Protein conformational transitions explored by mixed elastic network models[†]

Wenjun Zheng,[1] Bernard R. Brooks,[1] and Gerhard Hummer[2*]

[1] Laboratory of Computational Biology, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892

[2] Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892

ABSTRACT

We develop a mixed elastic network model (MENM) to study large-scale conformational transitions of proteins between two (or more) known structures. Elastic network potentials for the beginning and end states of a transition are combined, in effect, by adding their respective partition functions. The resulting effective MENM energy function smoothly interpolates between the original surfaces, and retains the beginning and end structures as local minima. Saddle points, transition paths, potentials of mean force, and partition functions can be found efficiently by largely analytic methods. To characterize the protein motions during a conformational transition, we follow "transition paths" on the MENM surface that connect the beginning and end structures and are invariant to parameterizations of the model and the mathematical form of the mixing scheme. As illustrations of the general formalism, we study large-scale conformation changes of the motor proteins KIF1A kinesin and myosin II. We generate possible transition paths for these two proteins that reveal details of their conformational motions. The MENM formalism is computationally efficient and generally applicable even for large protein systems that undergo highly collective structural changes.

INTRODUCTION

Dynamic transitions between different conformations are essential for the function of many proteins. Examples include allosteric transitions in enzymes,[1] force generations in motor proteins,[2] the opening and closing of ion channels, and the conformational changes induced by ligand binding to enzymes and receptors. A detailed molecular description of the structural changes during those transitions should facilitate our understanding of the underlying mechanisms and may aid in the design of ligands that modulate transition equilibria and rates. Direct experimental investigations of the dynamic transitions are practically limited to observations of sufficiently populated states, including the beginning and end states, and high-population intermediates. As a consequence, transition states with their low population are largely inaccessible to experiments, with the possible exception of perturbative approaches such as Φ-value analysis in protein folding.[3] Even single-molecule techniques, while able to detect molecular transitions,[4] do not yet have the spatial and temporal resolution to fully capture the structure and dynamics during transition events.

Faced with these experimental challenges, it appears highly desirable to develop general computational techniques that allow us to explore in atomic detail the structural changes occurring during molecular transitions. Molecular simulations appear ideally suited to study protein conformational dynamics. However, most transition processes of interest occur on time and length scales inaccessible to standard all-atom molecular dynamics simulations.[5]

Among several alternatives,[6–8] one promising approach to overcome the time scales limitations of simulations is to use coarse-grained models with simplified representations and energy functions.[9] Structure-based elastic network models (ENMs)[10] have been successfully applied to describe protein conformational changes.[11–21] In ENMs, amino acids are represented as beads connected by harmonic springs. Low-frequency normal modes then yield directions of collective large-amplitude motions.

By construction, the ENM potential function has only a single minimum corresponding to a known protein structure. To explore conformational

transitions, information about the modes connecting the beginning and end structure can be extracted by projecting the more compact state onto the normal modes of the more open state. Going beyond this prescription, Kim et al.[22,23] have used a distance-matrix based interpolation to generate intermediate structures between two end structures. Another study by Miyashita et al.[24] used an ENM-based iterative search to generate plausible transition paths with minimal structural distortions and energy cost.

Recently, Maragakis and Karplus[25] and Best et al.[26] independently proposed two closely related formalisms to interpolate between two structure-based potentials. In both formalisms, the structure-based potentials of the beginning and end states of a molecular transition are combined to obtain effective energy surfaces that retain the initial and final structures of the transition as dominant and smoothly connected local minima. In the "plastic network model" (PNM) of Maragakis and Karplus,[25] two ENM potentials $E_1(x)$ and $E_2(x)$, are combined following the valence bond formulation of quantum chemistry, $E(x) = [E_1(x) + E_2(x) + \varepsilon_1 + \varepsilon_2 - \{[E_1(x) - E_2(x) + \varepsilon_1 - \varepsilon_2]^2 + \beta^{-2}\}^{1/2}]/2$. Best et al.[26] mixed two Gō potentials using Boltzmann weights of statistical mechanics to describe the α-to-β transition of an arc repressor mutant: $\exp[-\beta E(x)] = \exp\{-\beta[E_1(x) + \varepsilon_1]\} + \exp\{-\beta[E_2(x) + \varepsilon_2]\}$, with energy offsets ε_1 and ε_2, and a mixing parameter β. The PNM mixing scheme[25] was recently explored in combination with a Gō model by Okazaki et al.[27]

In this work we extend the formalism of Best et al.[26] to mixing of ENMs for known structures of two protein conformational states. The resulting mixed-ENM (MENM) energy function has two minima connected by a saddle point (SP). For the harmonic ENM potentials, the Boltzmann-weighted mixing allows us to evaluate free energies, potentials of mean force (PMF), and even transition paths using largely analytic methods because the partition functions retain their Gaussian character.

We will demonstrate the use of the MENM method by examining two biologically important conformational transitions in two motor proteins (myosin II and KIF1A kinesin), which were analyzed by conventional ENM methods previously.[13,28,29] First, we will study the power-stroke transition of myosin II from the prepowerstroke state (PDB code: 1VOM[30]) to the postpowerstroke rigor-like state (PDB code: 1Q5G[31]). This transition takes place after ATP hydrolysis in response to actin binding, resulting in the converter/lever arm undergoing a large-scale rotation that is accompanied by the closure of the actin-binding cleft within the 50 kDa domain.[31] Second, we study a related force-generating transition in KIF1A kinesin from the ADP-bound state (PDB code: 1I5S[32]) to the ATP-like state (PDB code: 1I6I[32]). In this transition, the neck linker (analogous to the lever arm in myosin) undergoes a disorder-to-order transition to dock against the motor domain in response to the ATP-binding at the nucleotide-binding site.[32]

The general procedure is summarized as follows. First, we construct the MENM potential based on the two ENM potentials built from the known beginning and end structures of the transition. We then determine the SP and three alternative "transition paths:" the steepest-descent (SD) path; the path traced by the SP as the mixing parameter β is continuously varied; and a minimum free energy path defined in terms of 1D and 2D PMFs. As it turns out, the SP transition path is independent of the parameters β and ε_i of MENM surface and the specific model used for mixing. From the parameter-independent SP path, we extract information about the time-ordering of the forming and breaking of residue contacts during the transitions. Even though the main focus of the paper is on illustrating the MENM method, we briefly discuss the biological implications of the results.

METHODS

Potential function of ENM

In ENMs, proteins are represented as elastic bodies. The ENM potentials are constructed by using the C_α coordinates of a protein in its native structure. A harmonic potential with a uniform force constant C accounts for pair-wise interactions between all C_α atoms that are within a cutoff distance R_C. The potential energy in the elastic network representation of a protein is

$$E(\vec{x} - \vec{x}_0) = \frac{1}{2} \sum_{d_{ij}^0 < R_C} C(d_{ij} - d_{ij}^0)^2, \tag{1}$$

where $\vec{x}$ is a $3N$-dimensional vector representing the Cartesian coordinates of the N C_α atoms, $\vec{x}_0$ is the corresponding vector of the C_α positions in the native (crystal or NMR) structure, d_{ij} is the Euclidian distance between the C_α atoms i and j, and d_{ij}^0 is the corresponding distance in the native structure.

We expand the potential function in Eq. (1) to second order about the minimum by computing its Hessian matrix H:

$$E(\delta\vec{x}) \approx \frac{1}{2}\delta\vec{x}^T \cdot H \cdot \delta\vec{x}, \tag{2}$$

where $\delta\vec{x} = \vec{x} - \vec{x}_0$. Despite the drastic simplification of representing the complex protein structure by an effective harmonic potential, the resulting model has led to useful descriptions of large-amplitude protein motions in terms of low-frequency normal modes of the Hessian H.[11–21,33]

Fitting *B* factors to calibrate the force constant *C*

In its simplest form, the ENM has two parameters: the cutoff distance R_C and the force constant C of the harmonic springs. We calibrate C (in units of $k_B T$) by fitting the isotropic crystallographic B factor B_i of residue i in a given crystal structure as follows:

$$\frac{B_i}{8\pi^2} = \langle u_i^2 \rangle_{\text{isotropic}} = \frac{k_B T_{\text{crystal}}}{3} \sum_m \frac{\vec{v}_{m,i}^2}{\lambda_m}, \quad (3)$$

where k_B is Boltzmann constant, $\vec{v}_{m,i}$ is the 3D component of the eigenvector of mode m for residue i, λ_m is the eigenvalue of mode m, and T_{crystal} is the temperature at which the structure was determined (here $T_{\text{crystal}} = 100$ K). Therefore, a conversion is needed from T_{crystal} to $T = 300$ K in the fitting procedure.

A fit of the ENM spring constant C produces $C = 1.7 \, k_B T$ for the kinesin PDB structures 1I5S and 1I6I (with the fitting crosscorrelation coefficient = 0.5), and 2.8 $k_B T$ for the myosin PDB structures 1VOM and 1Q5G (crosscorrelation coefficient = 0.6). In the following, all energies will be expressed in units of $k_B T$, where $T = 300$ K.

