Supporting information: NMR relaxation in proteins with fast internal motions and slow conformational exchange: model free framework and Markov state simulations

Junchao Xia, Nan-jie Deng, and Ronald M. Levy*

Department of Chemistry and Chemical Biology and BioMaPS Institute for Quantitative Biology, Rutgers, the State University of New Jersey, 610 Taylor Road, Piscataway, NJ 08854

E-mail: ronlevy@lutece.rutgers.edu

*To whom correspondence should be addressed

1 Kinetic Network Model and Stochastic Simulation

1.1 Molecular Dynamics Simulation

Twenty MD simulations of HIV-1 PR dimer each lasting 20 ns were performed using the molecular simulation package IMPACT,[1] the OPLS-AA force field (2005),[2] and the AGBNP implicit solvent model.[3,4] All simulations started at the same temperature of 285K, the same initial conformation from the crystal structure (PDB: 1HHP) but different seeds of random numbers for the velocities. More details about the parameter settings and simulations can be found in our previous work.[5]

The equilibration stage of all simulations lasted 450 ps with a time step of 1.5 fs and gradually releasing of distance restraints. The subsequent 20 $\times$ 20 ns production runs were obtained with the time step of 1.0 fs. For each production simulation, the conformational snapshots were saved every 1 ps and $20 \times 20000 = 400000$ snapshots were stored in total.

1.2 Building the Markov State Kinetic Network Model

To build the network, we chose 200,000 snapshots (every 2 ps from all 20 $\times$ 20 ns trajectories) and calculated the C-α RMSD matrix of the two flaps (residue 43 to 58, and 142 to 157) between each pair of snapshots which requires storage of 4×10^{10} matrix elements. The snapshots were clustered into 82,291 conformational nodes using an RMSD criteria of 0.5 $\mathring{A}$. Namely two snapshots are grouped into the same node if their RMSD is smaller than 0.5 $\mathring{A}$. Then a network of conformational nodes was created by connecting all pairs of clustered nodes which have the node RMSDs $<$ 1.2 $\mathring{A}$.

In the kinetic network model the time evolution of probability $P(t)$ for each network node obeys the master equation[6,7]

$$\frac{dP_i(t)}{dt} = \sum_j k_{ij} P_j(t) - \sum_j k_{ji} P_i(t) \tag{1}$$

where k_{ij} is the rate constant for transition from state j to state i defined by

$$k_{ij} = C_{ij}(\frac{P_{i,eq}}{P_{j,eq}})^{1/2}. \tag{2}$$

The k_{ij} satisfy detailed balance $k_{ij}P_{j,eq} = k_{ji}P_{i,eq}$, and $C_{ij} = C_{ji}$. The rate k_{ij} can be expressed in terms of the branching probability $P_{j\to i}$ and the mean lifetime at node j, T_j,

$$k_{ij} = \frac{P_{j\to i}}{T_j}. \tag{3}$$

Eq. 3 suggests a way to parametrize k_{ij} based on the lifetimes estimated from many short MD trajectories. The branching probability $P_{j\to i}$ depends approximately on the RMSD between neighboring nodes i and j, Δr_{ij}. From running many short MD trajectories, we found that[8] on average the probability of jumping to a neighboring node at Δr_{ij} can be fitted with $P_{j\to i} \propto \frac{\Delta r_{ij}^{-6}}{<\Delta r_{ij}^{-6}>_j}$. Additionally, $P_{j\to i}$ decreases approximately linearly with the number of neighboring nodes of node j, i.e. $P_{j\to i} \propto \frac{1}{N_j^{Nb}}$. Taking these considerations together, we have

$$P_{j\to i} \propto \frac{\Delta r_{ij}^{-6}}{N_j^{Nb}}. \tag{4}$$

Hence the k_{ij} is approximated by

$$k_{ij} = C_{ij}(\frac{P_{i,eq}}{P_{j,eq}})^{1/2} = \frac{\Delta r_{ij}^{-6}}{<\Delta r_{ij}^{-6}>} \frac{1}{N_j^{Nb}} \frac{1}{T_{j,MD}} (\frac{P_{i,eq}}{P_{j,eq}})^{1/2}. \tag{5}$$

