

Methods of Mixed ENM for mapping transition paths

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

¹Laboratory of Computational Biology, National Heart, Lung, and Blood Institute,
National Institutes of Health, Bethesda, MD 20892, USA

²Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and
Kidney Diseases, National Institutes of Health, Building 5, Room 132, Bethesda, MD
20892, USA.

Potential function of elastic network model

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

$$E(\bar{x} - \bar{x}_0) = \frac{1}{2} \sum_{d_{ij}^0 < R_c} C (d_{ij} - d_{ij}^0)^2, \quad (1)$$

where \bar{x} is a $3N$ -dimensional vector representing the Cartesian coordinates of the N C_α atoms, \bar{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 can expand the above potential function to second order by computing its Hessian matrix H :

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

where $\delta\bar{x} = \bar{x} - \bar{x}_0$. 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 .

Fitting B factors to calibrate 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:

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

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.

Mixed potential function constructed from two ENMs

Given two protein structures \bar{x}_1 and \bar{x}_2 corresponding to the beginning and end states of a conformational transition, respectively, we construct two ENMs $E_1(\bar{x} - \bar{x}_1)$ and $E_2(\bar{x} - \bar{x}_2)$ with Hessians H_1 and H_2 according to Eq. (1) and Eq. (2). We then define the MENM potential function:

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

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 $E(\bar{x})$ 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 $E(\bar{x}) \rightarrow E_1(\bar{x}) + E_2(\bar{x}) + \varepsilon_1 + \varepsilon_2$. For small but finite T_m , $E(\bar{x})$ is approximately $E_1(\bar{x}) + \varepsilon_1$ near \bar{x}_1 and $E_2(\bar{x}) + \varepsilon_2$ near \bar{x}_2 , with a smooth interpolation between E_1 and E_2 elsewhere in configuration space. The MENM energy surface $E(\bar{x})$ has two dominant local minima corresponding to the beginning and end structures of the transition, and connects them with a saddle point whose height can be tuned by the mixing temperature T_m . $E(\bar{x})$ thus seems well suited to explore the transition between two protein conformations that represent different functional states. Eq. (4) can be easily generalized to mixing of more than two ENMs. 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, saddle points, and transition paths. In particular, the partition function of the MENM system at temperature $T=300$ K is given by

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

where

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

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.

Parameterization of MENM

The MENM potential has two parameters: T_m (mixing temperature) and $\varepsilon_2 - \varepsilon_1$ (difference in energy offsets). First, $\varepsilon_2 - \varepsilon_1$ is determined by the relative populations of the two end

states at equilibrium, as measured by experiments. Within the harmonic approximation to the partition functions of the two states, the ratio of populations in states 1 and 2 is given by

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

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

Secondly, a bound on T_m can in principle be determined from the experimental rate of transitions as follows:

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

Here the pre-factor ω_0 is assumed to have an upper bound of 1 ps^{-1} and the free energy barrier ΔG is approximated by E_{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 and efficiency of running PATH-ENM, we only consider the extreme case of weak mixing ($T_m = T$) with $\varepsilon_2 = \varepsilon_1$.

Saddle points and minima of mixed energy surface

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

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

where

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

$$f_2(\vec{x}) = 1 - f_1(\vec{x}).$$

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

$$\vec{x}_{sp} = [f_1(\vec{x}_{sp})H_1 + (1 - f_1(\vec{x}_{sp}))H_2]^{-1} [f_1(\vec{x}_{sp})H_1\vec{x}_1 + (1 - f_1(\vec{x}_{sp}))H_2\vec{x}_2] \quad (11)$$

where the matrix inverse is restricted to the sub-space spanned by the eigenvectors with non-zero 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

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

Eq. (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 above 1D reduction of the multi-dimensional SP equation can be generalized to a general mixing potential

