Supporting Information

Korkut and Hendrickson 10.1073/pnas.0907684106

SI Text

Algorithm Validation

Robustness of the VAMM Algorithm. Tests were made for the robustness of VAMM transition pathway parameters by comparing pathway calculations made by using the standard parameter set, no. 1, with ones using systematically varied parameter sets (Table S2). In each case, the deviation (rmsd) was computed between the terminal intermediate, as generated from open ADK with the standard set and the particular test set. In all cases, except lower-mode range (10 modes), the rmsd between the terminal intermediate states is within 1 Å compared with the terminal intermediate obtained with the native parameter set.

Step Size. For both smaller and larger step sizes, the rmsd is below 1 Å. However, with very small step sizes, as in case 2 ($c = 0.001$), the efficiency of the pathway decreases linearly (≈800 steps vs. 178 with $C = 0.005$), and the rmsd is higher than those with a large step size. We conclude that too-small step sizes should be avoided, because this decreases efficiency of the algorithm without contributing to the geometric accuracy of the generated models. Too-large step sizes should also be avoided so as not to generate unrealistic structures. The current value of 0.005 is a balanced choice.

Mode Range. The rmsd between 20 modes and 50 modes is 0.47 Å, indicating that increasing mode range beyond 20 modes has no significant affect. However, a mode range of 10 modes generates terminal intermediates significantly different from all other test sets. This is consistent with the finding shown in Fig. S2, where it is shown that higher modes (>10) participate in the later stages of the transition pathway. Thus, it is important that a sufficient subset of normal modes should be included. From the current analysis and previous works, we conclude that inclusion of 20 modes is a balanced approach, and system is robust when a sufficient amount of normal-mode information is used during the pathway calculation.

Update Interval. Secondary-structure update frequency has no effect on the terminal intermediate state.

Methods

θ vs. τ and θ^+ vs. θ^- Contour Plots. The favored and allowed regions for virtual bond angles (θ) and virtual dihedral angles (τ) are determined in a similar fashion as for those of the Ramachandran plot, as determined by the MolProbity software (1). The θ^+, θ^-, and τ angles are calculated for all of the residues in the Top500 database used to construct the allowed and favored φ-ψ space in MolProbity. This database contains 500 crystal structures with resolutions better than 1.8 Å. Residues with B factors higher than 40 are excluded from the analysis. Differently from the MolProbity, Gly and Pro residues are analyzed together with other residues for simplicity. Smoothly contoured boundaries for θ^- vs. τ and θ^- vs. θ^+ distributions are obtained using a density-dependent smoothing function (2). This function helps to represent sparse regions of the distribution as smooth, low-density, and continuous regions, whereas the sharp transitions in the distribution are preserved. The formulation of the smoothing function is identical to that used for MolProbity. The density of each point on the distribution is sum of contributions of all N points and represented by ρ:

$$\rho(\tau,\theta) = \sum_{i}^{N} \sigma_i(\tau,\theta) \qquad [S1]$$

where each contribution σ_i is computed by

$$\sigma_i(\tau,\theta) = \frac{\pi}{\alpha_i^2(\pi^2-4)}\left\{\cos\left(\frac{\pi x_i}{\alpha_i}\right)+1\right\}, \qquad x_i < \alpha_i \qquad [S2]$$

$$\sigma_i(\tau,\theta) = 0, \qquad \text{otherwise}$$

and x_i is given by

$$x_i = \sqrt{(\tau-\tau_i)^2+(\theta-\theta_i)^2} \qquad [S3]$$

The final distribution ρ is found in two iterations of the above formulation. In the first iteration, a uniform $\alpha_i = \alpha_0$ parameter is adopted. In the second iteration, α is determined by

$$\alpha_i = k\rho_1(\tau,\theta)^{-\lambda/n} \qquad [S4]$$

and then final ρ_2 is calculated with Eq. **S1**. The $\rho(\theta^-, \theta^+)$ distributions are calculated by the same procedure. In all calculations, the parameters are set as $\alpha_0 = 10$, $\lambda = 0.5$, $n = 2$, and $k = 13$. These values are identical to those for φ-ψ calculations of all atom models and help to preserve the steep transitions and spread the distribution on the sparse parts of the distribution. The contour levels are determined based on the percentage of the sum of the ρ values in the contoured areas to the sum of all ρ values. The favored contour level is plotted to the 95% of the data, and the allowed region is contained at 99% level.

