

Reaction paths based on mean first-passage times

Sanghyun Park

Beckman Institute and Department of Physics, University of Illinois at Urbana–Champaign, Urbana, Illinois 61801

Melih K. Sener

Beckman Institute, University of Illinois at Urbana–Champaign, Urbana, Illinois 61801

Deyu Lu and Klaus Schulten^{a)}

Beckman Institute and Department of Physics, University of Illinois at Urbana–Champaign, Urbana, Illinois 61801

(Received 9 January 2003; accepted 6 March 2003)

Finding representative reaction pathways is important for understanding the mechanism of molecular processes. We propose a new approach for constructing reaction paths based on mean first-passage times. This approach incorporates information about all possible reaction events as well as the effect of temperature. As an application of this method, we study representative pathways of excitation migration in a photosynthetic light-harvesting complex, photosystem I. The paths thus computed provide a complete, yet distilled, representation of the kinetic flow of excitation toward the reaction center, thereby succinctly characterizing the function of the system. © 2003 American Institute of Physics. [DOI: 10.1063/1.1570396]

I. INTRODUCTION

In many chemical and biological reactions, initial (reactant) and final (product) states are known but reaction pathways connecting the two are not. Examples range in complexity from single-particle Brownian motion to conformational changes of proteins, such as protein folding.^{1,2} Finding reaction pathways is one of the most fundamental challenges in chemistry and molecular biology.^{3,4} It is important for unraveling reaction mechanisms and for the calculation of reaction rates.⁵

The origin of the concept of reaction path is found in studies of simple chemical reactions, where a reaction is a transition from one potential energy minimum (reactant) to another (product). Such a reaction is usually considered to traverse across a saddle point that minimizes the potential energy barrier between the two minima, and the reaction path is constructed by locating the saddle point and then following the steepest descent of the potential surface from the saddle point. However, as researchers begin to study more complex reactions, the validity and practicality of the steepest-descent path is being called into question.³

In most cases of interest, reactions take place at finite temperature and therefore are stochastic. Every reaction event follows a different path and takes a different amount of time. Among all possible paths from the reactant to the product, we seek the path that is representative of all reaction events and thereby characterizes the reaction.⁶ In this regard, the steepest-descent path has the important drawback of not including the effect of temperature. Since reactions are driven by thermal fluctuations, temperature dependence of reaction paths should be taken into account. For example, suppose there is a direct path with high energy barriers and a roundabout path with low energy barriers. At high tempera-

ture (compared to the barriers) reaction events will occur most likely along the direct path, while the steepest-descent path will be the roundabout path regardless of temperature.

There have been efforts to find a better formulation of reaction path, like the maximum-flux path^{7,8} and the most probable path.⁹ These methods succeeded to a certain extent in elucidating reaction mechanisms, but they do not fully satisfy the criterion of representativity. The method of most probable path comes closest to satisfying the criterion. In the method, an ensemble of reaction events of a fixed time interval is considered. A probability is then assigned to each event, and the path followed by the most probable event is taken as the reaction path.¹⁰ But, it is not clear how to choose the time interval beforehand and whether an ensemble of reaction events of a single time interval suffices to represent the reaction.

This paper presents a new formulation of reaction path. While previous approaches attempted to quantify *paths*, we use the concept of reaction coordinate which quantifies *states*. Reaction coordinate is a function of states that describes where in the progress of a reaction a state is located. A natural measure of the progress of reaction is provided by the mean first-passage time (MFPT) $\tau(\mathbf{x})$ from state \mathbf{x} to the given product; the shorter the MFPT $\tau(\mathbf{x})$ is, the closer the state \mathbf{x} is to the product. The MFPT depends on the energy landscape, the temperature, as well as the boundary conditions, and most important, it is an average over all reaction events.^{11,12} Once the MFPT $\tau(\mathbf{x})$ is determined for all states, reaction paths can be constructed by following the direction along which τ decreases most rapidly.

II. REACTION PATHS BASED ON MEAN FIRST-PASSAGE TIMES

The following describes the calculation of MFPTs and the construction of reaction paths for two different settings:

^{a)}Electronic mail: kshulte@ks.uiuc.edu

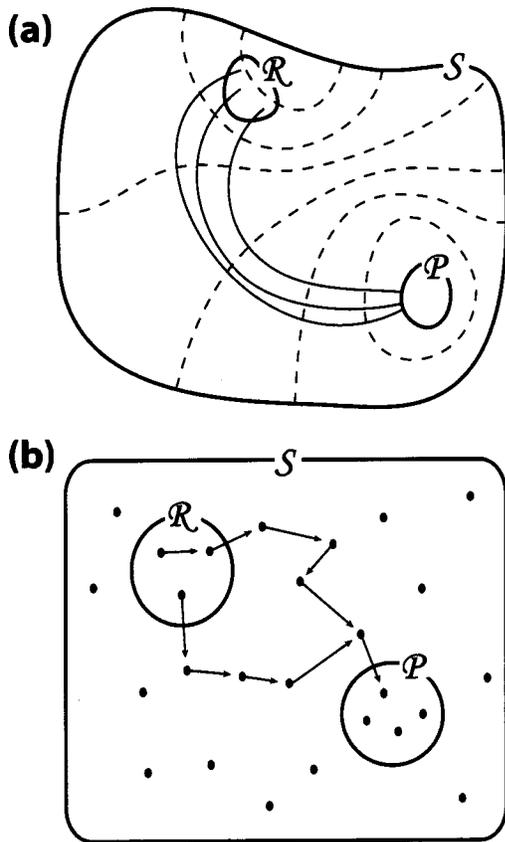


FIG. 1. Schematic illustration of reaction paths. S is the set of all accessible states, \mathcal{R} the reactant, and \mathcal{P} the product. (a) A continuous system. Dashed lines are contours of the reaction coordinate (the MFPT to the product \mathcal{P}) and the solid lines connecting \mathcal{R} and \mathcal{P} are reaction paths. See Sec. II A. (b) A discrete system. Dots are accessible states and the arrows denote reaction paths. See Sec. II B.

diffusion on a potential surface and discrete kinetics. Figure 1 shows a schematic illustration of reaction path in these two settings. The equation for the MFPT [Eq. (15)], whose derivation we outline in the following, was first derived in Refs. 13 and 14 and a more detailed account can be found in Ref. 11.