Since $C_{ij} = C_{ji}$, to enforce the detailed balance, we symmetrize the C_{ij} as $C_{ij} = C_{ji} \approx \frac{1}{2}(\frac{\Delta r_{ij}^{-6}}{<\Delta r_{ij}^{-6}>_j N_j^{Nb} T_{j,MD}} + \frac{\Delta r_{ji}^{-6}}{<\Delta r_{ji}^{-6}>_i N_i^{Nb} T_{i,MD}})$. Finally we obtain

$$k_{ij} \approx \frac{\Delta r_{ij}^{-6}}{2}(\frac{1}{<\Delta r_{ij}^{-6}>_j N_j^{Nb} T_{j,MD}} + \frac{1}{<\Delta r_{ji}^{-6}>_i N_i^{Nb} T_{i,MD}})(\frac{P_{i,eq}}{P_{j,eq}})^{1/2}. \tag{6}$$

The node lifetime $T_{i,MD}$ can be estimated by performing short MD simulations starting each node.[8] In this paper, we built a very high resolution network and found $T_{i,MD}$ for several representative nodes located in different macrostates are similar, with $< (T_{i,MD} + T_{j,MD})/2 >= 4.8$ ps with a variance of 0.5 ps, which reproduces the average short time behavior of the motion in the MD simulation.

1.3 Stochastic Simulations on the Network

Stochastic simulations on the kinetic network were implemented using the Gillespie algorithm[9–11] as described previously.[5,12,13] Briefly, the simulation starts from a node selected randomly and jumps to neighboring nodes for some waiting time. The waiting time at a given node i is an exponential random variable whose mean equals the inverse of the sum of the exiting rates from that node to neighboring nodes. The probability that the system subsequently lands on a connected node j is proportional to the microscopic rate from node i to j. Such random walks on the network satisfy the master equation 1.

To our best knowledge, this is the first time to calculate rotational correlation functions from a kinetic network trajectory. In the KMC trajectory, the snapshot representing a node and the waiting time on this node were saved once a transition to a new node occurs. Because the transition time from node to node is a random variable according to the Gilliespie algorithm, the snapshots are not evenly distributed on a time series from our network simulation. (There are a total of 3,123,612 snapshots in our 10 microseconds trajectory.) Instead, to compute the correlation functions, we need to use a bin width with a constant time interval (5ps) to remap the time series, namely we assigned the saved snapshots to each bin with the closest time to that from the Gilliespie simulation.

2 NMR Relaxation Theory

2.1 General Theory

The rotational time correlation function of a nuclear spin-spin vector is defined as

$$C_m^{(l)}(t) = < D_{m0}^{(l)}(\Omega_{LF}^t) D_{m0}^{(l)*}(\Omega_{LF}^0) > = \frac{4\pi}{2l+1} < Y_m^{(l)*}(\Omega_{LF}^t) Y_m^{(l)}(\Omega_{LF}^0) >, \quad (7)$$

where $Y_m^{(l)}$ is the spherical harmonics function, and $D_{mn}^{(l)}$ is the Wigner rotation D-matrix element. $\Omega_{LF}^t = (\alpha(t), \beta(t), \gamma(t))$ are Euler angles specifying the time-dependent orientation of the unit vector connecting the two nuclei in the laboratory coordinate system. The brackets $< \cdots >$ represent the ensemble average over the initial and final states at time 0 and t.

When the system tumbles in an isotropic solution, the correlation function does not depend on the index m and can be rewritten as

$$C^{(l)}(t) = \frac{1}{5} \sum_{m=-l}^{m=l} C_m^{(l)}(t) = \frac{1}{5} \cdot \frac{4\pi}{2l+1} \sum_{m=-l}^{m=l} < Y_m^{(l)*}(\Omega_{LF}^t) Y_m^{(l)}(\Omega_{LF}^0) > = \frac{1}{5} < P^{(l)}(\vec{n}(t) \cdot \vec{n}(0)) >, \quad (8)$$

where $P^{(l)}$ is the Legendre polynomials of order l. $\vec{n}(t)$ and $\vec{n}(0)$ represent the spin-spin unit vector at time 0 and t. To describe the relaxation in nuclear magnetic resonance, $l = 2$ is sufficient.