Strain energy of intermediate states. The strain energy of intermediate states during conformational transition is calculated using an approach similar to the double-well potential described in the plastic network model (3). In this model, two simple harmonic potentials are linked to each other to define an energy surface for each residue pair with two basins corresponding to the two conformational states. The equilibrium distances in this harmonic potential are given by the crystal structure distances of the structures.

$$V_{1ij} = k \times (|\vec{R}_{i,j}| - |\vec{R}^0_{1,i,j}|)^2 \qquad [S5]$$

$$V_{2ij} = k \times (|\vec{R}_{i,j}| - |\vec{R}^0_{2,i,j}|)^2$$

$$V_{ij} = \frac{V_{1ij}+V_{2ij}+\sqrt{(V_{1ij}-V_{2ij})^2+\varepsilon^2}}{2}$$

$$V = \sum_{i,j}^{N} V_{ij}$$

R_{ij} is the intermediate-state distance for residue pairs of i and j, R^0_{1ij} and R^0_{2ij} are the corresponding crystal structure distances of the two alternating structures used for the pathway calculation, k is the spring constant and set to 1 kcal/molÅ^2, and ε is a small number that ensures the smoothening of the transition between the two energy wells. V_{1ij} and V_{2ij} are the harmonic potentials between residues i and j associated with the first and second energy wells, respectively, and V_{ij} is the double-well potential between the two residues whose sum over all included residues yields the total strain energy, V. Two different cutoff criteria are taken into consideration to account for energy strain related,

respectively, to the breaking of native contacts and to the clash of residues that are not in contact in the native states. First, residues that are <8 Å apart in the native state are included in the strain energy calculation to account for possible breakage of native contacts. Second, the interactions between the residues that are >8 Å apart in the native states but which come closer than 4 Å during the pathway calculation are also included in the strain energy to account for residue clashes.

1. Davis IW, et al. (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35:W375–W383.
2. Lovell SC, et al. (2003) Structure validation by C alpha geometry: Phi,psi and C beta deviation. *Proteins* 50:437–450.
3. Maragakis P, Karplus M (2005) Large amplitude conformational change in proteins explored with a plastic network model: Adenylate kinase. *J Mol Biol* 352:807–822.