A. Diffusion on a potential surface

We first consider a reaction described as diffusion on a potential surface $U(\mathbf{x})$ defined on a continuous space S . The reactant and the product are specified by disjoint regions, \mathcal{R} and \mathcal{P} , respectively, in S . Let $P(\mathbf{y}, \mathbf{x}; t)$ and $\mathbf{J}(\mathbf{y}, \mathbf{x}; t)$ be, respectively, the probability density and the probability current at \mathbf{y} of the diffusing particle, starting at \mathbf{x} , after time interval t . The probability density obeys the Smoluchowski equation¹⁵

$$\partial_t P(\mathbf{y}, \mathbf{x}; t) = D \nabla_{\mathbf{y}} \cdot \{ e^{-\beta U(\mathbf{y})} \nabla_{\mathbf{y}} [e^{\beta U(\mathbf{y})} P(\mathbf{y}, \mathbf{x}; t)] \}, \quad (1)$$

where D is the diffusion coefficient and β is the inverse temperature. The differential operator $\nabla_{\mathbf{y}}$ acts on functions of \mathbf{y} . From comparing the continuity equation

$$\partial_t P(\mathbf{y}, \mathbf{x}; t) = -\nabla_{\mathbf{y}} \cdot \mathbf{J}(\mathbf{y}, \mathbf{x}; t) \quad (2)$$

and the Smoluchowski equation [Eq. (1)], an expression for the probability current follows:

$$\mathbf{J}(\mathbf{y}, \mathbf{x}; t) = -D e^{-\beta U(\mathbf{y})} \nabla_{\mathbf{y}} [e^{\beta U(\mathbf{y})} P(\mathbf{y}, \mathbf{x}; t)]. \quad (3)$$

The Smoluchowski equation must be accompanied by initial and boundary conditions. The initial condition is $P(\mathbf{y}, \mathbf{x}; 0) = \delta(\mathbf{y} - \mathbf{x})$. Since we are only interested in first passages to the product region \mathcal{P} , it is convenient to consider the diffusion process on the region $S \setminus \mathcal{P}$, the complement of \mathcal{P} . Therefore, there are two boundaries to take into account: the boundary of S and that of \mathcal{P} [Fig. 1(a)].¹⁶ Let us denote these boundaries by ∂S and $\partial \mathcal{P}$, respectively. Since only first passages to the product region \mathcal{P} need to be considered, it is assumed that the diffusing particle is absorbed on $\partial \mathcal{P}$:

$$P(\mathbf{y}, \mathbf{x}; t) = 0, \quad \text{for } \mathbf{y} \in \partial \mathcal{P}. \quad (4)$$

For ∂S , we assume the reflecting boundary condition for simplicity's sake:¹⁷

$$\mathbf{J}(\mathbf{y}, \mathbf{x}; t) \parallel \partial S \quad \text{for } \mathbf{y} \in \partial S, \quad (5)$$

meaning that the current \mathbf{J} is tangential to the boundary ∂S .

From the Smoluchowski equation [Eq. (1)] we derive the backward Smoluchowski equation which is used for calculating the MFPT. Since diffusion is a Markov process, the probability density $P(\mathbf{y}, \mathbf{x}; t)$ can be written as

$$P(\mathbf{y}, \mathbf{x}; t) = \int_{S \setminus \mathcal{P}} d\mathbf{z} P(\mathbf{y}, \mathbf{z}; s) P(\mathbf{z}, \mathbf{x}; t - s) \quad (6)$$

for an arbitrary intermediate time $s (0 < s < t)$. Taking the derivative with respect to t leads to

$$\begin{aligned} \partial_t P(\mathbf{y}, \mathbf{x}; t) &= \int_{S \setminus \mathcal{P}} d\mathbf{z} P(\mathbf{y}, \mathbf{z}; s) \partial_t P(\mathbf{z}, \mathbf{x}; t - s) \\ &= \int_{S \setminus \mathcal{P}} d\mathbf{z} P(\mathbf{y}, \mathbf{z}; s) D \nabla_{\mathbf{z}} \\ &\quad \cdot \{ e^{-\beta U(\mathbf{z})} \nabla_{\mathbf{z}} [e^{\beta U(\mathbf{z})} P(\mathbf{z}, \mathbf{x}; t - s)] \}. \end{aligned} \quad (7)$$

After carrying out the integration by parts twice, one obtains

$$\begin{aligned} \partial_t P(\mathbf{y}, \mathbf{x}; t) &= D \int_{S \setminus \mathcal{P}} d\mathbf{z} e^{\beta U(\mathbf{z})} P(\mathbf{z}, \mathbf{x}; t - s) \nabla_{\mathbf{z}} \\ &\quad \cdot [e^{-\beta U(\mathbf{z})} \nabla_{\mathbf{z}} P(\mathbf{y}, \mathbf{z}; s)] \end{aligned} \quad (8)$$

with the adjoint boundary conditions

$$P(\mathbf{y}, \mathbf{x}; t) = 0 \quad \text{for } \mathbf{x} \in \partial \mathcal{P}, \quad (9)$$

$$\nabla_{\mathbf{x}} P(\mathbf{y}, \mathbf{x}; t) \parallel \partial S \quad \text{for } \mathbf{x} \in \partial S. \quad (10)$$

The limit $s \rightarrow t$ leads to the backward Smoluchowski equation,

$$\partial_t P(\mathbf{y}, \mathbf{x}; t) = D e^{\beta U(\mathbf{x})} \nabla_{\mathbf{x}} \cdot [e^{-\beta U(\mathbf{x})} \nabla_{\mathbf{x}} P(\mathbf{y}, \mathbf{x}; t)], \quad (11)$$

where we have used $\lim_{s \rightarrow t} P(\mathbf{z}, \mathbf{x}; t - s) = \delta(\mathbf{z} - \mathbf{x})$.

