Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.
This journal is © the Owner Societies 2016

Supplementary data

Enhanced Conformational Sampling Technique Provides Energy Landscape View of Large-Scale Protein Conformational Transitions

Qiang Shao*

Drug Discovery and Design Center, CAS Key Laboratory of Receptor Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai, 201203, China

*To whom correspondence should be addressed. qshao@mail.shcnc.ac.cn

Computational Methods

Evaluation of the Relaxation Time of Protein Conformational Transition. In the present study, we attempt to assess the relaxation time of protein conformational change using Kramers' theory of unimolecular reaction rates in solution.[1-3] Kramers' theory assumes that the dynamics of the reaction can be described by one-dimensional diffusion along a reaction coordinate in which both the reactant well (in the case of protein conformational change, the minimum in the free-energy profile corresponding to the endpoint structure) and barrier top are parabolic (see the free-energy profile of nCaM in Fig. 4B as an example). The relaxation time for protein conformational change is given by:

$$\tau = \frac{2\pi k_B T}{\omega_{\min}\omega_{\max}D_{\max}} \exp\left(\frac{\Delta G^*}{k_B T}\right) \approx \frac{2\pi k_B T}{\omega_{\min}^2 D_{\min}} \exp\left(\frac{\Delta G^*}{k_B T}\right) \quad (1)$$

where $\omega_{\min}$ and $\omega_{\max}$ are frequencies that characterize the curvature of the free-energy profile at the endpoint structure well and (inverted) barrier top, respectively; $D_{\min}$ and $D_{\max}$ are the diffusion constant of protein at the two abovementioned positions; ΔG^* is the height of the free-energy barrier, k_B is Bolzmann's constant and T is the temperature. It has been reported that the approximation of $\omega_{\min} \approx \omega_{\max}$ and $D_{\min} \approx D_{\max}$ could be reasonable for the estimation of the mean folding time for small proteins.[3-5]

ΔG^* and $\omega_{\min}$ can be measured from one-dimensional free-energy profile (e.g., Fig. 4B for nCaM). To measure the diffusion constant $D_{\min}$, non-enhanced conventional MD simulations with explicit solvent were performed. AMBER14 package[6] was used for the simulations with FF14SBonlysc all-atom force field[7] modeling protein and TIP3P model[8] mimicking water, respectively. The functional (closed or open or semi-open) structures of the three proteins under study were solvated in cubic TIP3P water boxes and the simulation systems were neutralized by adding an appropriate number of counterions. Each constructed system was minimized for 5000

steps with the protein fixed using a harmonic restraint (using a force constant of 10 kcal/mol/Å^2 to apply to protein heavy atoms). Subsequently, the system was heated to 300 K in 2 ns and equilibrated at 300 K for 5 ns with a harmonic restraint (force constant = 10 kcal/mol/Å^2) applied to protein heavy atoms. Finally, the equilibrium simulation (production run) for each system was performed without any constraint for 20 ns. The production run was performed at a constant temperature of 300 K and a constant pressure of 1 atm. The integration time step was set to be 2 fs, and the temperature was regulated using Langevin dynamics with the collision frequency of 2 ps^{-1}. All covalent bonds involving hydrogen atoms were fixed by SHAKE algorithm. Periodic boundary conditions were used to avoid edge effects, and the particle mesh Ewald method[9] was applied to treat the long-range electrostatic interaction. The cutoff distance for long-range electrostatic and van der Waals energy terms was set to 10.0 Å.

The time series of the mean square displacement for protein molecule was calculated using *ptraj* command in AMBER software and the diffusion constant was evaluated by

$$\lim \left\langle \left| r(t' + t) - r(t') \right|^2 \right\rangle = 6Dt \qquad (2)$$

The simulation parameters of the explicit-solvent conventional MD simulations for determining the diffusion constant of protein ($D_{\min}$) and all parameters involved in the calculation of the relaxation time (Eq. (1)) are organized in Table S2.