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

In general, the stationary point is given by

$$\nabla E(\bar{x}) = \frac{\partial g}{\partial E_1} \nabla E_1(\bar{x} - \bar{x}_1) + \frac{\partial g}{\partial E_2} \nabla E_2(\bar{x} - \bar{x}_2) \quad (14)$$

$$= f_1(\bar{x}) \cdot \nabla E_1(\bar{x} - \bar{x}_1) + f_2(\bar{x}) \cdot \nabla E_2(\bar{x} - \bar{x}_2) = 0$$

where

$$f_1(\bar{x}) = \frac{\frac{\partial g}{\partial E_1}}{\frac{\partial g}{\partial E_1} + \frac{\partial g}{\partial E_2}}, \quad (15)$$

$$f_2(\bar{x}) = 1 - f_1(\bar{x}).$$

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 $\bar{x}_{sp}(f_1)$ solved from Eq. (11) is independent of the mathematical form of the mixing function $g(E_1, E_2)$. In particular, the MENM mixing potential (Eq. (4)) satisfies $\frac{\partial g}{\partial E_1} \cdot \frac{\partial g}{\partial E_2} > 0$ and so the SPs should fall along the same 1D SP path disregarding the choice of mixing schemes.

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 above formalism 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 $r(\bar{x})$ for a given direction \vec{n} :

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

So

$$r(\bar{x}_1) = 0, \quad (17)$$

$$r(\bar{x}_2) = (\bar{x}_2 - \bar{x}_1) \cdot \vec{n} \equiv D_{12}.$$

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

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

where

$$\begin{aligned} Z(r) &= \int e^{-\beta_T E(\bar{x})} \delta(r(\bar{x}) - r) d\bar{x} \\ &= \int \left[e^{-\beta(E_1(\bar{x}-\bar{x}_1)+\epsilon_1)} + e^{-\beta(E_2(\bar{x}-\bar{x}_2)+\epsilon_2)} \right]^{N_T} \delta(r(\bar{x}) - r) d\bar{x}. \end{aligned} \quad (19)$$

The average conformation at given r is:

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

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

$$r = r_{sp} = (\bar{x} - \bar{x}_1) \cdot \bar{n}_{sp} / [(\bar{x}_2 - \bar{x}_1) \cdot \bar{n}_{sp}], \quad r' = r_{12} = (\bar{x} - \bar{x}_1) \cdot \bar{n}_2 / [(\bar{x}_2 - \bar{x}_1) \cdot \bar{n}_2]. \quad (21)$$

The 2D PMF is then

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

$$\begin{aligned} Z(r, r') &= \int e^{-\beta_T E(\bar{x})} \delta(r(\bar{x}) - r) \cdot \delta(r'(\bar{x}) - r') d\bar{x} \\ &= \int \left[e^{-\beta(E_1(\bar{x} - \bar{x}_1) + \varepsilon_1)} + e^{-\beta(E_2(\bar{x} - \bar{x}_2) + \varepsilon_2)} \right]^{N_T} \delta(r(\bar{x}) - r) \cdot \delta(r'(\bar{x}) - r') d\bar{x} \end{aligned} \quad (23)$$

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

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

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 steepest-descent (SD) path³⁴ that connects the two minima. The two branches of the steepest-descent path satisfy $\dot{\bar{x}} = -\nabla E(\bar{x}) / |\nabla E(\bar{x})|$ with initial values $\bar{x}_{\pm}(0) = \bar{x}_{sp} \pm \varepsilon \bar{n}_d$ where \bar{n}_d is the direction of the eigenvector of the local Hessian with negative eigenvalue, and $\varepsilon \rightarrow 0$. The SP path is divided into frames at intervals of $\sim 0.1 \text{ \AA}$ of RMSD.

Eq. (11) provides a simpler definition of a transition path. If we use the Boltzmann weight f_l as a parameter within $[0, 1]$, $\bar{x}_{sp}(f)$ traces a path that connects the minima and saddle points as f_l is varied from 1 to 0. Here, we use 100 steps at intervals of $\delta f_l = 0.01$. 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 $E(\bar{x}) = g(E_1(\bar{x} - \bar{x}_1) + \varepsilon_1, E_2(\bar{x} - \bar{x}_2) + \varepsilon_2)$. 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 \bar{n}_1 and \bar{n}_2 in configuration space to define a transition state and a ‘‘PMF transition path’’. For that, we analytically compute the two-dimensional potential of mean force (PMF) along directions \bar{n}_{sp} and $\bar{x}_2 - \bar{x}_1$ in conformation space. The direction \bar{n}_{sp} corresponds to the unstable mode at the saddle point, and a 1D-projection on this direction gives to a good approximation the highest PMF barrier; \bar{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 $\text{PMF}(r, r')$ 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.