Mixed potential function constructed from two ENMs

Given two protein structures $\vec{x}_1$ and $\vec{x}_2$ corresponding to the beginning and end states of a conformational transition, respectively, we construct two ENMs $E_1(\vec{x} - \vec{x}_1)$ and $E_2(\vec{x} - \vec{x}_2)$ with Hessians H_1 and H_2 according to Eqs. (1) and (2). Following Best *et al.*,[26] we then define the MENM potential function:

$$E(\vec{x}) = -\beta^{-1} \, \ln[e^{-\beta(E_1(\vec{x}-\vec{x}_1)+\varepsilon_1)} + e^{-\beta(E_2(\vec{x}-\vec{x}_2)+\varepsilon_2)}], \quad (4)$$

where $\varepsilon_1, \varepsilon_2$ are energy offsets and $\beta = 1/k_B T_m$ is the inverse of the mixing temperature T_m that determines the extent of mixing between the two ENMs. In the limit $T_m \rightarrow +0$, the MENM potential $E(\vec{x})$ is the minimum $\min\{E_1 + \varepsilon_1, E_2 + \varepsilon_2\}$ of $E_1 + \varepsilon_1$ and $E_2 + \varepsilon_2$; for $T_m \rightarrow \infty$, the mixed potential approaches the sum $E(\vec{x}) \rightarrow E_1(\vec{x}) + E_2(\vec{x}) + \varepsilon_1 + \varepsilon_2$. For small but finite T_m, $E(\vec{x})$ is approximately $E_1(\vec{x}) + \varepsilon_1$ near $\vec{x}_1$ and $E_2(\vec{x}) + \varepsilon_2$ near $\vec{x}_2$, with a smooth interpolation between E_1 and E_2 elsewhere in configuration space. The MENM energy surface $E(\vec{x})$ has two dominant local minima corresponding to the beginning and end structures of the transition, and connects them with a SP, whose height can be tuned by the mixing temperature T_m. $E(\vec{x})$ thus seems well suited to explore the transition between two protein conformations that represent different functional states. Equation (4) can be easily generalized to mixing of more than two ENMs.

The mixing scheme of Eq. (4) was first applied to Gō potentials.[26] A related method, based on a quantum-mechanical mixing of two ENM energy surfaces, was proposed by Maragakis and Karplus as discussed in the Intro-duction.[25] Here, with the statistical–mechanical mixing of two free energy surfaces, we can take advantage of the largely retained Gaussian character of the MENM theory and find analytical expressions for free energies, PMFs, SPs, and transition paths. In particular, the partition function of the MENM system at temperature $T = 300$ K is given by

$$Z = \int e^{-\beta_T E(\vec{x})} \, d\vec{x}$$
$$= \int \left[e^{-\beta(E_1(\vec{x}-\vec{x}_1)+\varepsilon_1)} + e^{-\beta(E_2(\vec{x}-\vec{x}_2)+\varepsilon_2)} \right]^{N_T} d\vec{x}, \quad (5)$$

where

$$\beta_T = \frac{1}{k_B T}, \quad N_T = \frac{\beta_T}{\beta} \quad (6)$$

are the inverse sampling temperature, and the ratio of mixing and sampling temperatures, respectively. In general, $N_T \neq 1$ because the mixing temperature can be much higher than $T = 300$ K to achieve a reasonably low energy barrier (see Results). Nevertheless, for integer temperature ratios N_T this partition function can be computed analytically from Gaussian integrals obtained after a binomial expansion. We note that, in principle, thermodynamic quantities such as entropies at integer N_T could be calculated by using the binomial theorem and breaking up the configuration-space integral into regions where one or the other Boltzmann factor dominates in Eq. (4). More practically, to calculate entropies for strong mixing (small T/T_m) beyond a simple two-state approximation, one can use the analytical expressions for the free energies and partition functions for integer $N_T = T_m/T$ and estimate the entropy from the temperature derivative of an interpolation.

Parameterization of MENM

The MENM potential has two parameters: T_m (mixing temperature) and $\varepsilon_2 - \varepsilon_1$ (difference in energy offsets). The energy difference $\varepsilon_2 - \varepsilon_1$ is determined by the relative populations of the two end states at equilibrium, as obtained from experiment. Within the harmonic approximation to the partition functions of the two states, the ratio of populations in States 1 and 2 is given by

$$P_1/P_2 = e^{-\beta_T(\varepsilon_1-\varepsilon_2)} [\det(H_2)/\det(H_1)]^{1/2}. \quad (7)$$

For simplicity or without population information available experimentally, we set $\varepsilon_2 = \varepsilon_1$.

A bound on the mixing temperature T_m can, in principle, be obtained from the experimental rate of transitions:

$$\text{rate} \approx \omega_0 \, e^{-\frac{\Delta G}{k_B T}} \leq 1 \, (\text{ps}^{-1}) \times e^{\frac{E_{\text{sp}}}{k_B T}} \quad (8)$$

Here the prefactor ω_0 is assumed to have an upper bound of 1 ps^{-1} and the free energy barrier ΔG is

approximated by $E_{\rm sp}$ (the potential energy at the SP). Clearly, Eq. (4) is a very coarse approximation, with the implicit assumption that large-amplitude equilibrium fluctuations obey a simple dynamics and are realistically described by the combined energy surface. One can thus expect only a rough estimate of the barrier height.

For convenience, we here do not use rate information; instead, to demonstrate the effects of large versus small $T_{\rm m}$, we discuss two extreme cases: weak mixing ($T_{\rm m} = T$) and strong mixing ($T_{\rm m} = T_{\rm strong}$). $T_{\rm strong}$ is defined as the maximal $T_{\rm m}$ at which both beginning and end-state structures remain approximately local minima (or stationary points) of the mixed potential (see next subsection). In practice, $T_{\rm strong}$ is determined by varying the mixing temperature $T_{\rm m}$ in discrete steps until the double-well structure disappears at high $T_{\rm m}$. For kinesin we use $T_{\rm strong} = 810\ T$ as the "strong-mixing" temperature, and for myosin we use $T_{\rm strong} = 2160\ T$ (where $T = 300$ K).

SPs and minima of mixed energy surface

We can locate the energy minima and SP for the MENM system by setting the gradient of the mixed potential function to zero. From Eq. (4), we have

$$\nabla E(\vec{x}) = f_1(\vec{x})\nabla E_1(\vec{x} - \vec{x}_1) + f_2(\vec{x})\nabla E_2(\vec{x} - \vec{x}_2) = 0, \tag{9}$$

where

$$f_1(\vec{x}) = \frac{e^{-\beta(E_1(\vec{x}-\vec{x}_1)+\varepsilon_1)}}{e^{-\beta(E_1(\vec{x}-\vec{x}_1)+\varepsilon_1)} + e^{-\beta(E_2(\vec{x}-\vec{x}_2)+\varepsilon_2)}}, \tag{10}$$
$$f_2(\vec{x}) = 1 - f_1(\vec{x}).$$

With $\nabla E_1(\vec{x} - \vec{x}_1) = H_1(\vec{x} - \vec{x}_1)$ and $\nabla E_2(\vec{x} - \vec{x}_2) = H_2(\vec{x} - \vec{x}_2)$, we find that the SP (and minima) can be expressed as

$$\vec{x}_{\rm sp} = \left[f_1(\vec{x}_{\rm sp})H_1 + (1 - f_1(\vec{x}_{\rm sp}))H_2\right]^{-1}$$
$$\times \left[f_1(\vec{x}_{\rm sp})H_1\vec{x}_1 + (1 - f_1(\vec{x}_{\rm sp}))H_2\vec{x}_2\right] \tag{11}$$

where the matrix inverse is restricted to the subspace spanned by the eigenvectors with nonzero eigenvalues. By combining Eqs. (10) and (11), we find that the SP can be found by solving a one-dimensional "fixed-point" equation for f_1

$$f_1(\vec{x}_{\rm sp}(f_1)) = f_1. \tag{12}$$

Equation (12) can be solved efficiently (for example, by using 1D bisection) without a full-scale minimization of the gradient in the high-dimensional conformation space. Its solutions include all SPs and local minima of the potential, among which the unstable fixed point(s) of Eq. (12) can be identified as SPs.

The earlier 1D reduction of the multidimensional SP equation can be applied to a general mixing function $g(E_1, E_2)$ defining an effective potential

$$E(\vec{x}) = g(E_1(\vec{x} - \vec{x}_1) + \varepsilon_1, \quad E_2(\vec{x} - \vec{x}_2) + \varepsilon_2) \tag{13}$$

In general, the stationary points satisfy

$$\nabla E(\vec{x}) = \frac{\partial g}{\partial E_1}\nabla E_1(\vec{x} - \vec{x}_1) + \frac{\partial g}{\partial E_2}\nabla E_2(\vec{x} - \vec{x}_2)$$
$$= f_1(\vec{x})\nabla E_1(\vec{x} - \vec{x}_1) + f_2(\vec{x})\nabla E_2(\vec{x} - \vec{x}_2) = 0 \tag{14}$$

where

$$f_1(\vec{x}) = \frac{\frac{\partial g}{\partial E_1}}{\frac{\partial g}{\partial E_1} + \frac{\partial g}{\partial E_2}}, \quad f_2(\vec{x}) = 1 - f_1(\vec{x}). \tag{15}$$

If $\frac{\partial g}{\partial E_1} \cdot \frac{\partial g}{\partial E_2} > 0$, then $f_1 \in [0, 1]$; so the 1D SP path $\vec{x}_{\rm sp}(f_1)$ solved from Eq. (11) is independent of the mathematical form of the mixing function $g(E_1, E_2)$. In particular, both the MENM mixing potential [Eq. (4)] and the Hessian approximation to the PNM formulation[25]: $E(x) = [E_1(x) + E_2(x) + \varepsilon_1 + \varepsilon_2 - \{[E_1(x) - E_2(x) + \varepsilon_1 - \varepsilon_2]^2 + \beta^{-2}\}^{1/2}]/2$ satisfy $\frac{\partial g}{\partial E_1} \cdot \frac{\partial g}{\partial E_2} > 0$ and so the SPs should fall along the same 1D SP path for both mixing schemes.

We determine the SPs and SP paths for two limiting cases: weak mixing ($T_{\rm m} = T$) and strong mixing ($T_{\rm m} = T_{\rm strong}$). $T_{\rm strong}$ is found by gradually increasing $T_{\rm m}$ in steps of $10T$ until the minima of $E(\vec{x}_{\rm sp}(f_1))$ at $f_1 = 0, 1$ are shifted away from 0 or 1 by $\delta f > 0.01$.