References

1. H. A. Kramers, *Physica*, 1940, **7**, 284-304.
2. B. J. Berne, M. Borkovec and J. E. Straub, *J. Phys. Chem.*, 1988, **92**, 3711-3725.
3. N. D. Socci, J. N. Onuchic and P. G. Wolynes, *J. Chem. Phys.*, 1996, **104**, 5860-5868.
4. L. J. Yang, Q. Shao and Y. Q. Gao, *J. Phys. Chem. B*, 2009, **113**, 803-808.
5. J. Kubelka, J. Hofrichter and W. A. Eaton, *Curr. Opin. Struct. Biol.*, 2004, **14**, 76-88.
6. D.A. Case, V. Babin, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III,

T.A. Darden, R.E. Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu and P.A. Kollman, AMBER 14, University of California, San Francisco, 2014.

7. H. Nguyen, J. Maier, H. Huang, V. Perrone and C. Simmerling, *J. Am. Chem. Soc.*, 2014, **136**, 13959-13962.
8. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
9. T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089-10092.

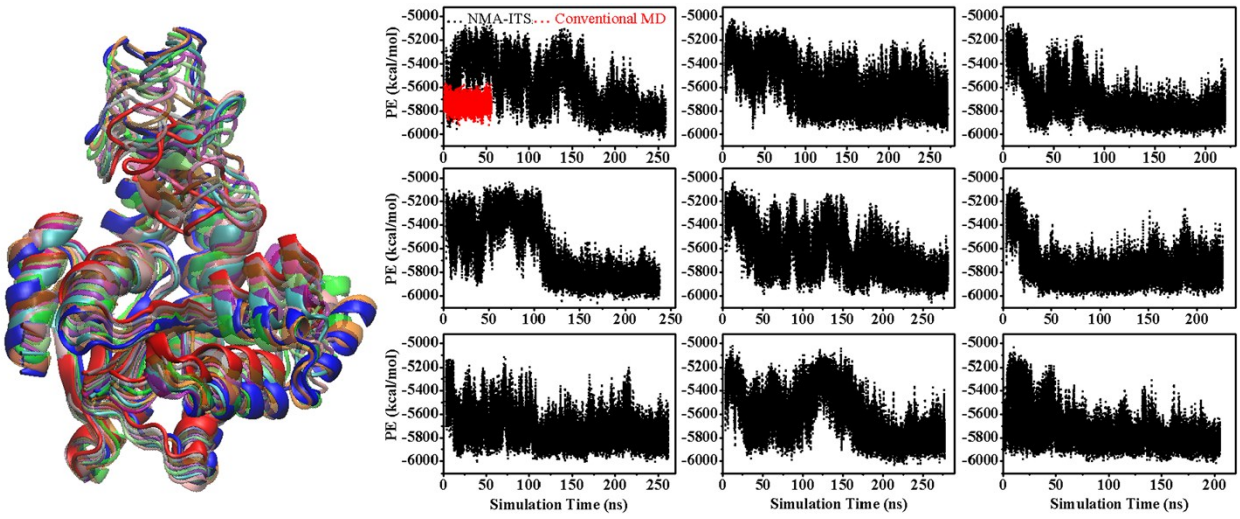

Figure S1. Left panel: nine structures from the closed to open states of AdK selected from the NMA measurement. Right panel: the corresponding nine ITSMD simulation trajectories represented by the time series of potential energies.