The probability that the diffusing particle reaches the product region \mathcal{P} between time t and $t + dt$ is equal to the surface integral $-dt \int_{\partial \mathcal{P}} d\mathbf{s}_{\mathbf{y}} \cdot \mathbf{J}(\mathbf{y}, \mathbf{x}; t)$ with respect to \mathbf{y} over the boundary of \mathcal{P} . The minus sign is due to the surface normal $\mathbf{s}_{\mathbf{y}}$ being defined to be pointing outward from the region \mathcal{P} . Thus, the MFPT $\tau(\mathbf{x})$ is given as

$$\begin{aligned} \tau(\mathbf{x}) = & - \int_0^\infty dt \int_{\partial\mathcal{P}} d\mathbf{s}_y \cdot \mathbf{J}(\mathbf{y}, \mathbf{x}; t) \\ & + \int_0^\infty dt \int_{\partial\mathcal{S}} d\mathbf{s}_y \cdot \mathbf{J}(\mathbf{y}, \mathbf{x}; t), \end{aligned} \quad (12)$$

where the second term can be added because the surface integral vanishes on the reflecting boundary $\partial\mathcal{S}$ [Eq. (5)]. By Gauss's theorem, the surface integrals are transformed to a volume integral:

$$\tau(\mathbf{x}) = \int_0^\infty dt \int_{\mathcal{S}\setminus\mathcal{P}} d\mathbf{y} \nabla_y \cdot \mathbf{J}(\mathbf{y}, \mathbf{x}; t). \quad (13)$$

We replace $\nabla_y \cdot \mathbf{J}(\mathbf{y}, \mathbf{x}; t)$ with $-\partial_t P(\mathbf{y}, \mathbf{x}; t)$ by using the continuity equation [Eq. (2)], integrate by parts (with respect to t), and obtain

$$\tau(\mathbf{x}) = \int_0^\infty dt \int_{\mathcal{S}\setminus\mathcal{P}} d\mathbf{y} P(\mathbf{y}, \mathbf{x}, t) \quad (14)$$

under the assumption that $\int_{\mathcal{S}\setminus\mathcal{P}} d\mathbf{y} P(\mathbf{y}, \mathbf{x}; t)$ decays more quickly than $1/t$. Finally, using the backward Smoluchowski equation [Eq. (11)] we find^{13,14}

$$D e^{\beta U(\mathbf{x})} \nabla \cdot [e^{-\beta U(\mathbf{x})} \nabla \tau(\mathbf{x})] = -1. \quad (15)$$

Therefore, the MFPT $\tau(\mathbf{x})$ can be determined by solving this inhomogeneous partial differential equation on the region $\mathcal{S}\setminus\mathcal{P}$. The boundary conditions for τ are extracted from the adjoint boundary conditions [Eqs. (9) and (10)] by using Eq. (14):

$$\tau(\mathbf{x}) = 0 \quad \text{for } \mathbf{x} \in \partial\mathcal{P}, \quad (16)$$

$$\nabla \tau(\mathbf{x}) \parallel \partial\mathcal{S} \quad \text{for } \mathbf{x} \in \partial\mathcal{S}. \quad (17)$$

Reaction paths are then constructed following the direction of $-\nabla\tau$, along which τ decreases most rapidly. Thus, a reaction path $\mathbf{x}(l)$, parametrized through the arclength l , satisfies

$$\frac{d\mathbf{x}}{dl} = - \frac{\nabla \tau}{|\nabla \tau|}. \quad (18)$$

Since any point in the reaction region \mathcal{R} can be a starting point of a reaction path, in general multiple reaction paths are obtained unless the reactant region is narrowed down to a single point. The resulting reaction paths are independent of the diffusion coefficient D because the latter affects only the overall scale of τ .

B. Discrete kinetics

Let us now consider a different setting, namely a reaction described by transitions between discrete states. Let \mathcal{S} be the set of all accessible states. The reactant \mathcal{R} and the product \mathcal{P} are given as disjoint subsets of \mathcal{S} . We denote by R_{ji} the transition rate from state i to j , and for later use in Sec. IV assume a probability loss rate λ_j for each state j . A discrete master equation can be written for the probability $P_{ji}(t)$ that the system, initially in state i , is in state j after time interval t :

$$\partial_t P_{ji}(t) = \sum_{k \neq j} R_{jk} P_{ki}(t) - \sum_{k \neq j} R_{kj} P_{ji}(t) - \lambda_j P_{ji}(t), \quad (19)$$

or in a matrix form,

$$\partial_t P_{ji}(t) = \sum_k K_{jk} P_{ki}(t). \quad (20)$$

The transition matrix K_{jk} is built as

$$K_{jk} = R_{jk} \quad \text{for } j \neq k \quad (21)$$

$$K_{jj} = - \sum_{l \neq j} R_{lj} - \lambda_j. \quad (22)$$

The master equation is accompanied by the initial condition, $P_{ji}(0) = \delta_{ji}$. If there is no probability loss ($\lambda_j = 0$ for all j), the total probability is conserved, namely $\partial_t \sum_j P_{ji}(t) = 0$.