For the parameters we choose here, there is only one SP. However, at higher mixing temperatures, a situation can arise with three minima connected by two SPs. We note that the formalism described above can be generalized to mixing of more than two ENMs: for the mixed potential from K ENMs, one needs to solve a $K-1$ dimensional fixed-point equation.

1D and 2D potential of mean force

We define the following reaction coordinate $r(\vec{x})$ for a given direction $\vec{n}$:

$$r(\vec{x}) = (\vec{x} - \vec{x}_1) \cdot \vec{n}. \tag{16}$$

So

$$\begin{aligned} r(\vec{x}_1) &= 0, \\ r(\vec{x}_2) &= (\vec{x}_2 - \vec{x}_1) \cdot \vec{n} \equiv D_{12}. \end{aligned} \tag{17}$$

The PMF at temperature T is computed as the free energy for the subspace constrained by $r(\vec{x}) = r$:

$$F(r) = -\frac{1}{\beta_T}\log(Z(r)), \tag{18}$$

where

$$Z(r) = \int e^{-\beta_T E(\vec{x})} \delta(r(\vec{x}) - r) d\vec{x}$$
$$= \int \left[e^{-\beta(E_1(\vec{x}-\vec{x}_1)+\varepsilon_1)} + e^{-\beta(E_2(\vec{x}-\vec{x}_2)+\varepsilon_2)} \right]^{N_T} \delta(r(\vec{x}) - r) d\vec{x}$$

The average conformation at given r is

$$\langle \vec{x} \rangle_r = \frac{1}{Z(r)} \int e^{-\beta_T E(\vec{x})} \vec{x} \delta(r(\vec{x}) - r) d\vec{x} \qquad (20)$$

We further compute the 2D PMF by using two sampling directions, $\vec{n}_1 = \vec{n}_{sp}, \vec{n}_2 = \vec{x}_2 - \vec{x}_1$, that define two reaction coordinates r and r', respectively:

$$\begin{aligned} r &= r_{sp} = (\vec{x} - \vec{x}_1) \cdot \vec{n}_{sp} / [(\vec{x}_2 - \vec{x}_1) \cdot \vec{n}_{sp}], \\ r' &= r_{12} = (\vec{x} - \vec{x}_1) \cdot \vec{n}_2 / [(\vec{x}_2 - \vec{x}_1) \cdot \vec{n}_2]. \end{aligned} \qquad (21)$$

The 2D PMF is then

$$F(r, r') = -\frac{1}{\beta_T} \log(Z(r, r')) \qquad (22)$$

$$Z(r, r') = \int e^{-\beta_T E(\vec{x})} \delta(r(\vec{x}) - r) \cdot \delta(r'(\vec{x}) - r') d\vec{x}$$
$$= \int \left[e^{-\beta(E_1(\vec{x}-\vec{x}_1)+\varepsilon_1)} + e^{-\beta(E_2(\vec{x}-\vec{x}_2)+\varepsilon_2)} \right]^{N_T}$$
$$\times \delta(r(\vec{x}) - r) \cdot \delta(r'(\vec{x}) - r') d\vec{x} \qquad (23)$$

Similarly, the average conformation at given (r, r') is

$$\langle \vec{x} \rangle_{r,r'} = \frac{1}{Z(r, r')} \int e^{-\beta_T E(\vec{x})} \vec{x} \delta(r(\vec{x}) - r) \delta(r'(\vec{x}) - r') d\vec{x}$$
$$(24)$$

For integer N_T, Eqs. (16) and (19) can be computed analytically from Gaussian integrals obtained after a binomial expansion (see Appendix).

Transition paths

Once the SP is found, one can start from the SP to trace the SD path that connects the two minima. The two branches of the SD path satisfy $\dot{\vec{x}} = -\nabla E(\vec{x})/|\nabla E(\vec{x})|$ with initial values $\vec{x}_\pm(0) = \vec{x}_{sp} \pm \varepsilon \vec{n}_d$, where $\vec{n}_d$ is the direction of the eigenvector of the local Hessian with negative eigenvalue, and $\varepsilon \to 0$. The SP path is divided into frames at intervals of ~ 0.1 Å of RMSD.

Equation (11) provides a simpler definition of a transition path. If we use the Boltzmann weight f_1 as a parameter within [0,1], $\vec{x}_{sp}(f_1)$ traces a path that connects the minima and SPs as f_1 is varied from 1 to 0. Here, we use 100 steps at intervals of $\delta f_1 = 0.01$. Interestingly, the resulting "SP path" is independent of the mixing temperature T_m and $\varepsilon_2 - \varepsilon_1$, which means that the path contains the minima and saddles, regardless of the choice of MENM parameters. Furthermore, it is essentially invariant to the choice of mixing potential $E(\vec{x}) = g(E_1(\vec{x} - \vec{x}_1) + \varepsilon_1, E_2(\vec{x} - \vec{x}_2) + \varepsilon_2)$, including the PNM mixing function[25] in a Hessian approximation. The SP path is particularly appealing for energetically driven transitions because it is the trace of the SP as $\varepsilon_2 - \varepsilon_1$, the thermodynamic driving force, is varied from $+\infty$ to $-\infty$. In addition, we also use the PMF along directions $\vec{n}_1$ and $\vec{n}_2$ in configuration space to define a transition state and a "PMF transition path." For that, we analytically compute the two-dimensional PMF along directions $\vec{n}_{sp}$ and $\vec{x}_2 - \vec{x}_1$ in conformation space. The direction $\vec{n}_{sp}$ corresponds to the unstable mode at the SP, and a 1D-projection on this direction gives to a good approximation the highest PMF barrier; $\vec{n}_2$ is the vector connecting the initial and final structure. The minimal free-energy path (PMF path) is then defined as $\langle \vec{x} \rangle_{r,r'_{min}(r)}$ [Eq. (24)], where $r'_{min}(r)$ corresponds to the minimum of (r, r') for given r. To a good approximation, the "optimal direction" $\vec{n}_{sp}$, defined as the eigenvector of the unstable mode at the SP, is given by the gradient of E_1 or E_2 at the SP.

RESULTS

We will illustrate the use of the MENM method by examining the biologically important conformational transitions in two motor proteins (myosin II and KIF1A kinesin), which were analyzed previously by ENM calculation.[13,28–29]

First, we study the power-stroke transition of myosin II from the prepowerstroke state (PDB code: 1VOM[30]) to the postpowerstroke rigor-like state (PDB code: 1Q5G[31]): it takes place after ATP hydrolysis in response to actin binding, resulting in the converter/lever arm undergoing a large-scale rotation, which is accompanied by the closure of the actin-binding cleft within the 50-kDa domain.[31]

Second, we study a similar force-generating transition in KIF1A kinesin from the ADP-bound state (PDB code: 1I5S) to the ATP-like state (PDB code: 1I6I): the neck linker (analogous to the lever arm in myosin) undergoes a disorder-to-order transition to dock against the motor domain in response to ATP-binding at the nucleotide-binding site.[32]

Our procedure is as follows. First, we construct the MENM potential based on the two ENM potentials built from the beginning and end structures (see Methods). Second, we determine the SP for weak mixing and strong mixing, and generate the SD path in each case (Methods) as well as the SP path. Third, we compute the 1D and
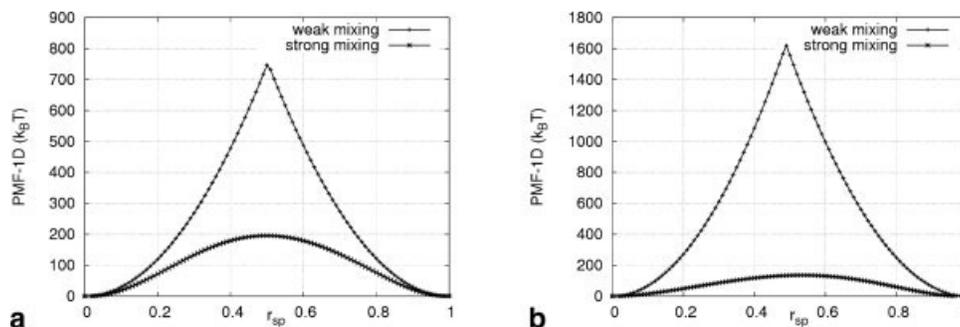
Figure 1

1D Potentials of mean-force (PMF) as a function of the reaction coordinate r$_{sp}$ for (**a**) kinesin and (**b**) myosin in the limits of weak mixing (+; top curve) and strong mixing (×; bottom curve), respectively.

2D PMF and generate transition paths based on free-energy sampling. Fourth, we compare the various transition paths that connect the beginning and end structures. Finally, we analyze the detailed structural changes along the parameter-independent SP path. The main focus will be placed on illustrating the method, although we will devote brief discussions to the biological implications of the results produced by the MENM method.

Calculation of the SP

SPs of the mixed potential surface govern the dynamical transitions that connect the two conformational basins. For both myosin and kinesin, we determine the SP for the mixed potential function (see Methods) for two sets of mixing temperatures: $T_m = T$ (weak mixing), and $T_m = T_{strong}$ (strong mixing). In both cases, we set the energy offset to zero, $\varepsilon_2 = \varepsilon_1$, resulting in a single SP located between the two minima at $\vec{x}_1$ and $\vec{x}_2$. The SPs determined by solving Eq. (12) for weak and strong mixing are nearly identical (RMSDs of 0.02 Å for kinesin and 0.2 Å for myosin). The larger shift of SP in myosin than kinesin may result from relatively stronger mixing, combined with the larger structural difference between the two structures of myosin that render the two ENM potentials to be more different and asymmetric. Note that as T_m is increased above T_{strong}, the energy surface changes qualitatively. The two minima shift significantly away from $\vec{x}_1$ and $\vec{x}_2$, respectively, and toward each other. For $T_m \gg T_{max}$, a new minimum develops near the original SP.