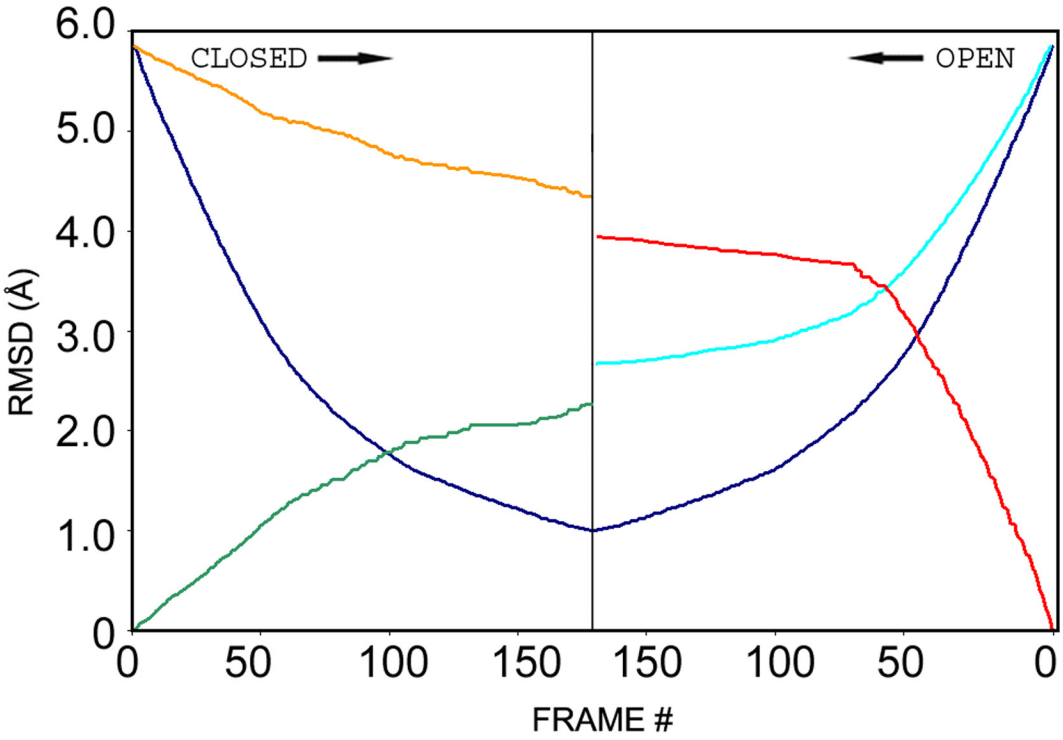


Fig. S1. Structural deviations along ADK transition pathways in VAMM-based computations. As in Fig. 2*B*, deviations (rmsd) are shown between the succession of intermediate states as they move, frame by frame, from the respective closed (*Left*) and open (*Right*) states toward similar terminal intermediates (blue), between the closed state and intermediates generated from the closed state (green), and between the open state and intermediates generated from the open state (red). Here, in addition, deviations (rmsd) are also shown between intermediates generated from the closed state and the initial open state (orange), and between intermediates generated from the open state and the initial closed state (cyan). The open conformation samples states that are closer to the alternative initial state (i.e., closed conformation) compared with the closed state. This is consistent with the fact that the open state undergoes a larger transition during the pathway calculation (red).

