Supporting Information: Characterization of a Dynamic String Method for the Construction of Transition Pathways in Molecular Reactions

Margaret E. Johnson and Gerhard Hummer*

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

*Email: Gerhard.Hummer@NIH.GOV

S1. Coordinate Transformations of Smoluchowski functions.

When transforming to a new system, the basis vectors $\mathbf{g}_i$ of the transformed coordinate system can be written as a linear combination of the original Cartesian standard basis $\hat{\mathbf{x}}_j$,

$$\mathbf{g}_i = \sum_j \frac{\partial x_j}{\partial z_i} \hat{\mathbf{x}}_j \qquad \text{(S1)}$$

(covariant basis vectors). The new basis vectors are in general local and dependent on the coordinates, as they would be for example in a transformation to polar coordinates.

Probability distributions (both equilibrium and time dependent) are scaled by the determinant of the Jacobian of the coordinate transformation[1],

$$p^z(\vec{z},t) = \left|\mathbf{J}_z(\vec{z})\right| p(\vec{x},t)_{\vec{x}=\vec{x}(\vec{z})} = \left|\mathbf{J}_x(\vec{x})\right|^{-1} p(\vec{x},t)_{\vec{x}=\vec{x}(\vec{z})}, \quad \text{(S2)}$$

where $J_{z,ij} = \partial x_i / \partial z_j$ and $J_{x,ij} = \partial z_i / \partial x_j$. This Jacobian factor is included in the probability distribution to retain the functional form of the Smoluchowski equation. For example, in polar coordinates the integral over the probability would be written as $\int dr d\theta p(r,\theta)$, consistent with the form of the Cartesian integral. The 'missing' familiar volume element r is absorbed into $p(r,\theta)$. Importantly, this additional Jacobian factor implies that although the potential transforms as a scalar, from the definition of the equilibrium probability, we see that:

$$p^z_{eq}(\vec{z}) = Z^{-1} \exp\left(-\beta V(\vec{x})\right) \left|\mathbf{J}_z(\vec{z})\right| = Z^{-1} \exp\left(-\beta (V(\vec{z}) - \beta^{-1} \ln\left|\mathbf{J}_z(\vec{z})\right|)\right)_{\vec{x}=\vec{x}(\vec{z})}. \quad \text{(S3)}$$

Hence, when one solves the Smoluchowski equation in the *z* coordinate space, the effective potential sampled is modified compared to the potential sampled in the original *x* coordinate space by,

$$V^z(\vec{z}) = \left(V(\vec{x}) - \beta^{-1} \ln\left|\mathbf{J}_z(\vec{z})\right|\right)_{\vec{x}=\vec{x}(\vec{z})} . \quad \text{(S4)}$$

The transformed diffusion tensor is given by[1,2] $D_{ij}(\vec{z}) = \sum_{lm} \frac{\partial z_i}{\partial x_l} \frac{\partial z_j}{\partial x_m} D_{lm}(\vec{x})_{\vec{x}=\vec{x}(\vec{z})}$, or in matrix notation:

$$\mathbf{D}^z(\vec{z}) = \mathbf{J}_x(\vec{x})\mathbf{D}(\vec{x})\mathbf{J}_x(\vec{x})^T_{\vec{x}=x(\vec{z})} . \quad \text{(S5)}$$

The derivatives of a scalar function *f* are related through the chain rule,

$\frac{\partial f}{\partial z_i} = \sum_j \frac{\partial f}{\partial x_j} \frac{\partial x_j}{\partial z_i}$, and can be written compactly as

$$\frac{\partial}{\partial \vec{z}} = \mathbf{J}_z(\vec{z})^T \frac{\partial}{\partial \vec{x}} . \quad \text{(S6)}$$

Again, the notation of the derivative will retain the same form in all coordinate systems, hence the avoidance of using the coordinate system-specific grad operator. Also, to calculate the change in the free energy *F* along some coarse variable of interest, the derivative is taken as in eq S6. For instance, to follow the gradient in the space of Φ, Ψ dihedral angles of a peptide backbone, positions are changed according to the vector ($\partial F/\partial\Phi$, $\partial F/\partial\Psi$).

S2. Committor Along a Path

For isotopic diffusion tensors, the equation for the committor (eq 4) simplifies to

$$\sum_i \frac{\partial q}{\partial x_i}\left(-\beta D(x)\frac{\partial V}{\partial x_i}+\frac{\partial D(x)}{\partial x_i}\right)+D(x)\sum_i \frac{\partial^2 q}{\partial x_i^2}=0. \quad \text{(S7)}$$

In isotropic systems both the potential and diffusion are scalar functions of the coordinates, and hence one can define the drift vector in terms of an effective potential, $W(x)=V(x)-\beta^{-1}\ln(D(x))$. The drift vector can now be written with components $\text{drift}_i=-\beta D(x)\frac{\partial W}{\partial x_i}$. We parameterize a path $\vec{x}(s)$ by its arc length varying from *0* to s_f. The unit tangent to a path pointing towards s_f has components dx_i/ds.