Calculation of PMF

To a high degree of accuracy, the direction $\vec{n}_{sp}$ of the unstable mode at the SP coincides with the gradients of both E_1 and E_2 at the SP. Figure 1 shows the 1D PMFs along the corresponding "optimal" coordinate r_{sp}, as evaluated by projection (see Methods). For weak mixing, the PMF consists of two harmonic potentials joined in the middle (Fig. 1; top curves). There are two minima at $r_{sp} = 0$ and 1. The maximum is at a cusp-like joint. For strong mixing, the two minima remain at $r_{sp} = 0$ and 1, but at the barrier the PMF has a more rounded shape (Fig. 1; bottom curves). The maximum of the PMF is much lower because of the increased mixing. However, its location shifts only slightly along r_{sp}, remaining at 0.5 for kinesin and moving from 0.49 to 0.53 for myosin, consistent with the earlier finding of a larger shift of the SP for myosin than kinesin.

We also compute the 2D PMF in the plane spanned by the unstable mode at the SP (Fig. 2) and the distance vector between the two structures. The vectors $\vec{n}_1 = \vec{n}_{sp}$ and $\vec{n}_2 = \vec{x}_2 - \vec{x}_1$ (with the component of $\vec{n}_2$ along $\vec{n}_{sp}$ projected out) define the reaction coordinates r_{sp} and r_{12}, respectively (see Methods). The 2D PMF provides a more detailed view of the free energy landscape and the transition paths than the 1D PMF, especially near the SP.

At weak mixing, the 2D PMF is dominated by two large harmonic free-energy wells centered at (0,0) and (1,1), which elongate predominantly along the r_{12} axis. The two basins border along a narrow region in the middle that extends parallel to the r_{12} axis. At strong mixing, the border region expands significantly while the two basins shrink accordingly. This is particularly noticeable in myosin because of the relatively strong mixing. Despite these changes in the PMF landscape, its SP only shifts slightly [remaining at (0.50, 0.49) for kinesin, shifting from (0.49, 0.56) to (0.53, 0.58) for myosin]. The larger shift of the SP for myosin is consistent with similar observations made in the previous sections.

Comparison of the three transition paths

On the basis of the calculations presented in the previous sections, we can generate three different transition paths that all connect the two structures $\vec{x}_1$ and $\vec{x}_2$ with the SP (see Supplementary Materials for movies of the SP paths): (1) The SD path follows the gradient from the
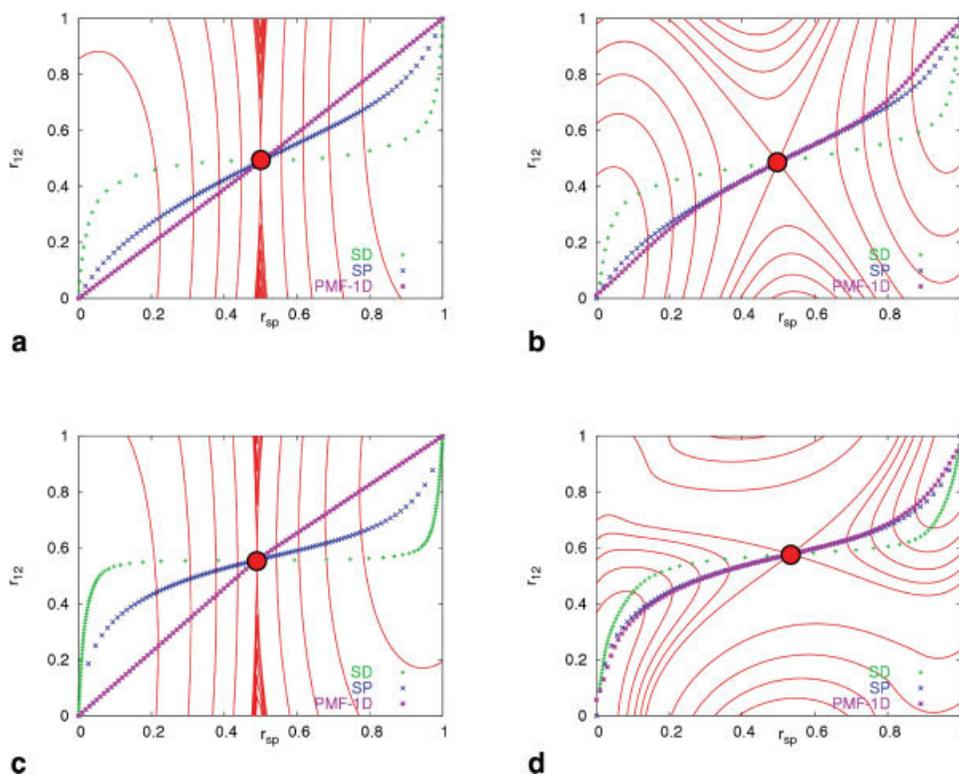
Figure 2

*2D Potentials of mean-force as a function of reaction coordinates r_{sp} and r_{12} for (**a**) weak mixing (kinesin), (**b**) strong mixing (kinesin), (**c**) weak mixing (myosin), and (**d**) strong mixing (myosin). Contour lines (red) are drawn at intervals of (a) 150, (b) 40, (c) 324, and (d) 27 $k_B T$. The three transition paths (SD, green +; SP, blue ×; and PMF-1D, purple *) are projected onto the 2D PMF surface. The initial and final structure correspond the bottom left and top right corners, respectively.*

SP toward the beginning and end structures. (2) The SP path traces $\vec{x}_{sp}(f_1)$ according to Eq. (11), as f_1 is varied from 1 to 0. Unlike the other two paths, it is invariant to parameter changes and even the detailed form of the mixing function (see Methods). (3) The PMF-1D path traces $\langle \vec{x} \rangle_r$ [Eq. (20)] as r is varied from 0 to 1. We find that the PMF-1D paths are essentially identical to the PMF-2D paths based on Eq. (24).

At weak mixing, the PMF-1D path consists of two roughly linear segments. It initially follows the direction given by $H_1^{-1}\vec{n}_{sp}$ or $\vec{x}_{sp} - \vec{x}_1$, then it turns around near $\vec{x}_{sp}$ to follow the direction of $-H_2^{-1}\vec{n}_{sp}$ or $\vec{x}_2 - \vec{x}_{sp}$ [Fig. 2(a,c)]. In general $\vec{x}_{sp}$ is not located on the straight line connecting $\vec{x}_1$ with $\vec{x}_2$ (for example, in the case of myosin). In case of strong mixing, this path is more nonlinear [Fig. 2(b,d)], reflecting contributions from many competing terms in the binomial expansion (see Methods).

To perform a quantitative comparison between the three different transition paths, we compute a distance between two paths P_1 and P_2, $D(P_1, P_2) = \max_i \{\min_j [\text{RMSD}(P_{1,i}, P_{2,j})]\}$, defined in terms of the RMSD between frames $P_{1,i}$ of path P_1 and $P_{2,j}$ of path P_2. Overall, the SD and PMF-1D paths deviate relatively little from the SP path, when compared with the measured RMSD of the conformational

transition (Table I). The deviations are smaller in the strong mixing limit than in the weak mixing limit, indicating that the three paths converge better at strong mixing. After being projected onto the 2D PMF plane, the SP path indeed lies very close to the PMF-1D path, especially in the strong-mixing limit (Fig. 2). Therefore, the SP path is not only parameter-independent, but also representative of the low free-energy path PMF-1D. We will thus primarily use the SP path.

Structural changes along the transition paths

The paths defined earlier contain detailed structural information about the conformational changes associated

Table I

Maximal Deviations $D(P_1, P_2)$ Between the Three Different Transition Paths (in Å Units)

		D(SP, SD)		D(SP, PMF-1D)	
	RMSD$_{transition}$	Weak	Strong	Weak	Strong
Kinesin	2.2	0.41	0.35	0.36	0.12
Myosin	6.6	0.72	0.70	0.82	0.71

Table II
High-Strain-Energy Residues (Top 10%) for Kinesin (top) and Myosin (bottom)

High-strain residues (reference to beginning-state)	High-strain residues (reference to end-state)
Kinesin: 4, 5, 6, 151, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 254, 255, 267, 268, 269, 270, 271, 272, 282, 285, 286, 320, 321, 324, 352	4, 5, 151, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 266, 267, 268, 269, 270, 271, 272, 273, 274, 286, 317, 318, 319, 320, 321, 351
Myosin: 19, 20, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31, 32, 82, 84, 92, 94, 97, 120, 121, 178, 179, 180, 181, 182, 186, 209, 212, 216, 231, 232, 233, 454, 457, 458, 459, 479, 480, 483, 484, 486, 487, 488, 489, 490, 495, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 573, 574, 590, 627, 677, 678, 679, 680, 682, 683, 685, 686, 687, 688, 689, 693, 747	16, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 93, 94, 95, 96, 97, 120, 121, 179, 180, 181, 182, 185, 186, 203, 209, 212, 216, 233, 234, 235, 236, 454, 457, 458, 479, 480, 481, 483, 484, 485, 486, 487, 491, 501, 503, 504, 506, 507, 508, 509, 510, 511, 573, 574, 590, 621, 628, 675, 678, 679, 680, 682, 683, 684, 685, 687, 689, 693, 694, 695, 720, 747

Note that the strain energy is calculated relative to the initial and final structures.

with the transitions. In the following, we will first analyze the evolution of the elastic strain energy and the amplitude of motion at each residue along the transition paths[25] (Table II). Both quantities are computed with reference to the beginning structure for all frames prior to the SP conformation, and relative to the end structure beyond the SP.