Following a similar procedure as used for the diffusion process in the preceding section, we derive an equation for the MFPT τ_i from state i to \mathcal{P} . Since we are only interested in first passages to \mathcal{P} , it is convenient to consider the subsystem consisting of the states in $\mathcal{S}\setminus\mathcal{P}$. This subsystem is described by the master equation

$$\partial_t P_{\beta\alpha}(t) = \sum_\gamma K_{\beta\gamma} P_{\gamma\alpha}(t) \quad (23)$$

with the initial condition $P_{\beta\alpha}(0) = \delta_{\beta\alpha}$. We use Greek subscripts to indicate that the states in \mathcal{P} are excluded. That is, Greek subscripts are assigned only to the states in $\mathcal{S}\setminus\mathcal{P}$. The transition matrix $K_{\beta\gamma}$ is extracted from the original transition matrix K_{jk} by eliminating the rows and columns belonging to the states in \mathcal{P} . Even when the total probability is conserved in the original system, the total probability in the subsystem is not conserved. The passage to \mathcal{P} accounts for this probability decrease.

Using the Markov property,

$$P_{\beta\alpha}(t) = \sum_\gamma P_{\beta\gamma}(s) P_{\gamma\alpha}(t-s) \quad \text{for } 0 < s < t, \quad (24)$$

one obtains

$$\begin{aligned} \partial_t P_{\beta\alpha}(t) &= \sum_\gamma P_{\beta\gamma}(s) \partial_t P_{\gamma\alpha}(t-s) \\ &= \sum_\gamma P_{\beta\gamma}(s) \sum_\delta K_{\gamma\delta} P_{\delta\alpha}(t-s). \end{aligned} \quad (25)$$

Taking the limit $s \rightarrow t$ leads to the backward master equation

$$\partial_t P_{\beta\alpha}(t) = \sum_\gamma P_{\beta\gamma}(t) K_{\gamma\alpha}. \quad (26)$$

Let ξ_β be the transition rate from state β to any state in \mathcal{P} :

$$\xi_\beta = \sum_{j \in \mathcal{P}} K_{j\beta}. \quad (27)$$

The MFPT τ_α from state α to the product set \mathcal{P} is then given as

$$\tau_\alpha = \int_0^\infty dt t \sum_\beta \xi_\beta \mathcal{P}_{\beta\alpha}(t) / \phi_\alpha, \quad (28)$$

$$\phi_\alpha = \int_0^\infty dt \sum_\beta \xi_\beta P_{\beta\alpha}(t). \quad (29)$$

Here ϕ_α is the total probability that the passage to \mathcal{P} will eventually occur. Because of the loss rate λ_j in each state, ϕ_α is in general less than one and the denominator in Eq. (28) is necessary for normalization. Finally, by using the backward master equation [Eq. (26)] and integrating by parts, we find

$$\sum_\alpha \phi_\alpha K_{\alpha\gamma} = -\xi_\gamma, \quad (30)$$

$$\sum_\alpha \tau_\alpha \phi_\alpha K_{\alpha\gamma} = -\phi_\gamma \quad (31)$$

under the assumption that $\sum_\beta \xi_\beta P_{\beta\gamma}(t)$ decays more rapidly than $1/t$. The solution to this coupled equation is given in terms of the inverse matrix $K_{\alpha\gamma}^{-1}$ of the matrix $K_{\alpha\gamma}$:

$$\phi_\alpha = -\sum_\gamma \xi_\gamma K_{\gamma\alpha}^{-1}, \quad (32)$$

$$\tau_\alpha = -\sum_\gamma \phi_\gamma K_{\gamma\alpha}^{-1} / \phi_\alpha. \quad (33)$$

These are the equations that we use in Sec. IV.

In the case of no loss ($\lambda_j=0$ for all j), ξ_γ is equal to $-\sum_\beta K_{\beta\gamma}$ and the solution can be stated in a simpler form:

$$\phi_\alpha = 1, \quad (34)$$

$$\tau_\alpha = -\sum_\gamma K_{\gamma\alpha}^{-1}. \quad (35)$$

In fact, when there is no loss, there is only one sink (through the product \mathcal{P}) in the system. The MFPT then may be derived simply by¹⁸

$$\begin{aligned} \tau_\alpha &= -\int_0^\infty dt t \partial_t \sum_\beta P_{\beta\alpha}(t) \\ &= \int_0^\infty dt \sum_{\beta\gamma} (e^{Kt})_{\beta\gamma} P_{\gamma\alpha}(0) \\ &= -\sum_\beta K_{\beta\alpha}^{-1}. \end{aligned} \quad (36)$$

In order to determine reaction paths based on the MFPT τ_i ($\tau_i=0$ for $i \in \mathcal{P}$), we draw an analogy to the diffusion on a continuous space considered earlier, where reaction paths follow the direction of the steepest descent of the MFPT, $-\nabla\tau$. Let us interpret these reaction paths as sequences of jumps of infinitesimal distance ϵ . Suppose a reaction path goes through a point \mathbf{x} . The next point \mathbf{x}' the reaction path visits is chosen out of the points lying at the distance of ϵ from \mathbf{x} . The choice is made in such a way that the decrease in the MFPT per distance, $[\tau(\mathbf{x}) - \tau(\mathbf{x}')]/\epsilon$, is maximized, which is the same as following $-\nabla\tau$. For the reaction described by discrete kinetics, a reaction path must be a se-

