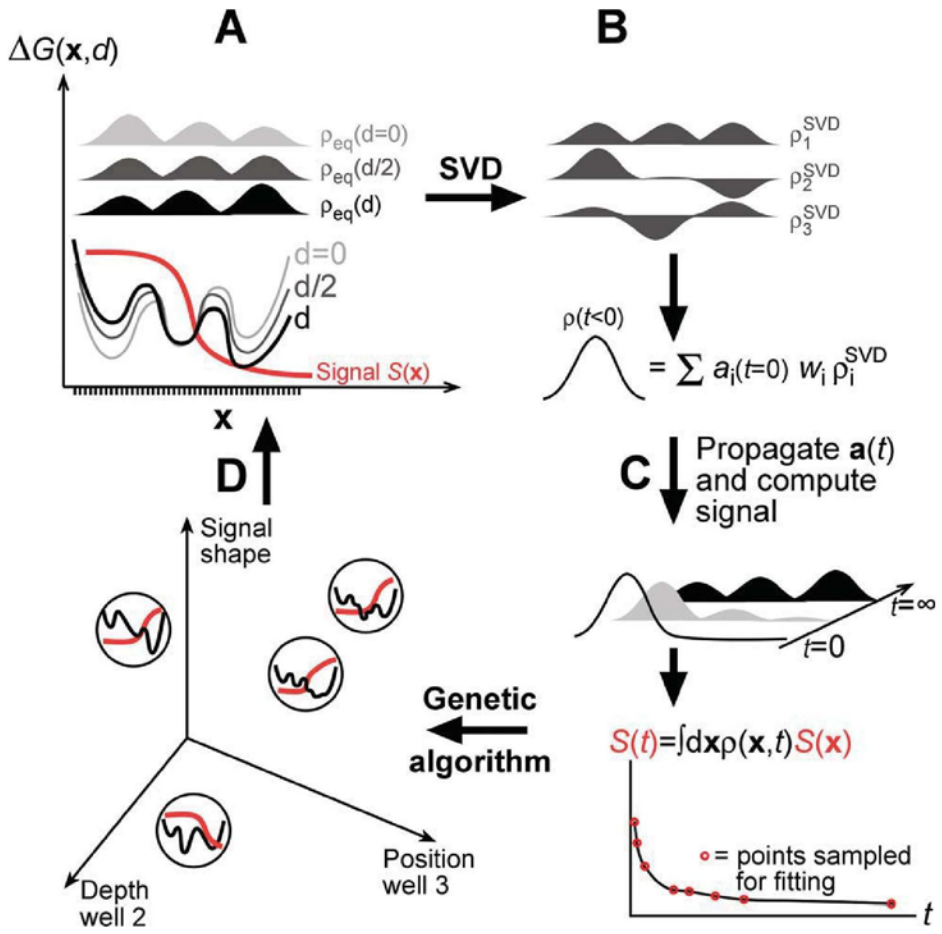

Figure 1. Schematic of the combined Smoluchowski propagation and genetic algorithm optimization. (A) The free energy $G(\mathbf{x},d)$ will be jumped from $d = 0$ (light) to d (black); $n = 3$ surfaces and the corresponding ρ_{eq} are calculated on a grid "x" Signal functions $S(\mathbf{x})$ are defined. (B) The matrix obtained from ρ_{eq} is singular value decomposed, yielding a small basis with expansion coefficient vector $\boldsymbol{a}(t)$ for $\rho(\mathbf{x},t)$. (C) $\boldsymbol{a}(t)$ is propagated by a master equation; signals are evaluated by integrating $S(t) = \int d\mathbf{x}\ \rho(\mathbf{x},t)\ S(\mathbf{x})$ only at time points sampled for data fitting. (D) Signals from a family of free energy surfaces/signal functions are compared with data by a genetic algorithm that selects the 'fittest' as well as creates new family members for the next generation. The procedure is then repeated until highly 'fit' surfaces that match the data well emerge.