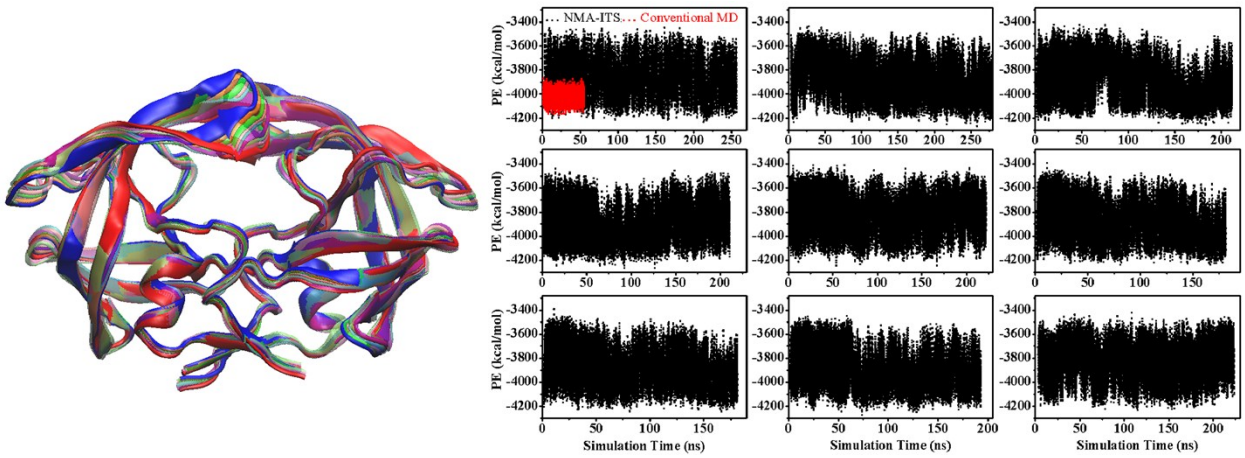

Figure S2. Left panel: nine structures from the closed to wide-open states of HIV-1 PR selected from the NMA measurement. Right panel: the corresponding nine ITSMD simulation trajectories represented by the time series of potential energies.

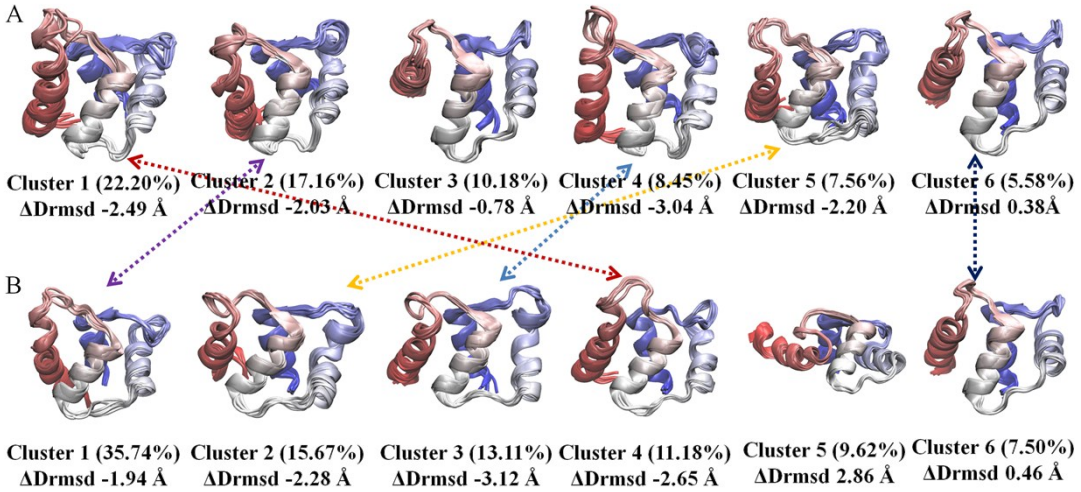

Figure S3. The top six clusters measured in (A) NMA-ITS and (B) explicit-solvent REMD simulations. Under the representative structure of each cluster is shown the cluster order, the average population present from the simulation, and the average value of ΔDrmsd. The common clusters observed in both simulations are connected by dashed arrows.