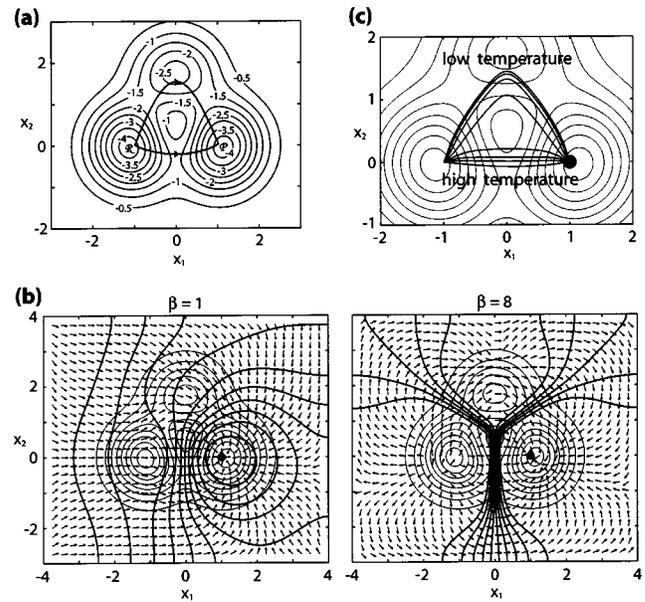


FIG. 2. Reaction paths of the diffusion on the three-hole potential. (a) A contour plot of the potential, with two channels connecting the reactant \mathcal{R} and the product \mathcal{P} . (b) The directions of $-\nabla\tau$ at selected grid points are plotted as arrows, for two different temperatures. Contours of the reaction coordinate τ (thick lines) and the potential (thin lines) are shown. (c) Temperature dependence of the reaction paths. Shown are eight reaction paths for eight different temperatures from bottom to top, $\beta=1, 2, 3, 4, 5, 6, 7, 8$. The reactant is the point $(-1, 0)$, and the product is the region indicated by the closed circle.

quence of transitions leading to the product. Suppose a reaction path goes through a state i . The next state j is chosen out of all the states to which a transition from state i can occur. According to the analogy with the continuous diffusion, the choice should be made such that the decrease in the MFPT per distance is maximized. However, in the case of discrete kinetics, no notion of distance is given *a priori*. Nevertheless, an analogue of distance is provided by the transition time $1/R_{ji}$ from state i to j , and one can choose the next state j that maximizes $R_{ji}(\tau_i - \tau_j)$. For the transition step from i to j , the transition time $1/R_{ji}$ may be interpreted as a cost, and the MFPT decrease $\tau_i - \tau_j$ as a gain. The scheme then amounts to maximizing for each step the ratio between these two times, namely the gain-cost ratio.

III. DIFFUSION ON THE THREE-HOLE POTENTIAL

We first apply the present method to the simplest non-trivial case, diffusion on a two-dimensional potential surface. For this purpose we choose the three-hole potential

$$\begin{aligned} U(x_1, x_2) &= 3e^{-x_1^2 - (x_2 - 1/3)^2} - 3e^{-x_1^2 - (x_2 - 5/3)^2} \\ &\quad - 5e^{-(x_1 - 1)^2 - x_2^2} - 5e^{-(x_1 + 1)^2 - x_2^2}, \end{aligned} \quad (37)$$

which was also studied by others regarding the temperature dependence of reaction paths.^{8,9} As can be seen from the contour plot in Fig. 2(a), the potential features two deep holes at $(-1, 0)$ and $(1, 0)$ and one shallow hole at $(0, 5/3)$. The two deep holes are considered as the reactant and the product. Roughly, there are two channels connecting the reactant and the product: the upper indirect channel via the

shallow hole and the lower direct channel. The upper channel is longer than the lower one but has lower energy barriers. It is therefore expected that the upper channel will be taken at low temperature and the lower channel at high temperature.

The MFPT $\tau(x_1, x_2)$ is calculated by solving Eq. (15) numerically. We take the region $(-4 \leq x_1 \leq 4, -3 \leq x_2 \leq 4)$ as the whole space \mathcal{S} , and assume that its boundary is reflecting [Eq. (17)]. For the reactant \mathcal{R} we take the point $(-1, 0)$, and for the product \mathcal{P} the small circular region of radius 0.1 centered at $(1, 0)$.¹⁹ The MFPT τ vanishes on the boundary of \mathcal{P} [Eq. (16)].

The numerical solution of Eq. (15) with these boundary conditions was obtained with MATLAB.²⁰ Figure 2(b) shows the solution for two different temperatures, $\beta=1$ and 8. The difference between the two temperatures is dramatic. At the high temperature ($\beta=1$) the arrows pointing in the direction of $-\nabla\tau$ flow more or less directly toward the product. At the low temperature ($\beta=8$), on the other hand, the flow is significantly distorted so that potential barriers are avoided, with a singular point produced around $(-1.5, -0.5)$. Also, the MFPT τ drops rapidly when barriers are crossed, as indicated by the contours of τ packed around saddle points. Figure 2(c) shows the reaction paths found at these two and other intermediate temperatures. As temperature is lowered the reaction path changes gradually from the lower channel to the upper channel, which agrees with the expected temperature dependence.

IV. EXCITATION MIGRATION IN PHOTOSYSTEM I

The task of determining representative pathways is also encountered in the study of the excitation migration in photosynthetic complexes, which can be described by discrete kinetics. In this section we apply the present method to the excitation migration in a photosynthetic complex, photosystem I (PSI).

Photosynthesis is carried out by pigment-protein complexes embedded in cell membranes.²¹ In such a complex, an aggregate of interacting pigments held in a fixed arrangement absorbs light and the resulting electronic excitation is used for separating charge across the cell membrane. The transmembrane potential induced by the charge separation is later used for the synthesis of ATP, the universal energy currency of a cell. However, in a typical light-harvesting complex most pigments do not participate directly in charge separation, instead they serve merely as light absorbing antenna and deliver their electronic excitation to a reaction center where charge separation takes place.