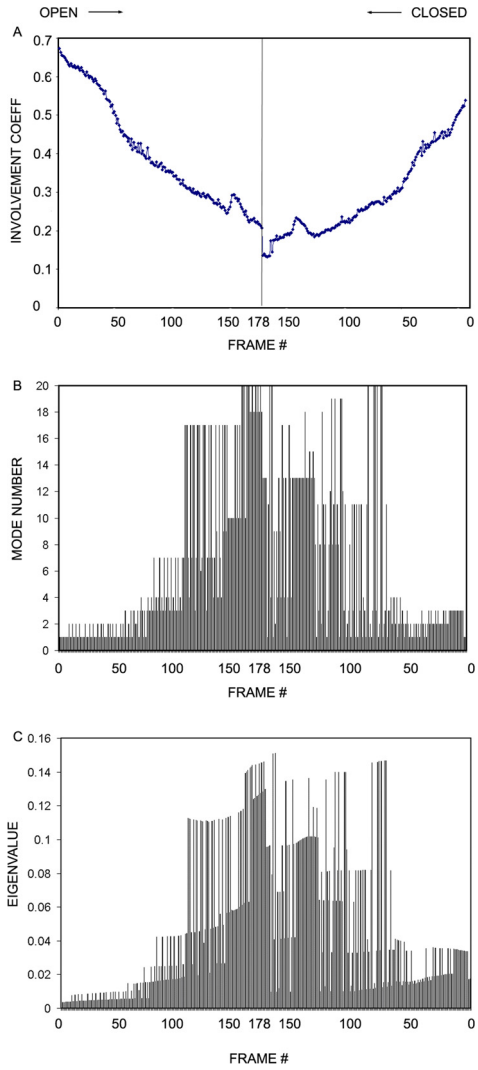


Fig. S2. Evolution of normal-mode-related parameters during the ADK conformational transition computed by the VAMM algorithm. (*A*) Involvement coefficients (Eq. **4**) for normal modes selected at each intermediate frame during the respective transitions from the open state (*Left*) and closed state (*Right*) toward similar terminal intermediates (*Middle*). (*B*) Selected normal modes, numbered from lowest frequency, during the respective ADK conformational transitions from open and closed states as in *A*. (*C*) Eigenvalues of selected normal modes during the ADK conformational transitions from open and closed states as in *A*. At each step during the transition pathway calculation, the normal modes computed from one conformational state that show highest involvement coefficients to their respective target state (i.e., the intermediate state computed from the alternative conformation of ADK) are selected, and this selection is repeated iteratively to reach a common intermediate state in the middle of the trajectory. At the beginning, low-frequency modes have the highest involvement coefficients (i.e., greatest engagement of fluctuations into the direction of the targeted alternative state), whereas later in the calculation, finer adjustments favor higher-frequency modes. Involvement coefficients get smaller as the calculation proceeds. This occurs because finer adjustments are required to converge the pathway to its final state. At the end of the calculation, terminal intermediate states generated from open and closed conformations converge to a similar state, where the only deviations are observed around the loops at the tip of the lid domains for ADK, and such deviations cannot be sampled as efficiently by normal mode analysis as the global deviations observed in the initial stages of the transition.

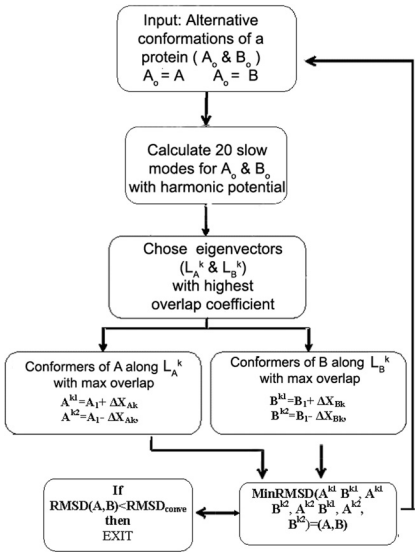


Fig. S3. Control algorithm. The algorithm for calculating the conformational transitions by the elastic network model using simple harmonic or bond-restrained harmonic potentials.