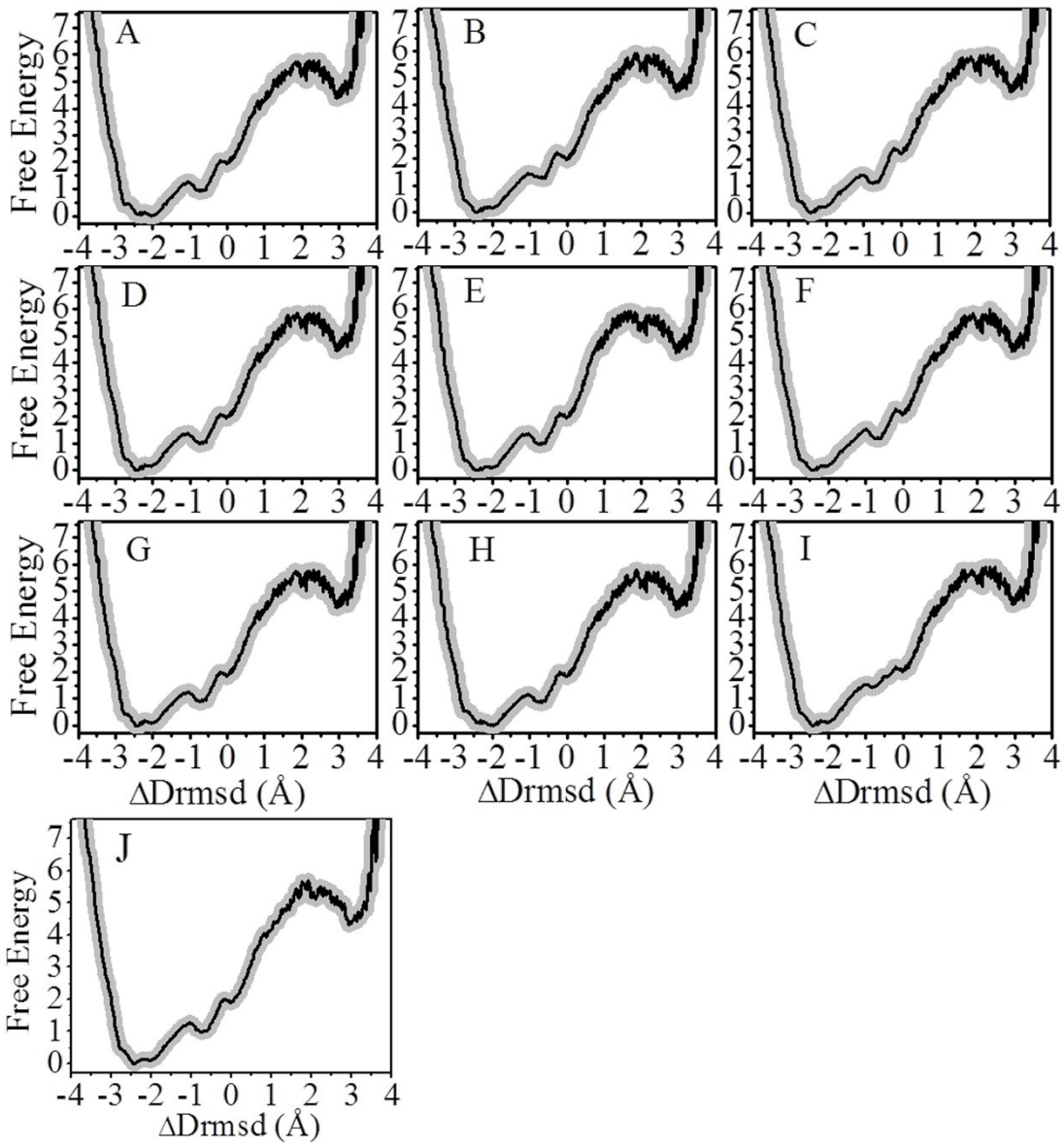

Figure S4. One-dimensional free-energy profile as the function of ΔDrmsd for the conformational change of nCaM (free-energy unit: k_BT). (A-I) The free-energy profiles generated by the data from either combination of eight out of nine ITSMD simulation trajectories. (J) The free-energy profile generated by the data from a total of twelve ITSMD simulation trajectories.