Among photosynthetic complexes, the photosynthetic unit of purple bacteria has been most extensively studied (for a review see Ref. 22). Recently, a high-resolution structure of PSI has been obtained from the cyanobacterium *Synechococcus elongatus*.²³ PSI along with photosystem II and their various satellite complexes constitutes the main machinery of oxygenic photosynthesis in plants, green algae, and cyanobacteria.²¹ PSI (Fig. 3) contains an aggregate of 96 chlorophylls as its primary light-harvesting pigments. The electronic excitation resulting from the absorption of a photon by a chlorophyll migrates to a special pair of chlorophylls (named P700 after their absorption peak in nm) in the

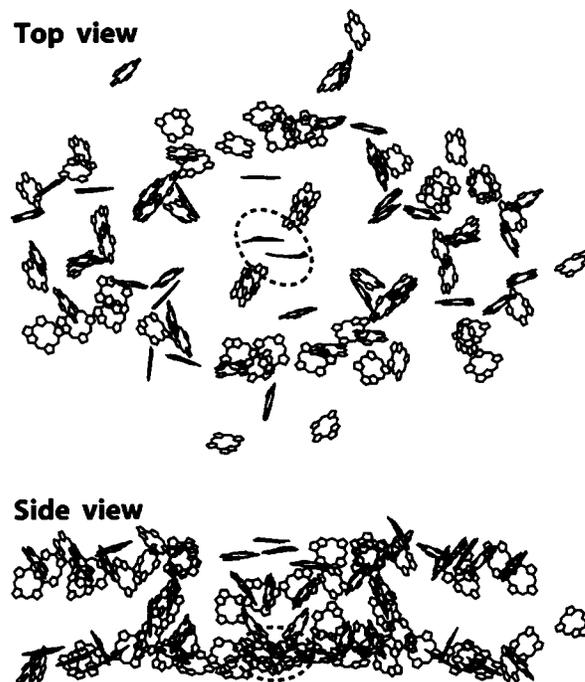


FIG. 3. Arrangement of the 96 chlorophylls in photosystem I (PSI), seen along the membrane normal (top view) and through the plane of the membrane (side view). The special pair of chlorophylls (P700) in the reaction center, at which charge separation is initiated, are marked. The figure made with VMD (Ref. 35).

center of the complex, where charge separation is initiated.²⁴ Figure 3 reveals that the chlorophylls of PSI are arranged without any apparent order (except for a pseudo- C_2 symmetry), which is in contrast to the highly symmetric circular arrangement of bacteriochlorophylls in the photosynthetic unit of purple bacteria.²² The apparently random arrangement of chlorophylls in PSI poses the question which pathways the excitation migration follows toward the reaction center.

The rates of interchlorophyll excitation transfer in PSI have been calculated based on Förster theory.^{25–27} Excitation follows stochastic trajectories along the excitation transfer network set by these rates. However, the obscure pattern of the excitation transfer network in PSI (cf. Fig. 6 in Ref. 25) instills a need for a simpler and more distilled picture of excitation migration. Besides, the functional aspect of the complex, namely that the P700 special pair is the target of the excitation migration, is not integrated in the network picture solely dictated by the transfer rates. In fact, the process of excitation migration can be considered a reaction: the reactant is the state where a pigment molecule that initially absorbed a photon is electronically excited, and the product is the state where the excitation resides in the special pair. Thus, the present method of reaction paths can determine the pathways representative of all excitation migration events reaching the reaction center, thereby providing a functional picture of the kinetic flow of excitation.

A. Mean first-passage time and excitation migration pathway

Excitation migration in PSI is a stochastic process governed by the rates of interchlorophyll transfer, dissipation