In a second step, we will analyze the time-ordering of residue contact forming/breaking along the transition path. As a result of the transition, "old" residue contacts specific to the beginning state are broken while "new" end-state specific contacts are formed. Experimentally, the relevance of residue contacts can be probed by adapting the Φ-value analysis of protein folding.[3] On the basis of the MENM calculations, we can identify new contacts that form early (i.e., before reaching the transition state) and old contacts that break early along the transition paths. Those contacts should strongly affect the transition rate and have correspondingly high Φ values. To assess the time-ordering of residue contact forming and breaking, we define f_{cont} as the relative frame number along the transition path at which the contact "breaks" (or "forms"). A contact between residue pair (i, j) is defined to break (form) when the pair distance d_{ij} between the two C_α atoms reaches 1.2 times the initial (final) value. For example, for a residue pair at an initial distance of $d_{ij} = 10$ Å and an end distance of 20 Å, with $d_{ij} < 12$ Å for the first 70 of 100 frames between beginning and end state, f_{cont} would be 0.7. For the new (old) contacts, we only consider those residue pairs whose minimal distance between pairs of heavy atoms is less than 4 Å in the end (beginning) state but not in the beginning (end) state, and the change in d_{ij} is at least 2 Å between the two states. The results are given in Table III.

Kinesin

A structural superposition of the ADP-bound and the ATP-like structures has revealed two regions with marked differences[32]: the first consists of loop L9 (residue range: 202–218), which includes the switch I element (residue range: 211–218) of the nucleotide-binding site; the second is the "switch II cluster" (residue range 248–324: including loop L11, helix α4, loop L12, helix α5, and loop L13) containing the switch II element (residue range: 248–253). The two switches are believed to be essential to sense the γ-phosphate of a bound nucleotide and trigger nucleotide-dependent conformational changes in motor proteins such as kinesins.[34] Our analysis aims to map the aforementioned two major structural changes to the sequence of dynamical motions that connect the beginning and end states of the kinesin transition.

First, we compare the structures of the SP conformation and the beginning (ADP-bound state) and end (ATP-like state) structures [Fig. 3(a)]. We focus on the significant structural differences in the loop L9 cluster, the switch II cluster and the α6 helix (residue range: 335–351). In the SP conformation, the loop L9 region resembles the end state more than the beginning state [Fig. 3(a)], indicating its high mobility during the transition. In the switch II cluster, the helix α4 appears to adopt a more mixed conformation: it is unwound at the N terminus (just like the ATP-like state structure), but at the C terminus its position is closer to the beginning state structure. Overall, it is structurally closer to the beginning state, indicating its lack of mobility in the transition. This result agrees with an earlier ENM-based normal mode analysis of the ADP-bound structure,[13] which found that the low-frequency normal modes do not sustain large-scale motions in the α4 helix. In the α6 helix, the SP conformation is closer to the end-state structure: both show a small translocation toward the C-terminus compared with the beginning state structure.

Next we examine the frame-by-frame change in strain energy[25] and amplitude of motions at each residue position [Fig. 4(a)]. Both quantities are computed with reference to the beginning (end) state structure for the frames prior to (after) the SP conformation, resulting for some residues in an abrupt change at the frame of the SP conformation. For structures near the SP, the strain

Table III
Time-Ordering of Breaking Beginning-State Residue Contacts (Column 1) and Forming End-State Contacts (Column 2) Along the SP Path for Kinesin (top), Myosin (bottom).

Pair	Distance change (Å)	F_{cont}	Pair	Distance change (Å)	F_{cont}
208, 155	11.60->20.54	0.18	204, 148	11.73->8.24	0.19
208, 202	8.22->17.69	0.10	204, 155	13.16->9.50	0.23
208, 203	6.86->14.90	0.07	211, 205	11.65->5.37	0.49
211, 150	9.58->15.33	0.31	211, 206	10.49->4.53	0.33
213, 148	8.21->11.26	0.68	212, 205	10.10->5.12	0.66
213, 150	7.18->11.15	0.43	213, 205	10.04->3.89	0.43
216, 211	7.72->12.51	0.59	214, 206	10.65->4.66	0.88
267, 211	9.57->17.07	0.34	269, 152	10.58->5.33	0.98
267, 251	8.32->11.27	0.84	271, 264	10.65->6.62	0.48
282, 6	7.16->10.59	0.87	271, 265	10.09->6.71	0.50
285, 5	7.16->11.40	0.92	272, 216	10.86->7.53	0.63
285, 6	6.97->10.91	0.89	272, 263	11.42->8.85	0.13
286, 6	5.63->10.31	0.71	273, 152	10.67->7.37	0.53
289, 5	7.40->10.72	0.97	321, 316	10.12->6.64	0.78
318, 243	6.87->10.19	0.72	347, 283	10.08->7.68	0.18
322, 84	10.37->12.39	1.00			
347, 275	9.45->11.73	0.90			
83, 28	5.66->10.09	0.54	20, 7	12.82->6.30	0.79
84, 28	5.89->11.17	0.11	23, 18	13.98->7.39	0.90
212, 201	6.42->10.92	0.25	24, 17	13.31->8.65	0.53
212, 202	8.06->13.22	0.28	233, 180	10.15->7.46	0.58
219, 190	8.07->10.41	0.48	482, 121	11.58->8.73	0.20
223, 190	11.36->15.06	0.50	485, 121	11.32->8.95	0.02
232, 181	6.29->10.40	0.76	489, 121	14.53->8.82	0.13
237, 186	7.56->11.54	0.53	505, 484	12.19->6.24	0.39
455, 185	7.87->11.26	0.41	507, 480	12.17->4.86	0.88
457, 181	4.62->10.40	0.01	633, 590	11.46->9.44	0.03
457, 185	8.12->11.82	0.66	674, 232	16.48->12.64	0.30
458, 179	6.83->10.00	0.64	678, 458	11.63->5.60	0.91
458, 181	5.57->13.43	0.06	680, 122	11.30->9.20	0.01
475, 179	8.05->12.55	0.28	683, 479	13.04->8.51	0.45
478, 178	7.57->11.00	0.72	686, 509	16.28->10.99	0.75
591, 586	6.79->10.90	0.23	687, 483	11.12->8.54	0.02
620, 589	10.88->16.09	0.74	688, 97	11.73->8.67	0.42
620, 590	10.22->13.05	0.98	693, 92	12.44->5.64	0.98
626, 534	7.48->12.04	0.49	694, 92	15.43->6.61	0.49
627, 620	6.35->10.25	0.08	694, 93	17.10->5.44	0.99
654, 479	10.15->12.29	0.98	694, 94	20.23->7.25	0.98
654, 482	10.15->12.23	0.99	694, 95	21.53->6.72	0.96
662, 232	7.65->15.02	0.04	704, 84	38.82->11.35	0.99
662, 233	7.38->13.07	0.01	747, 501	14.12->11.38	0.01
662, 234	8.63->13.00	0.03			
679, 482	8.69->10.88	0.86			
680, 487	10.07->14.53	0.45			
680, 508	6.97->15.19	0.36			
682, 97	7.12->10.11	0.30			
683, 487	8.73->12.83	0.64			
683, 506	6.32->10.42	0.88			
683, 508	8.49->13.04	0.80			
684, 487	7.74->10.41	0.89			
687, 491	9.09->13.00	0.93			
688, 121	6.68->10.68	0.27			

energy distribution is quite similar relative to the beginning and end structures, although some differences exist [Fig. 4(a): left panel]: relative to the beginning structure, the strain energy is more concentrated in the loop L9 cluster and the N-terminal part of the switch II cluster;

relative to the end structure, more strain energy is shifted to the switch II cluster (especially α4 and α5 helices). The amplitude plot [Fig. 4(a): right panel] points to a consistent picture: the motions in the loop L9 cluster dominate before the SP, while the motions in helices α4 and α5 significantly increase after the SP [Fig. 4(a)]. These results agree with the earlier observations based on structural comparisons [Fig. 3(a)].

Finally, we analyze the time-ordering of residue contact forming/breaking along the transition path. The residue contacts specific to the ADP-bound state break in two stages: near the ADP, the contacts involving residues 155 and 202, 203, 208 of loop L9 break early [colored in red, see Fig. 5(a) and Table III]; far from the ADP, the contacts involving residues 5, 6, 84, 251, 271, 275, 282, 285, 289, 322, 347 (most of them are part of the switch II cluster and nearby regions) break late [colored in purple, see Fig. 5(a) and Table III]. Such two-stage ordering is consistent with an earlier study[13]: without being favored by lowest normal modes, the observed large-scale motions in the α4 helix are delayed to near the end of the transition path. The earliest-broken contact involving the switch II cluster (contact 267-211) appears to provide a link between the loop L9 cluster (near ADP) and the switch II cluster (far from ADP), possibly playing a role in relaying the nucleotide-binding signals from the former to the latter.

Early-forming residue contacts specific to the ATP-like state involve residues from both the loop L9 cluster (148, 204), and the switch II cluster and nearby regions (263, 272, 283, 347) [colored red, see Fig. 5(b)]. The contact 347-283 between helix α6 and helix α4 forms early, which agrees with the recent finding of a hinge motion of the α6 helix encoded in a lowest normal mode of the ADP-bound structure.[28] The earliest-formed contact involving the switch II cluster [contact 272-263, see Fig. 5(b)] is adjacent to the early-broken contact 267-211, which may jointly relay the signals from the nucleotide-binding site to the switch II cluster.