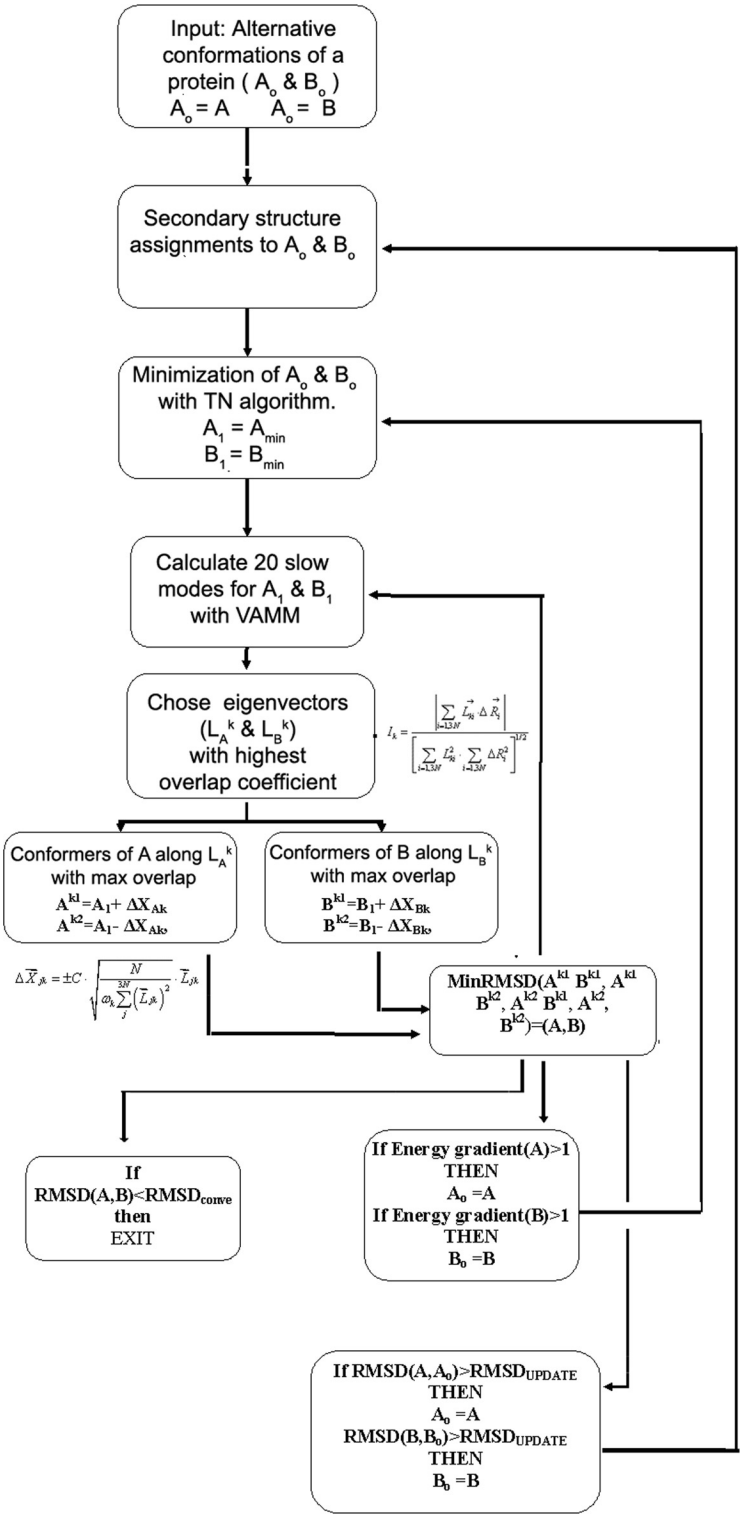
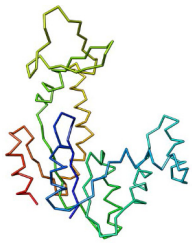


Fig. S4. VAMM algorithm. The algorithm for calculating conformational transitions with iterative normal modes based on the VAMM force field.



Movie S1. VAMM-based conformational transition of adenylate kinase. The transition trajectory is that shown in Fig. 2*B* and in Fig. 6 *A* and *B*. The movie begins at the open (apo) state and moves toward the closed (Ap5A) state along the blue path of Fig. 6*B*. The first part follows the computation from the open state through to the terminal intermediate generated from the open state. There is a discontinuity midway through the movie that corresponds to the juncture between that terminal intermediate and the one generated from the closed state. From that point to the end of the movie, the time course is the reverse of the trajectory computed from the closed state toward the terminal intermediate generated from the closed state.

Movie S1 (MPEG)

Table S1. Virtual bond angle and dihedral angle outliers

	τ vs. θ	θ^- vs. θ^+
Crystal structure	0	3
Minimum structure (VAMM)	0	2
Terminal intermediate state (VAMM)	0	1
Terminal intermediate state (simple harmonic potential)	9	14
Terminal intermediate state (bond-restrained harmonic potential)	24	15

Table S2. The variation of transition pathway parameters

Parameter set	Step size (*C*)	Update interval, Å	No. of modes included	Terminal state deviations, Å, rmsd
1	0.005	0.1	20	0
2	0.001	0.1	20	0.93
3	0.01	0.1	20	0.53
4	0.005	0.1	10	1.37
5	0.005	0.1	50	0.47
6	0.005	0.5	20	0.14
7	0.005	0.05	20	0.08