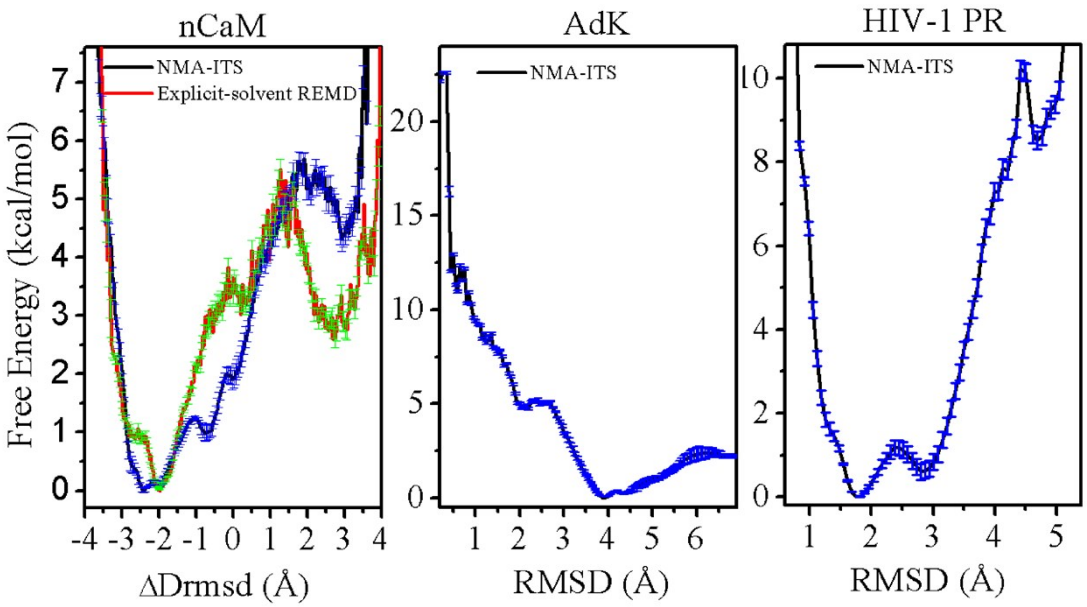

Figure S5. One-dimensional free-energy profiles and corresponding error bars calculated for nCaM, AdK, and HIV-1 PR at 300 K.

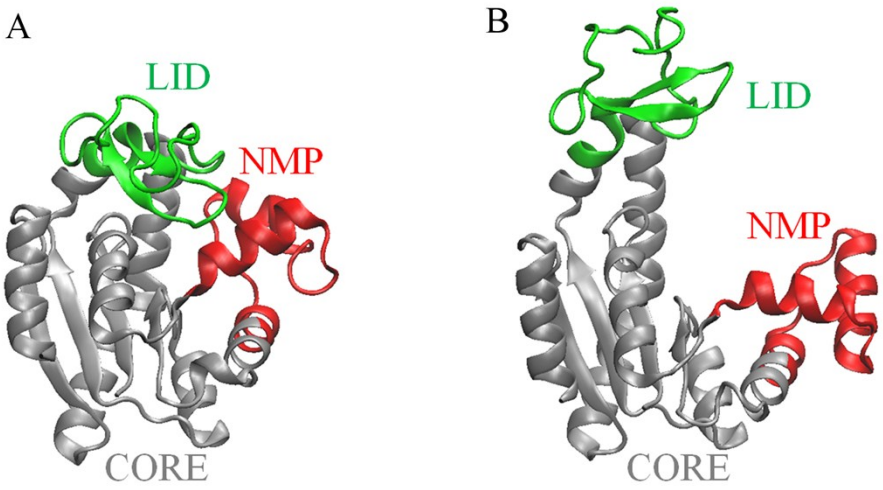

Figure S6. The crystal structures of AdK: (A) closed conformation (PDB code: 1AKE) and (B) open conformation (PDB code: 4AKE). The LID domain is colored in green, NMP domain is colored in red, and CORE domain is colored in silver, respectively.