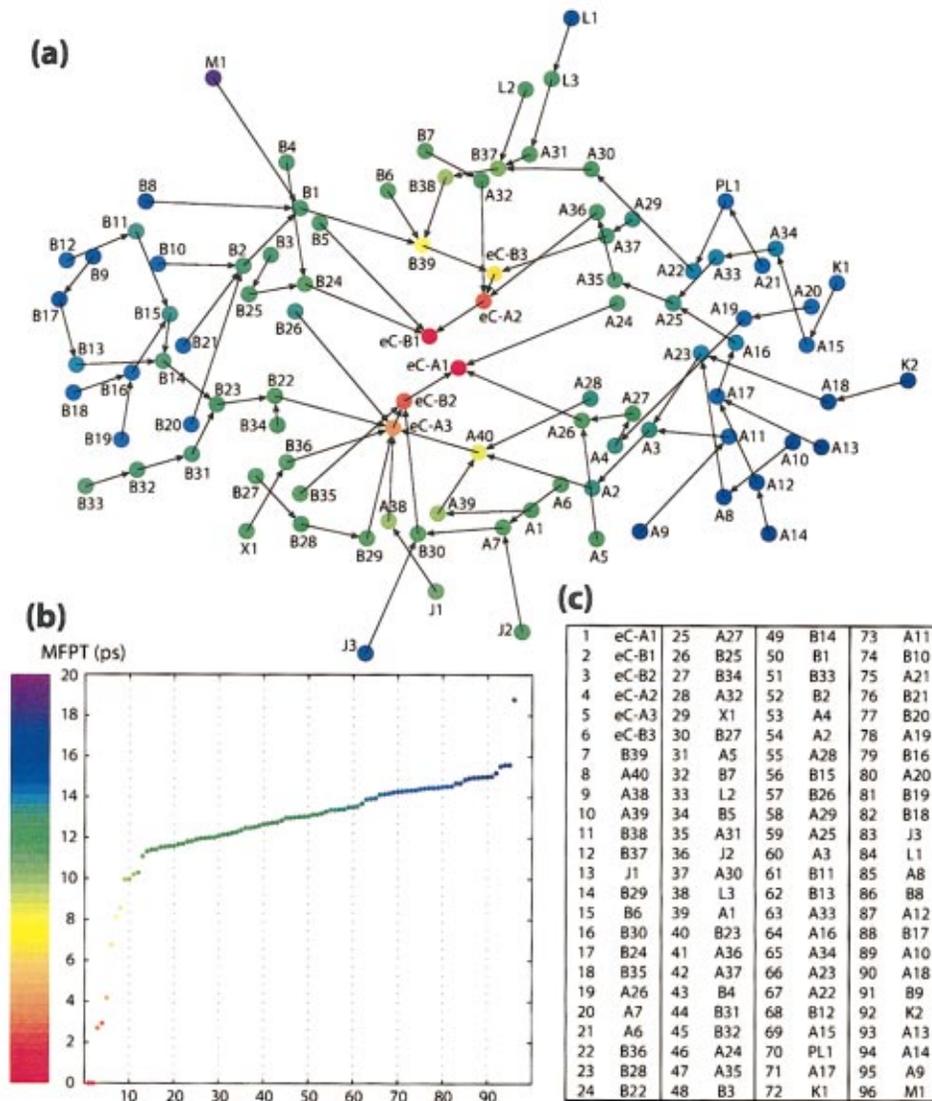


FIG. 4. (Color) MFPT and pathways of excitation migration in PSI. (a) The 96 chlorophylls, projected onto the membrane plane, are denoted by circles color-coded according to the MFPT to the special pair (eC-A1 and eC-B1). The excitation migration paths constructed based on the MFPT are shown as arrows. (b) The MFPTs are plotted in increasing order. The color-code scheme is the same as in (a). (c) List of chlorophylls sorted in order of increasing MFPT. The nomenclature (eC-A1, etc.) follows Ref. 23.

(conversion of excitation to heat), and the charge separation at the reaction center, and thus can be studied with master equations.^{26,28} The process can be described in terms of the probability $P_{ji}(t)$ that an excitation, initiated by light absorption at chlorophyll i , resides at chlorophyll j after a time interval t . The interchlorophyll transfer rates R_{ji} from chlorophyll i to chlorophyll j are calculated as explained in Ref. 26. For this calculation, we use the recently obtained site energies for chlorophylls²⁷ and the interchlorophyll electronic couplings determined by the full Coulomb computation that includes all multipole contributions to the coupling.²⁶ The dissipation rate is assumed to be the same ($1/\lambda_i = 1$ ns) at all the chlorophylls. Since we consider only first passages to the special pair,²⁹ the charge separation rate is not needed in the present model. Collecting these rates, we build the 96×96 transition matrix K_{ji} for the entire system [Eqs. (21) and (22)], and then extract the 94×94 transition matrix $K_{\beta\alpha}$ for the subsystem (corresponding to all the chlorophylls except the P700) by eliminating the rows and columns corresponding to the two P700 chlorophylls. The

MFPT τ_i (from chlorophyll i to the special pair) and the corresponding pathways then follow from the method described in Sec. II B.

B. Results and discussion

The obtained MFPT and paths are shown in Fig. 4. The 96 chlorophylls are sorted in the order of increasing MFPT and are listed in Fig. 4(c). The excitation migration paths shown in Fig. 4(a) exhibit a network without any apparent order, as expected from the disordered arrangement of the chlorophylls. There are various paths, including six direct paths to the special pair (from chlorophylls eC-A2, eC-B2, A24, A26, B5, and B24) and the most complicated path composed of nine steps (A21 \rightarrow PL1 \rightarrow A22 \rightarrow A30 \rightarrow B37 \rightarrow B38 \rightarrow B39 \rightarrow eC-B3 \rightarrow eC-A2 \rightarrow eC-B1). We emphasize that excitation does not necessarily follow these paths; they should rather be understood as representative paths. Overall, the pathways are more or less evenly distributed over the entire

complex, which is in accord with the robustness of the complex against removal of individual chlorophylls as discussed in Ref. 26.

As can be seen from Fig. 4(b), most of the chlorophylls (9th to 96th) have MFPTs around or above 10 ps, with the average over all being 12.5 ps. Chlorophylls A40 and B39 lie between these peripheral chlorophylls and the six central chlorophylls (eC-A1, eC-B1, eC-A2, eC-B2, eC-A3, and eC-B3). This supports to a certain extent the suggestion that chlorophylls A40 and B39 play the role of a bridge connecting the central and the peripheral chlorophylls.²³ However, among the found paths to the special pair, many go through neither chlorophyll A40 nor B39. Therefore, these two chlorophylls are not the only connection between the central chlorophylls and the periphery, and hence should not be considered as bottlenecks.

The last chlorophyll, M1, raises a question as it appears to be rather isolated from the rest [Fig. 4(b)]. This isolation does not imply that excitation will be trapped at this chlorophyll; the MFPT associated with this chlorophyll (18.8 ps) is still much shorter than the dissipation time of 1 ns. However, it seems inappropriate that the one chlorophyll is located relatively farther from the rest. The answer to this apparent puzzle lies in the observation that PSI exists at times as a trimeric structure *in vivo*.^{23,30} The chlorophyll M1 lies close to the boundary between monomers. In fact, we find that chlorophyll M1 is coupled to chlorophyll A30 in the next monomer with a coupling of 52.9 cm^{-1} , which is much stronger than the strongest coupling it has within its own monomer (6.7 cm^{-1} with B8). Hence, the chlorophyll M1 functionally belongs to the next monomer, not its own monomer.

V. CONCLUSION

We have presented a new method to construct reaction paths based on MFPTs that incorporates all reaction events, and illustrated how it captures important aspects of reactions, most notably temperature effects. We believe that MFPT is a natural choice for a reaction coordinate, and expect that our approach will give an insight into the study of reactions. The present method is particularly suitable for large reaction networks that are completely characterized by the method through a pathway graph visualizing the kinetic flow of the reaction.