Myosin

A recently solved crystal structure (PDB code: 1Q5G) of the myosin II motor domain in the rigor-like state was compared with the previously determined structures of myosin II in the prepowerstroke state (PDB code: 1VOM) and the postrigor state (PDB code: 1MMA). Three striking differences were identified[31]: first, the switch I region (residue range: 233–238) has moved 8 Å away from its original position, and the switch II (residue range: 454–459) is also in the open state. In comparison, 1VOM has both switch I and II closed (C/C); 1MMA has switch I closed and switch II open (C/O). Second, the 50-kDa domain undergoes a large rotation that alters the relative orientation between its two parts: the U50 and L50 subdomains, resulting in a new configu-
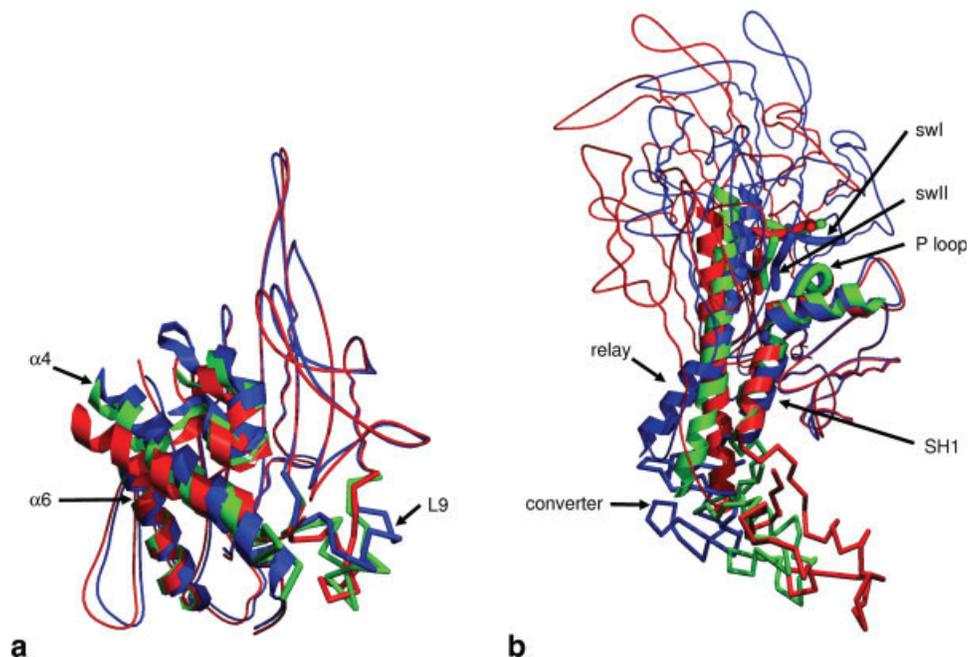
Figure 3

*Structural comparisons of the SP conformation (green) with the beginning (blue) and end structures (red) for (**a**) kinesin and (**b**) myosin. (a) Kinesin. For clarity, the SP structure (green) shows only regions with large motions: the L9 cluster (202–218), the switch II cluster (248–324), and the α6 helix (335–351). The structural alignment uses 40 β-strand residues from residues (90–150, 218–248) with RMSD = 0.36 Å between the beginning and end structures. (b) Myosin. For clarity, the SP structure (green) only shows the P-loop (179–186), switch I (233–238), switch II (454–459), relay helix (466–498), SH1 helix (668–690), and converter (690–747). Part of the N-terminal subdomain is not shown. The structural alignment uses 70 β-strand or α-helical residues from the core of the N-terminal domain (range: 88–198) with RMSD = 0.62 Å between the beginning and end structures.*

ration poised for actin-binding. Third, the lever arm and converter rotate from an up to a down position consistent with a force-generating power-stroke. Here we focus on the transition from 1VOM (beginning state) to 1Q5G (end state).

First, we perform a structural comparison of the SP conformation with the beginning and end structures [Fig. 3(b)]. We focus on the significant structural differences in the nucleotide-binding site (including the P-loop, the switch I and the switch II), the relay helix, the SH1 helix, and the converter. In the SP conformation, both switches adopt positions intermediate between the beginning and the end state, with the P loop being well aligned in all three structures. The relay helix (analogous to the helix α4 in kinesin) also adopts an intermediate conformation. However, unlike the helix α4 in kinesin, the C terminus of the relay helix (together with the converter) is closer to the end-state structure. In the SH1 helix (analogous to the α6 helix in kinesin), the SP conformation is closer to the end structure than the beginning structure (similar to what is observed in the α6 helix of kinesin).

Next we examine the evolution in strain energy and amplitude of motions at each residue position along the transition path [Fig. 4(b)]. The distribution of the strain energy of the SP structure relative to the beginning and end structures is similar, except for some subtle differences [Fig. 4(b): left panel]: relative to the beginning structure, the strain energy is more concentrated in the N terminal domain, the C-terminal end of the relay helix (residue range: 500–510), and the SH1 helix. This is consistent with our previous study[29] on an ENM-based fluctuation analysis of the beginning state structure that identified these regions as flexible hinges. The amplitude plot [Fig. 4(b): right panel] shows that the converter and the SH1 helix move more before the SP than after the SP. These results agree with the earlier structural comparison.

Finally, we analyze the time-ordering of residue contact forming/breaking along the transition path. Early-breaking residue contacts specific to the beginning state are mostly near the nucleotide-binding site [including residues 181 of the P loop, 232, 233, 234 of switch I, 457, 458 of switch II, and 662, see Fig. 5(c)], and a few near the actin-binding site (620, 627) and the N-terminal subdomain (28, 84). These residues were previously found at the hinge regions on the myosin motor domain.[29]

Early-forming residue contacts specific to the end state [Fig. 5(d)] are mostly distant from nucleotide-binding site, involving residues on or near the relay helix (482, 483, 485, 489, 501), actin binding site (590, 633), SH1
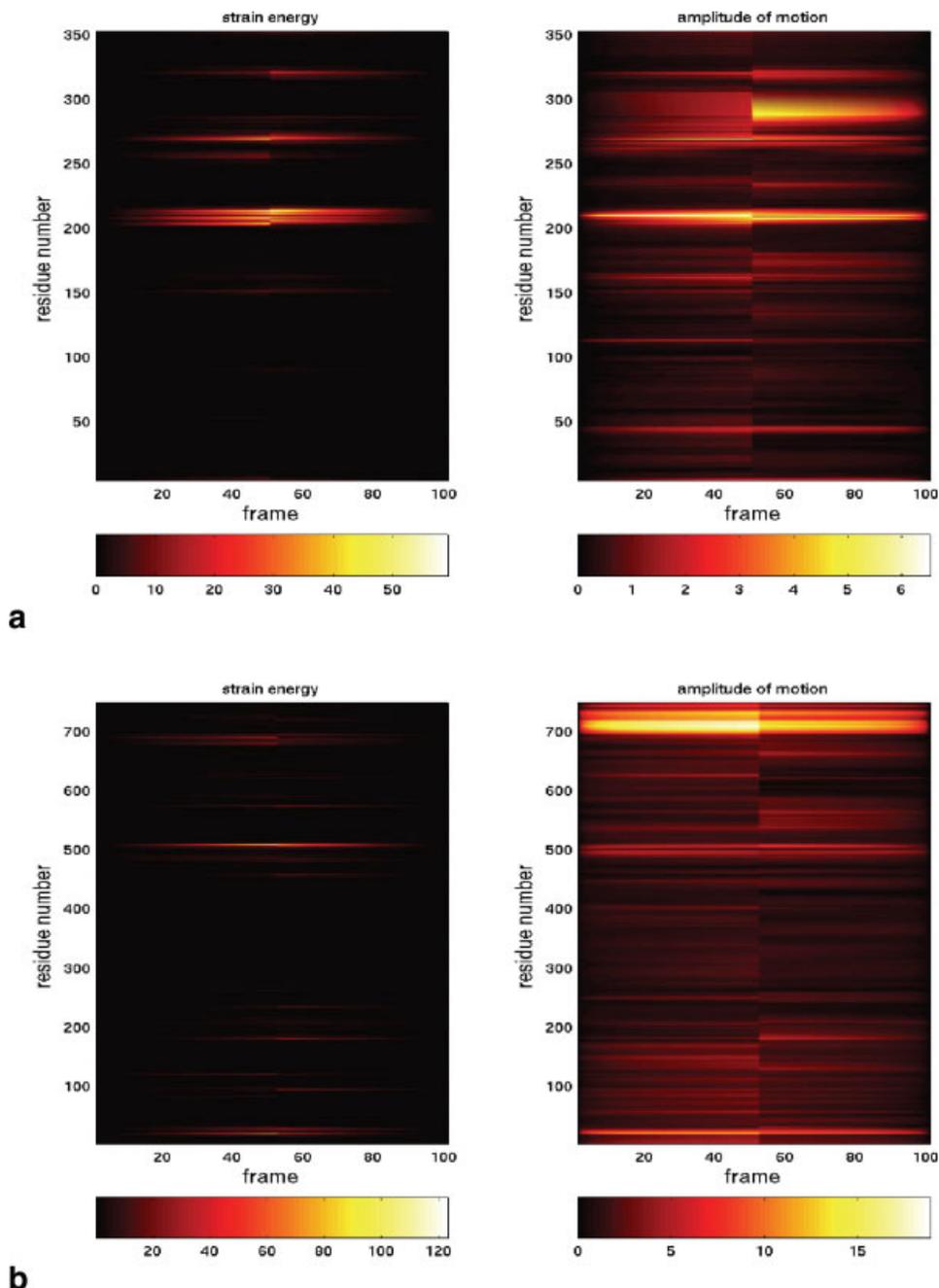
Figure 4

*Frame-by-frame change in "strain energy" (k_BT) and amplitude of motions (Å) at each residue position for (**a**) kinesin and (**b**) myosin. Both quantities are computed with reference to the beginning (end) state structure for the frames prior to (after) the SP conformation.*

helix (680, 687), N-terminal subdomain (121, 122), and converter (747). These residues are also located in the hinge regions.[29]

Unlike kinesin, the relay helix and the SH1 helix in myosin are involved in several early-forming residue contacts (Table III), because their flexibility and mobility are sustained by the lowest ENM-derived normal modes.[28]

These differences may lead to different force-generation mechanisms between these two molecular motors.