To derive the equation for *q* along a path, we start by noting the general relation for the change in *q* along a path parameterized by its arc length *s*:

$$\frac{dq}{ds}=\sum_i \frac{\partial q}{\partial x_i}\frac{dx_i}{ds}. \quad \text{(S8)}$$

Along the path tangent to $\partial q/\partial \vec{x}$ one therefore has $||\partial q/\partial \vec{x}||=dq/ds$ (where ||...|| denotes the Euclidian vector norm), because the tangent vector satisfies $\frac{dx_i}{ds}=\frac{\partial q}{\partial x_i}\left\|\partial q/\partial \vec{x}\right\|^{-1}$ and *q* is increasing along the path to s_f. The first sum in eq S7 is a dot product between the vectors $\partial q/\partial \vec{x}$ and **drift**. This dot product is therefore equivalent to $-\beta D(x)\sum_i \frac{\partial q}{\partial x_i}\frac{\partial W}{\partial x_i}=-\beta D(x)\frac{dq}{ds}\left\|\frac{\partial W}{\partial \vec{x}}\right\|\cos(\phi(s))$, where $\phi(s)$ is the angle between the drift vector and the tangent to the path. However, using eq S8 for *W* instead of *q*, we see that $\left\|\frac{\partial \mathrm{W}}{\partial \vec{x}}\right\|\cos(\phi(s))=\frac{d\mathrm{W}}{ds}$, since the tangent vectors have unit length. Therefore, the first sum in eq S7 simplifies to $\sum_i \frac{\partial q}{\partial x_i}\text{drift}_i=-\beta D(x)\frac{dq}{ds}\frac{d\mathrm{W}}{ds}$.

For the second sum in eq S7 we first define the second derivative with respect to *s*:

$$\frac{d^2q}{ds^2} = \sum_{ij} \frac{dx_i}{ds}\frac{dx_j}{ds}\frac{\partial^2 q}{\partial x_i \partial x_j} + \sum_i \frac{\partial q}{\partial x_i}\frac{d^2 x_i}{ds^2}. \quad \text{(S9)}$$

Since $\partial q / \partial \vec{x}$ is tangential to the path, the second sum is a dot product between the tangent and the normal of the path (components d^2x_i/ds^2) and because these two vectors are always perpendicular, the sum is zero. The results up to this point are exact. To simplify the first sum in eq S9, we assume that the isocommittor surface around the path remains perpendicular for finite width, so that as one moves a small distance along the isocommittor surface, the gradient of *q* has the same magnitude as it does along the path. Then taking a discrete limit of the gradients of *q* around the path demonstrates that this sum is equivalent to $\sum_i \frac{\partial^2 q}{\partial x_i^2}$ if the path is straight.

Comparing with eq S7, we see that, under the above assumptions, we have $-\beta \frac{dW}{ds}\frac{dq}{ds} + \frac{d^2q}{ds^2} = 0$ along the path. Integration with $q(0) = 0$ gives the result in eq 12 of the main text.

Without the above assumption, we can write,

$$\frac{d^2q}{ds^2} = \sum_i \frac{d^2q}{dx_i^2} + \left[\sum_{ij} \frac{dx_i}{ds}\frac{dx_j}{ds}\frac{d^2q}{dx_i dx_j} - \sum_i \frac{d^2q}{dx_i^2} \right] \quad \text{(S10)}$$

The term in brackets captures changes in the curvature of *q* along the path due to the non-zero curvature of the path itself. With the condition that the isocommittor surfaces remain perpendicular for some epsilon around the path, this term in brackets is zero for straight paths. Hence eq 12 of the main text is exact for straight

paths. If we assume this deviation term in brackets is a function of the path rather than the committor function, and define it along the path as *g(s)*, then comparing with eq S7 we get $-\beta\frac{dW}{ds}\frac{dq}{ds}+\frac{d^2q}{ds^2}=g(s)$. The solution to this equation is given by

$$q(s)=\frac{\int_0^s ds'\exp(+\beta W(s'))\left[1+\int_0^{s'} ds''\exp(-\beta W(s''))g(s'')\right]}{\int_0^{sf} ds'\exp(+\beta W(s'))\left[1+\int_0^{s'} ds''\exp(-\beta W(s''))g(s'')\right]}. \quad \text{(S11)}$$

For straight paths *g(s)*=0 and the equation simplifies to eq 12 of the main text.

S3. Equation for the MRP with arbitrary diffusion.