As an application of the method, we have found representative paths of excitation migration in PSI. The MFPT and the paths provide a complete yet distilled picture of the excitation migration toward the reaction center, thereby characterizing the function of the system. We expect that our methodology will be useful for various photosynthetic complexes as more of their high-resolution structures become available.

ACKNOWLEDGMENTS

The authors are grateful to Paul Grayson for valuable discussions. This work has been supported by National Institutes of Health Grant No. PHS 5 P41 RR05969.

- ¹V. S. Pande, A. Y. Grosberg, T. Tanaka, and D. S. Rokhsar, *Curr. Opin. Struct. Biol.* **8**, 68 (1998).
- ²P. Eastman, N. Grøbech-Jensen, and S. Doniach, *J. Chem. Phys.* **114**, 3823 (2001).
- ³R. Elber, in *Recent Developments in Theoretical Studies of Proteins*, edited by R. Elber (World Scientific, Singapore, 1996).
- ⁴J. E. Straub, in *Computational Biochemistry and Biophysics*, edited by O. M. Becker, A. D. MacKerell, Jr., B. Roux, and M. Watanabe (Dekker, New York, 2001).
- ⁵P. Hänggi, P. Talkner, and M. Borkovec, *Rev. Mod. Phys.* **62**, 251 (1990).
- ⁶An alternative approach is to sample and analyze a multitude of reaction events (Refs. 31 and 32).
- ⁷M. Berkowitz, J. D. Morgan, J. A. McCammon, and S. H. Northrup, *J. Chem. Phys.* **79**, 5563 (1983).
- ⁸S. Huo and J. E. Straub, *J. Chem. Phys.* **107**, 5000 (1997).
- ⁹R. Elber and D. Shalloway, *J. Chem. Phys.* **112**, 5539 (2000).
- ¹⁰This amounts to minimizing an action integral, and similar methods have been used to solve boundary-value problems in classical mechanics (Refs. 33 and 34).
- ¹¹C. W. Gardiner, *Handbook of Stochastic Methods*, 2nd ed. (Springer, Berlin, 1985).
- ¹²S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, New York, 2001).
- ¹³D. L. Weaver, *Phys. Rev. B* **20**, 2558 (1979).
- ¹⁴A. Szabo, K. Schulten, and Z. Schulten, *J. Chem. Phys.* **72**, 4350 (1980).
- ¹⁵M. von Smoluchowski, *Ann. Phys. (Paris)* **48**, 1103 (1915).
- ¹⁶The location of the reactant region \mathcal{R} does not affect the calculation of the reaction coordinate $\tau(\mathbf{x})$; it is involved only in the construction of reaction path.
- ¹⁷Extension to more general boundary conditions is possible (Refs. 11 and 14).
- ¹⁸W. Nadler and K. Schulten, *J. Chem. Phys.* **82**, 151 (1985).
- ¹⁹This choice of the product region seems to be close enough to the limit of single-point product region. As the product region shrinks to zero volume, the MFPT τ will diverge but $-\nabla\tau/|\nabla\tau|$, which is what we need for constructing reaction paths, will still be well defined. Nevertheless, the product region must have finite volume in numerical calculations.
- ²⁰The MathWorks, Inc., *Partial Differential Equation Toolbox User's Guide* (2002).
- ²¹R. E. Blankenship, *Molecular Mechanisms of Photosynthesis* (Blackwell Science, Malden, 2002).
- ²²X. Hu, T. Ritz, A. Damjanović, F. Autenrieth, and K. Schulten, *Q. Rev. Biophys.* **35**, 1 (2002).
- ²³P. Jordan, P. Fromme, H. T. Witt, O. Klukas, W. Saenger, and N. Krauß, *Nature (London)* **411**, 909 (2001).
- ²⁴In the literature the term *reaction center chlorophylls* refers typically to the set of six chlorophylls in the center (including the pair P700) which are part of the electron transfer chain (Ref. 23).
- ²⁵M. Byrdin, P. Jordan, N. Krauß, P. Fromme, D. Stehlik, and E. Schlodder, *Biophys. J.* **83**, 433 (2002).
- ²⁶M. K. Şener, D. Lu, T. Ritz, S. Park, P. Fromme, and K. Schulten, *J. Phys. Chem. B* **106**, 7948 (2002).
- ²⁷A. Damjanović, H. M. Vaswani, P. Fromme, and G. R. Fleming, *J. Phys. Chem. B* **106**, 10251 (2002).
- ²⁸T. Ritz, S. Park, and K. Schulten, *J. Phys. Chem. B* **105**, 8259 (2001).
- ²⁹The first passage corresponds to the first term in the so-called sojourn expansion whose higher-order terms cover subsequent escapes from and returns to the reaction center (Ref. 26).
- ³⁰E. J. Boekema, J. P. Dekker, M. G. van Heel, M. Rögner, W. Saenger, I. Witt, and H. T. Witt, *FEBS Lett.* **217**, 283 (1987).
- ³¹L. R. Pratt, *J. Chem. Phys.* **85**, 5045 (1986).
- ³²P. G. Bolhuis, C. Dellago, P. L. Geissler, and D. Chandler, *J. Phys.: Condens. Matter* **12**, A147 (2000).
- ³³R. Olender and R. Elber, *J. Chem. Phys.* **105**, 9299 (1996).
- ³⁴D. Passerone and M. Parrinello, *Phys. Rev. Lett.* **87**, 108302 (2001).
- ³⁵W. Humphrey, A. Dalke, and K. Schulten, *J. Mol. Graphics* **14**, 33 (1996).