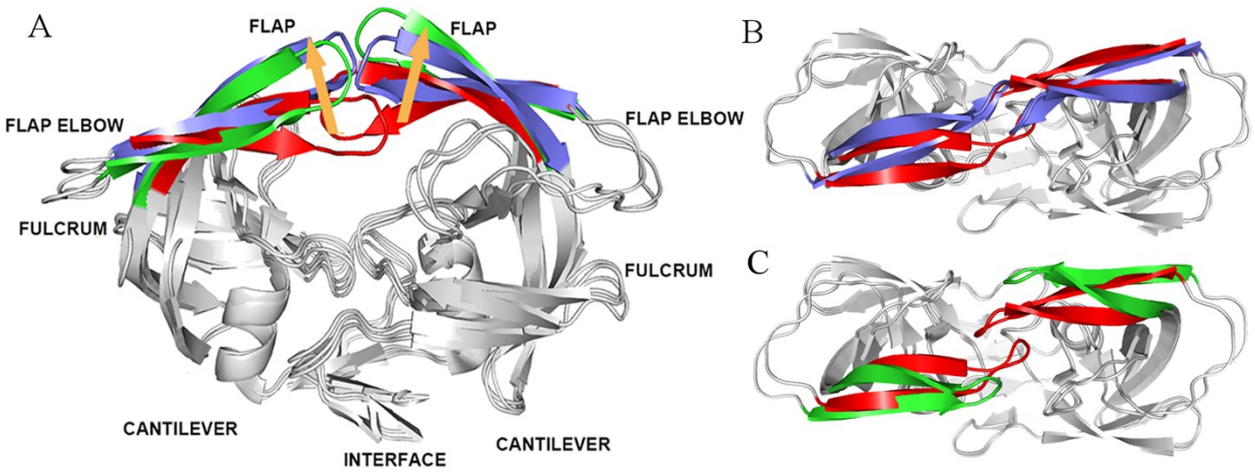

Figure S7. (A) Superposition of the crystal structures of the closed (1T3R), semi-open (1HHP), and wide-open (1TW7) forms of HIV-1 PR (arrows indicate the upward direction of changes during the flap opening). (B-C) Top view of the closed form (red flaps) superposed onto the semi-open form (blue flaps) and the wide-open form (green flaps), respectively. B and C indicate the relative orientation of the two flaps (the handedness) in the lateral direction in various structures of HIV-1 PR.