Berkowitz *et al.*[3] solved for the MRP solution by using the Euler-Lagrange formalism to maximize the flux over all possible paths of different arc lengths s_f. They specifically solved the resulting equation for cases with isotropic diffusion. To extend their formalism to general diffusion tensors, we reproduce here their Euler-Lagrange equations

$$0=\frac{d}{dt}\left[\frac{\partial}{\partial\dot{x}_k}\left(\exp(\beta V)\sum_{ij}\dot{x}_i\dot{x}_j f_{ij}/\left(\sum_i\dot{x}_i^2\right)^{1/2}\right)\right]-\frac{\partial}{\partial x_k}\left[\exp(\beta V)\sum_{ij}\dot{x}_i\dot{x}_j f_{ij}/\left(\sum_i\dot{x}_i^2\right)^{1/2}\right] \quad \text{(S12)}$$

where *k* is the dimension component out of *N* dimensions, the friction tensor **f** is related to **D** through **fD**=k_BT, the dot indicates differentiation with respect to *t*, which varies from 0 to 1 along the path. The actual path length of the optimal path is the arc length *s* related to *t* through $\frac{ds}{dt}=\left(\sum_i^N\frac{dx_i}{dt}^2\right)^{1/2}$. An important constraint when solving for the MRP is that at each point along the string, the arc length has

been properly chosen such that $\sum_i^N \left(\frac{dx_i}{ds}\right)^2 = 1$. This means that $\sum_i^N \frac{dx_i}{ds}\frac{d^2x_i}{ds^2} = 0$.

The second derivatives are defined by $\frac{d^2s}{dt^2} = \frac{ds}{dt}^{-1} \sum_i^N \frac{dx_i}{dt}\frac{d^2x_i}{dt^2}$, and

$$\frac{d^2x_i}{dt^2} = \left(\frac{ds}{dt}\right)^2 \frac{d^2x_i}{ds^2} + \frac{dx_i}{ds}\frac{d^2s}{dt^2}.$$

Expanding eq S12 and using the additional relations, one obtains the equations

$$0 = -\sum_{ij} \frac{dx_i}{ds}\frac{dx_j}{ds}\frac{\partial f_{ij}}{\partial x_k} - \frac{\partial \beta V}{\partial x_k}\sum_{ij}\frac{dx_i}{ds}\frac{dx_j}{ds} f_{ij} + 2\frac{d\beta V}{ds}\sum_i \frac{dx_i}{ds} f_{ik} - \frac{dx_k}{ds}\frac{d\beta V}{ds}\sum_{ij}\frac{dx_i}{ds}\frac{dx_j}{ds} f_{ij}$$

$$-\frac{dx_k}{ds}\sum_{ij}\frac{dx_i}{ds}\frac{dx_j}{ds}\frac{df_{ij}}{ds} + 2\sum_i f_{ik}\frac{d^2x_i}{ds^2} - \frac{d^2x_k}{ds^2}\sum_{ij}\frac{dx_i}{ds}\frac{dx_j}{ds} f_{ij} - 2\frac{dx_k}{ds}\sum_{ij} f_{ij}\frac{dx_i}{ds}\frac{d^2x_j}{ds^2}. \quad \text{(S13)}$$

If the diffusion is diagonal but anisotropic, the equation simplifies to

$$\frac{d^2x_k}{ds^2}\left(2f_{kk} - \sum_i f_{ii}\left(\frac{dx_i}{ds}\right)^2\right) + \frac{dx_k}{ds}\left[\frac{d\beta V}{ds}\left(2f_{kk} - \sum_i f_{ii}\left(\frac{dx_i}{ds}\right)^2\right) + 2\frac{df_{kk}}{ds} - \sum_i \frac{df_{ii}}{ds}\left(\frac{dx_i}{ds}\right)^2\right]$$

$$-2\frac{dx_k}{ds}\left(\sum_i \frac{dx_i}{ds} f_{ii}\frac{d^2x_i}{ds^2}\right) - \frac{\partial \beta V}{\partial x_k}\sum_i\left(\frac{dx_i}{ds}\right)^2 f_{ii} - \sum_i\left(\frac{dx_i}{ds}\right)^2\frac{\partial f_{ii}}{\partial x_k}. \quad \text{(S14)}$$

For isotropic diffusion, this again simplifies to eq 7 of the main text. It is important that the path be found as a function of *s* rather than *t*, because the equations for the components in *t* are not all linearly independent.

(1) Risken, H. *The Fokker-Planck Equation*; Springer: Berlin, 1996.
(2) Moro, G. J.; Cardin, F. *Phys. Rev. E* **1997**, *55*, 4918-4934.

(3) Berkowitz, M.; Morgan, J. D.; McCammon, J. A.; Northrup, S. H. *J. Chem. Phys.* **1983**, *79*, 5563-5565.