CONCLUDING REMARKS

We have explored a formalism to study large-scale conformational changes in proteins with known beginning
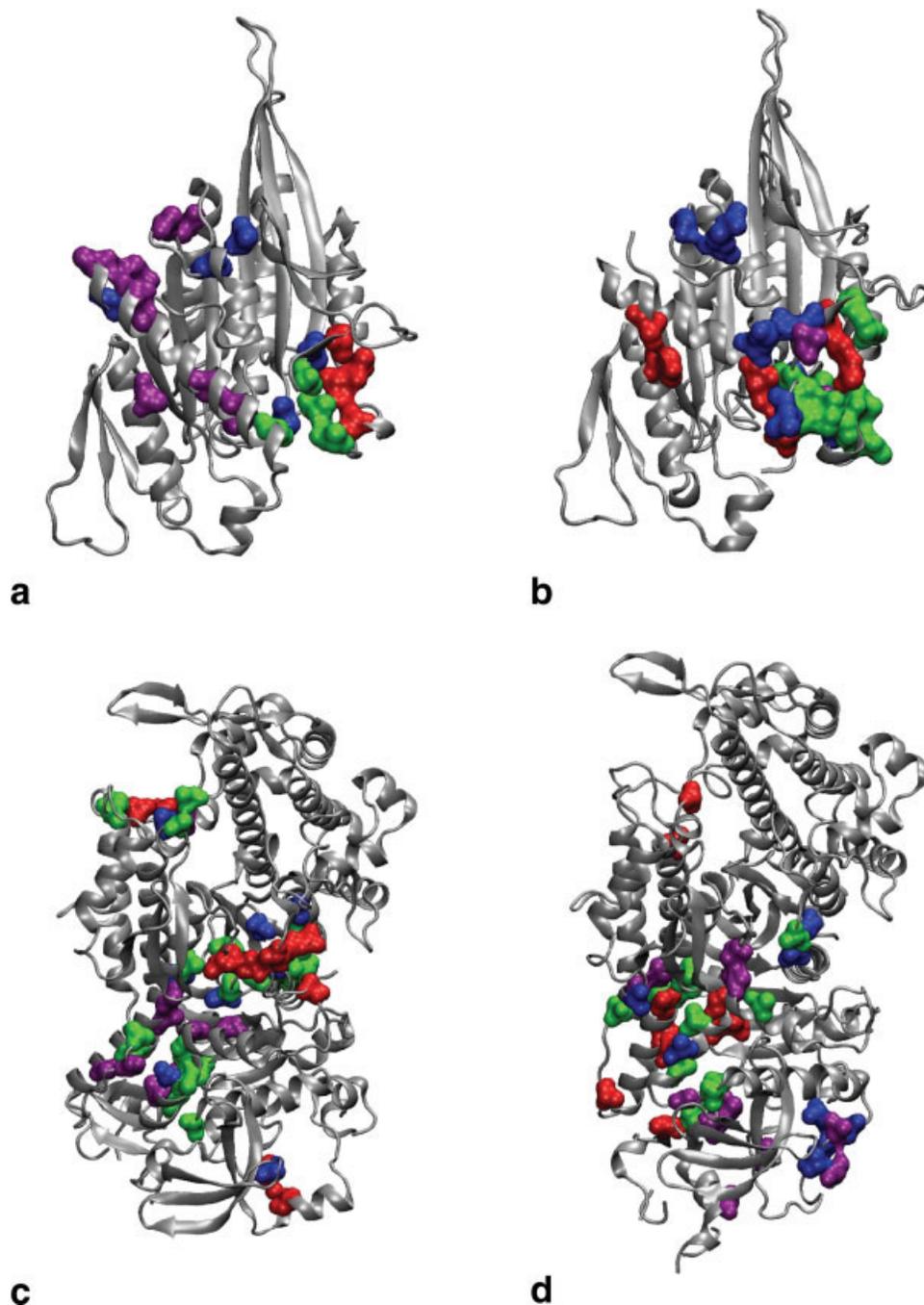
Figure 5

*Time-ordering of residue-contact forming and breaking. (**a**) Kinesin contact breaking; (**b**) kinesin contact forming; (**c**) myosin contact breaking; and (**d**) myosin contact forming. Residue contacts are shown as atomic surfaces. Residues involved in the contacts that form or break early are shown in red [f_{cont} in (0, 0.2)]. Contacts breaking/forming at intermediate times are shown in green [f_{cont} in (0.2, 0.5)] and blue [f_{cont} in (0.5, 0.8)]. Contacts breaking/forming late are shown in purple [f_{cont} in (0.8, 1)].*

and end structures. ENM harmonic potentials for the known structures of the transition end points are mixed into a smooth potential function. The resulting coarse-grained energy surface retains the beginning and end struc-

tures as local minima, connected by a single saddle. The MENM method is applicable if the transition is dominated by collective motions. Such transitions are usually well described by the ENM-derived normal modes. The

MENM formalism is computationally efficient, as it only requires the construction and inversion of Hessian matrices, and is essentially not limited by the size of proteins. It is based on a mathematically simple form of a mixed potential, whose SP and conformation-dependent free energies are easy to compute analytically or numerically. Despite its formal simplicity, the MENM method generates highly nontrivial energy surfaces and transition paths that are different from ad hoc interpolations and can be used to predict time-ordering of multiple movements involved in the conformational changes. The MENM energy surface can be parameterized by X-ray crystallographic B factors, relative populations, and transition rates. But even without this information, one can characterize the structural transition using the parameter-independent SP path. Finally, it is straightforward to generalize this method to more than two structures.

The relative mathematical simplicity of the MENM surfaces permits analytical treatments as well as fast computational approaches. However, their applicability is limited to cases where local (or global) protein unfolding is not a relevant factor in the conformational transition. If such unfolding is potentially important, as was found in the case of the arc-repressor mutant transition between helical and sheet forms,[26] one may want to use instead (or in addition) multistate Gō-type models introduced by Best et al.[26]

The MENM formalism allows us to explore large-scale conformational transitions with limited computational effort by using a coarse-grained model with a simplified energy function. The resulting SP structures and transition paths can be further explored and characterized using molecular dynamics simulations combined with, for instance, transition path sampling[35] or dynamics projected onto collective modes.[36] The MENM calculations also provide experimentally testable predictions for the contributions of specific amino-acid contacts and other structural elements to the transition state ensemble.

An MENM web server is available at http://enm.lobos.nih.gov.

ACKNOWLEDGMENTS

REFERENCES

1. Kern D, Zuiderweg ERP. The role of dynamics in allosteric regulation. Curr Opin Struct Biol 2003;13:748–757.
2. Geeves MA, Holmes KC. Structural mechanism of muscle contraction. Annu Rev Biochem 1999;68:687–728.
3. Fersht AR, Matouschek A, Serrano L. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. J Mol Biol 1992;224:771–782.
4. Rhoades E, Cohen M, Schuler B, Haran G. Two-state folding observed in individual protein molecules J Am Chem Soc 2004;126:14686–14687.
5. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nat Struct Biol 2002;9:646–652.
6. Krebs WG, Gerstein M. The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework. Nucleic Acids Res 2000;28:1665–1675.
7. Fischer S, Karplus M. Conjugate peak refinement: an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. Chem Phys Lett 1992;194:252–261.
8. Zuckerman DM. Simulation of an ensemble of conformational transitions in a united-residue model of calmodulin. J Phys Chem B 2004;108:5127–5137.
9. Tozzini V. Coarse-grained models for proteins. Curr Opin Struct Biol 2005;15:144–150.
10. Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 1996;77:1905–1908.
11. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 2001;80:505–515.
12. Delarue M, Sanejouand YH. Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model. J Mol Biol 2002;320:1011–1024.
13. Zheng W, Doniach S. A comparative study of motor protein motions using a simple elastic network model. Proc Natl Acad Sci USA 2003;100:13253–13258.
14. Navizet I, Lavery R, Jernigan RL. Myosin flexibility: structural domains and collective vibrations. Proteins 2004;54:384–393.
15. Song G, Jernigan RL. An enhanced elastic network model to represent the motions of domain-swapped proteins. Proteins 2006;63:197–209.
16. Van Wynsberghe A, Li GH, Cui Q. Normal-mode analysis suggests protein flexibility modulation throughout RNA polymerase's functional cycle. Biochemistry 2004;43:13083–13096.
17. Li GH, Cui Q. Analysis of functional motions in Brownian molecular machines with an efficient block normal mode approach: myosin-II and Ca2+-ATPase. Biophys J 2004;86:743–763.
18. Krebs WG, Alexandrov V, Wilson CA, Echols N, Yu H, Gerstein M. Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic. Proteins 2002;48:682–695.
19. Bahar I, Rader AJ. Coarse-grained normal mode analysis in structural biology. Curr Opin Struct Biol 2005;15:586–592.
20. Ma JP. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. Structure 2005;13:373–380.
21. Tama F, Brooks CL. Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. Annu Rev Biophys Biomol Struct 2006;35:115–133.
22. Kim MK, Chirikjian GS, Jernigan RL. Elastic models of conformational transitions in macromolecules. J Mol Graph Model 2002;21:151–160.
23. Kim MK, Jernigan RL, Chirikjian GS. Rigid-cluster models of conformational transitions in macromolecular machines and assemblies. Biophys J 2005;89:43–55.
24. Miyashita O, Onuchic JN, Wolynes PG. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. Proc Natl Acad Sci USA 2003;100:12570–12575.
25. Maragakis P, Karplus M. Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase. J Mol Biol 2005;352:807–822.
26. Best RB, Chen YG, Hummer G. Slow protein conformational dynamics from multiple experimental structures: the helix/sheet transition of arc repressor. Structure 2005;12:1755–1763.