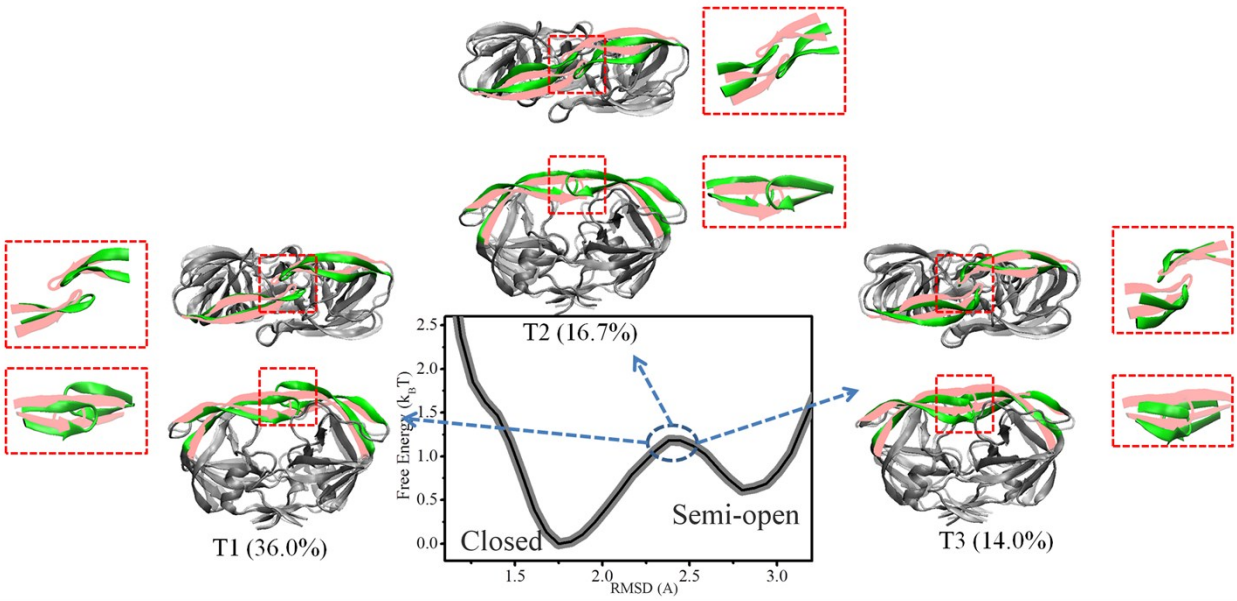

Figure S8. Superposition of the three popular conformations (T1-T3) of the transition state (green flaps) onto the closed state (red flaps) of HIV-1 PR. Upper: top view; lower: side view. Expanded view of the flap regions is presented for all three conformations to show the detailed difference in flap orientations. The percentages of the three conformations in the structure ensemble of the transition state are also provided.

States	Flap-Flap Interaction Pairs				Flap-80s (80's) Loop Interaction Pairs		
	50-50'	50-47' (47-50')	50-53' (53-50')	54-50' (50-54')	79-50' (50-79')	81-50' (50-81')	84-50' (50-84')
Closed	74.3%	83.2(95.0)%	6.6(5.6)%	92.3(90.0)%	6.5(10.5)%	20.2(56.7)%	94.5(80.7)%
Semi-open	67.5%	0.32(4.5)%	38.8(28.6)%	2.8(5.0)%	0.8(1.1)%	0.4(0.04)%	12.2(0.5)%
Curled-in	6.2%	0.0(5.4)%	60.4(5.7)%	1.8(9.1)%	0.6(0.0)%	0.1(0.0)%	0.4(0.0)%
Fully open	0.0%	0.28 (0.0)%	0.0(0.0)%	0.0(0.0)%	0.0(0.0)%	0.0(0.0)%	0.0(0.0)%

Table S1. Survival probability of individual hydrophobic interactions among flaps and 80s (80's) loops of HIV-1 PR in important states.

Protein Transition	ω_{min}	D_{min} ($10^{-12}m^2/s$)	ΔG^* (k_BT)	Simulation System	Simulation Box Size ($Å^3$)	Simulation Length (ns)
nCaM (Closed → Open)	0.186	5.65	5.5	Closed State of NCaM in water	58.3×58.5×58.0	20
HIV-1 PR Closed → Semi-open	1.70	2.26	1.2	Closed State of HIV-1 PR in water	82.0×64.7×65.7	20
HIV-1 PR Semi-open→ Fully open	2.65	2.62	9.1	Semi-open State of HIV-1 PR in water	60.9×61.4×81.1	20
AdK Lid Opening	0.093	2.83	1.5	Closed State of AdK in water	60.3×62.2×65.0	20
AdK Lid Closing	0.022	3.00	3.0	Open State of AdK in water	60.3×73.4×73.7	20

Table S2. Parameter determination for the evaluation of the relaxation time associated with the conformational transitions under study. Left: the values of the parameters used in Eq. (1); right: simulation parameters of the explicit-solvent conventional MD simulations for the determination of diffusion constants.