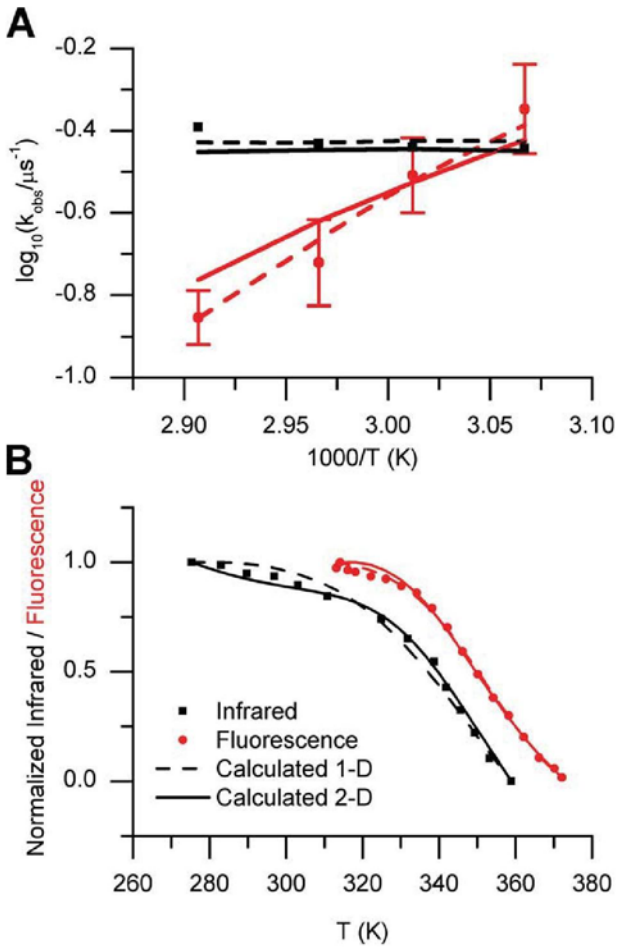

Figure 2. Measured and computed signals for α_3D. (A) Infrared kinetic rates (black points) and fluorescence kinetic rates (red points), compared to a 1-D (dashed) and 2-D (solid) fit. The 2-D fit has a more realistic diffusion coefficient (see text). The actual kinetic data fitted consisted of relaxation decays sampled at many data points, see for example Figure 2 in Ref. 10. (B) Infrared thermal unfolding data and fluorescence thermal unfolding data and fits using the same labeling.

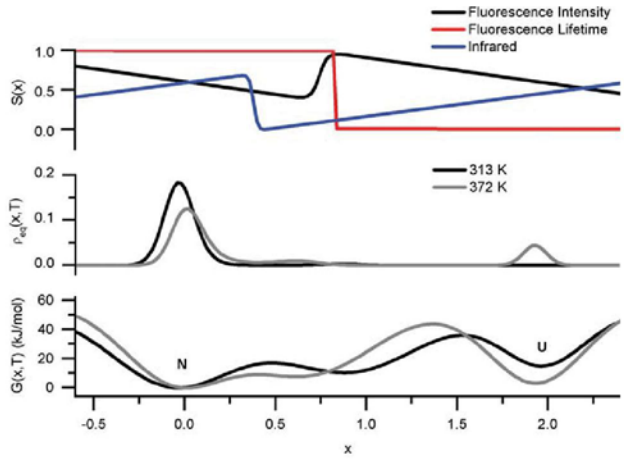

Figure 3. 1-D signal functions (A), equilibrium probability distribution at two temperatures (313 and 372 K) (B), and best free energy surface discovered by the genetic algorithm (C). The native well is labeled 'N' and the least folded well 'U.' The reaction coordinate corresponds roughly to the change in radius of gyration from the native value in nm.

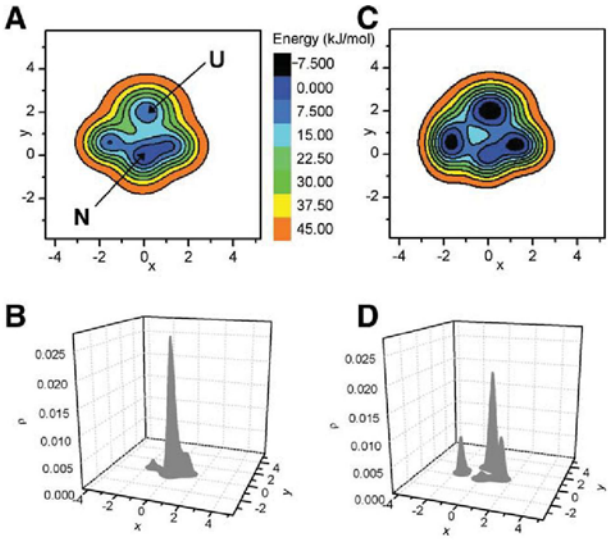

Figure 4. 2-D free energy surface at two different temperatures (313 and 372 K), together with equilibrium populations. The native well is labeled 'N' and the least folded well 'U.' The temperature range covers the thermal titration detected by fluorescence in Figure 2B.