27. Okazaki K, Koga N, Takada S, Onuchic JN, Wolynes PG. Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: structure-based molecular dynamics simulations. Proc Natl Acad Sci USA 2006;103:11844–11849.

28. Zheng W, Brooks BR. Probing the local dynamics of nucleotide binding pocket coupled to the global dynamics: myosin versus kinesin. Biophys J 2005;89:167–178.

29. Zheng W, Brooks BR. Identification of dynamical correlations within the myosin motor domain by the normal mode analysis of an elastic network model. J Mol Biol 2005;346:745–759.

30. Smith CA, Rayment I. X-ray structure of the magnesium(II).ADP. vanadate complex of the *Dictyostelium discoideum* myosin motor domain to 1.9 A resolution. Biochemistry 1996;35:5404–5417.

31. Reubold TF, Eschenburg S, Becker A, Kull FJ, Manstein DJ. A structural model for actin-induced nucleotide release in myosin. Nat Struct Biol 2003;10:826–830.

32. Kikkawa M, Sablin EP, Okada Y, Yajima H, Fletterick RJ, Hirokawa N. Switch-based mechanism of kinesin motors. Nature 2001;411:439–445.

33. Hinsen K. Analysis of domain motions by approximate normal mode calculations. Proteins 1998;33:417–429.

34. Kull FJ, Endow SA. Kinesin: switch I & II and the motor mechanism. J Cell Sci 2002;115(Part 1):15–23.

35. Dellago C, Bolhuis PG, Csajka FS, Chandler D. Transition path sampling and the calculation of rate constants. J Chem Phys 1998;108:1964–1977.

36. García AE. Large-amplitude non-linear motions in proteins. Phys Rev Lett 1992;68:2696–2699.

APPENDIX

1D Potential of Mean Force

The PMF at temperature $1/\beta_T$ along a reaction coordinate R that is obtained by projection onto a normalized sampling vector $\vec{n}$, $R(\vec{x}) = (\vec{x} - \vec{x}_1) \cdot \vec{n}$, is defined as

$$F(R) = -\frac{1}{\beta_T} \log(Z(R)),$$

with

$$Z(R) = \int e^{-\beta_T E(\vec{x})} \delta(R(\vec{x}) - R) d\vec{x}$$

$$= \int \left[e^{-\beta(E_1(\vec{x}-\vec{x}_1)+\varepsilon_1)} + e^{-\beta(E_2(\vec{x}-\vec{x}_2)+\varepsilon_2)} \right]^{N_T} \delta(R(\vec{x}) - R) d\vec{x}$$

$$= \sum_{m=0}^{N_T} \binom{N_T}{m} \cdot Z_m.$$

In the binomial expansion of the last step, we assumed that

$$N_T = \frac{\beta_T}{\beta},$$

is an integer number. The Z_m can be computed as Gaussian integrals:

$$Z_m(R)$$

$$= \int e^{-\beta[m(E_1(\vec{x}-\vec{x}_1)+\varepsilon_1)+(N_T-m)(E_2(\vec{x}-\vec{x}_2)+\varepsilon_2)]} \delta(R(\vec{x}) - R) d\vec{x}$$

$$= e^{-\beta E_m^*} \int e^{-\frac{\beta}{2}(\vec{x}-\vec{x}_m^*)^T H_m(\vec{x}-\vec{x}_m^*)} \delta(R(\vec{x}) - R) d\vec{x}$$

$$= e^{-\beta E_m^*} \int e^{-\frac{\beta}{2}(\vec{x}^T H_m \vec{x})} \delta(R(\vec{x} + \vec{x}_m^*) - R) d\vec{x}$$

$$= e^{-\beta E_m^*} \int e^{-\frac{\beta}{2} x_P^T H_{PP} x_P + 2 x_P^T H_{PQ} \vec{x}_Q + \vec{x}_Q^T H_{QQ} \vec{x}_Q} \delta(x_P - x_P^*) dx_P \, d\vec{x}_Q$$

$$= e^{-\beta E_m^* - \frac{\beta}{2} x_P^{*T} H_{PP} x_P^*} \int e^{-\frac{\beta}{2}(2 x_P^{*T} H_{PQ} \vec{x}_Q + \vec{x}_Q^T H_{QQ} \vec{x}_Q)} d\vec{x}_Q$$

$$= e^{-\beta E_m^* - \frac{\beta}{2} \vec{x}^{*T} H_m \vec{x}^*} \int e^{-\frac{\beta}{2}(\vec{x}_Q - \vec{x}_Q^*)^T H_{QQ}(\vec{x}_Q - \vec{x}_Q^*)} d\vec{x}_Q$$

$$= \left(\frac{2\pi}{\beta}\right)^{n_Q/2} \cdot \frac{e^{-\beta E_m^* - \frac{\beta}{2} \vec{x}^{*T} H_m \vec{x}^*}}{\sqrt{\det(H_{QQ})}}$$

$$= \left(\frac{2\pi}{\beta}\right)^{n_Q/2} \cdot \frac{e^{-\beta E_m^* - \frac{\beta}{2} \frac{(R-R_m^*)^2}{\vec{n}^T H_m^{-1} \vec{n}}}}{\sqrt{\det(H_M) \cdot (\vec{n}^T H_m^{-1} \vec{n})}}$$

with n_Q the number of nonzero eigenvalues of H_{QQ}, and the determinant being their product,

$$P = \vec{n} \cdot \vec{n}^T, \ Q = I - P,$$
$$x_P = \vec{n}^T \cdot \vec{x}, \ \vec{x}_Q = Q\vec{x}$$

$$H_{PP} = \vec{n}^T H_m \vec{n}, \ H_{PQ} = \vec{n}^T \cdot H_m Q, \ H_{QQ} = Q H_m Q$$

and

$$\vec{x}^* = \vec{x}_Q^* + x_P^* \cdot \vec{n}$$
$$= -(R - R_m^*) \cdot H_{QQ}^{-1} H_{QP} + (R - R_m^*)\vec{n}$$
$$= \frac{(R - R_m^*) \cdot H_m^{-1} \vec{n}}{\vec{n}^T H_m^{-1} \vec{n}}$$

where $R_m^* = (\vec{x}_m^* - \vec{x}_1) \cdot n$. H_m and $\vec{x}_m^*$ are defined in the next subsection.

Note: $\vec{n}$ must have zero overlap with the six translation/rotation zero-modes of H_1 (and H_2)

2D Potential of Mean Force

The 2D PMF can be computed similarly for two sampling vectors $\vec{a}$ and $\vec{b}$:

$$Z(R_1, R_2) = \int e^{-\beta_T E(\vec{x})} \delta(R_1(\vec{x}) - R_1) \cdot \delta(R_2(\vec{x}) - R_2) d\vec{x}$$

$$= \int \left[e^{-\beta(E_1(\vec{x}-\vec{x}_1)+\varepsilon_1)} + e^{-\beta(E_2(\vec{x}-\vec{x}_2)+\varepsilon_2)} \right]^{N_T} \delta(\vec{x} \cdot \vec{a} - R_1)$$

$$\times \delta(\vec{x} \cdot \vec{b} - R_2) d\vec{x} = \sum_{m=0}^{N_T} \binom{N_T}{m} \cdot Z_m,$$

where $N_T = \frac{\beta_T}{\beta}$ and

$$Z_m = \int e^{-\beta[m(E_1(\vec{x}-\vec{x}_1)+\varepsilon_1)+(N_T-m)(E_2(\vec{x}-\vec{x}_2)+\varepsilon_2)]}\delta(\vec{x}\cdot\vec{a}-R_1)$$

$$\times \delta(\vec{x}\cdot\vec{b}-R_2)\mathrm{d}\vec{x}$$

$$= e^{-\beta E_m^*}\int e^{-\beta[\frac{1}{2}(\vec{x}-\vec{x}_m^*)^T H_m(\vec{x}-\vec{x}_m^*)]}\delta(\vec{x}\cdot\vec{a}-R_1)\cdot\delta(\vec{x}\cdot\vec{b}-R_2)\mathrm{d}\vec{x}$$

$$= \left(\frac{2\pi}{\beta}\right)^{n_Q/2}\frac{e^{-\beta(E_m^*+E_{\min}(R_1-R_{m,1}^*,R_2-R_{m,2}^*))}}{\sqrt{\det(QH_mQ)}}$$

where

$$E_m^* = m(E_1(\vec{x}_m^*-\vec{x}_1)+\varepsilon_1)+(N_T-m)(E_2(\vec{x}_m^*-\vec{x}_2)+\varepsilon_2)$$

$$\vec{x}_m^* = H_m^{-1}\cdot[mH_1\vec{x}_1+(N_T-m)H_2\vec{x}_2]$$

$$H_m = mH_1 + (N_T-m)H_2$$

$$R_{m,1}^* = (\vec{x}_m^*-\vec{x}_1)\cdot\vec{a}$$

$$R_{m,2}^* = (\vec{x}_m^*-\vec{x}_1)\cdot\vec{b}$$

$$E_{\min}(R_1,R_2) = \frac{1}{2}\sum_{m=0}^{N_T}[R_1 \quad R_2]\left\{\begin{array}{cc}\vec{a}H_m^{-1}\vec{a} & \vec{a}H_m^{-1}\vec{b} \\ \vec{a}H_m^{-1}\vec{b} & \vec{b}H_m^{-1}\vec{b}\end{array}\right\}^{-1}\begin{bmatrix}R_1 \\ R_2\end{bmatrix}$$

$$P = \vec{a}\cdot\vec{a}^T + \vec{b}\cdot\vec{b}^T, \quad Q = I - P$$