Interpretation of NMR relaxations usually requires simplified structure models, from the early single rigid rotor,[14] to the anisotropic rotor plus internal motions,[15–17] to the model-free approach including the overall rotation part and the fast internal motion represented by a generalized order parameter and an exponential decay with a characteristic time,[18–20] and to the more recent conformational exchange model between states with different overall rotation diffusion tensors.[21] Those simplifications vary from model to model and largely depend on the problem being studied, but they are based on a common fact that Ω_{LF}^t, the time-dependent orientation of a spin-spin vector in the laboratory coordinates frame can be decomposed into an overall rotation Ω_{LP}^t in the principal frame, a series of Euler angles (Ω_{P1}, Ω_{12}, Ω_{23}, $\cdots$, Ω_{ii+1}, $\cdots$, and Ω_{N-1N}) in the N successive internal rotation frames, and Ω_{NF}^t describing the relative motions of spin vector in the N frame.

Correspondingly, the Wigner D-matrix can be described by

$$D^{(2)}_{q0}(\Omega_{LF}) = \sum_{a=-2}^{2}\sum_{b_1=-2}^{2}\sum_{b_2=-2}^{2}\cdots\sum_{b_i=-2}^{2}\cdots\sum_{b_N=-2}^{2} D^{(2)}_{qa}(\Omega_{LP})D^{(2)}_{ab_1}(\Omega_{P1})D^{(2)}_{b_1b_2}(\Omega_{12})\cdots D^{(2)}_{b_{i-1}b_i}(\Omega_{i-1i})\cdots D^{(2)}_{b_{N-1}b_N}(\Omega_{N-1N})D^{(2)}_{b_N0}(\Omega_{NF}). \quad (9)$$

For many cases, the overall rotational motions are assumed to be independent of the internal motions and they can be separated. Furthermore, a common approximation is that there are also no correlations among any internal motions. Hence the correlation function of Eq. 7 is simplified as[22] as

$$\begin{aligned} C^{(2)}_q(t) &= < D^{(2)}_{q0}(\Omega^t_{LF})D^{(2)*}_{q0}(\Omega^0_{LF}) > \\ &= \sum_{a=-2}^{2}\sum_{a'=-2}^{2} < D^{(2)}_{qa}(\Omega^t_{LP})D^{(2)*}_{qa'}(\Omega^0_{LP}) > \sum_{b_1=-2}^{2}\sum_{b'_1=-2}^{2} < D^{(2)}_{ab_1}(\Omega^t_{P1})D^{(2)*}_{a'b'_1}(\Omega^0_{P1}) > \\ &\sum_{b_2=-2}^{2}\sum_{b'_2=-2}^{2} < D^{(2)}_{b_1b_2}(\Omega^t_{12})D^{(2)*}_{b'_1b'_2}(\Omega^0_{12}) > \cdots \sum_{b_i=-2}^{2}\sum_{b'_i=-2}^{2} < D^{(2)}_{b_{i-1}b_i}(\Omega^t_{i-1i})D^{(2)*}_{b'_{i-1}b'_i}(\Omega^0_{i-1i}) > \cdots \\ &\sum_{b_N=-2}^{2}\sum_{b'_N=-2}^{2} < D^{(2)}_{b_{N-1}b_N}(\Omega^t_{N-1N})D^{(2)*}_{b'_{N-1}b'_N}(\Omega^0_{N-1N}) >< D^{(2)}_{b_N0}(\Omega^t_{NF})D^{(2)*}_{b'_N0}(\Omega^0_{NF}) > . \end{aligned} \quad (10)$$

2.2 Model-Free Approach

The model-free approach[18,19] assumes that the reorientational time correlation function of a spin-spin vector can be separated into a slow overall rotation part and a fast internal one that is independent of the former. The overall rotation part generally describes the rotational motions of a rigid rotor using the principal coordinate frame of the molecule or the diffusion tensor. The internal part depicts the relative motions of a spin-spin vector in the local coordinate frame specified for each spin-spin vector and is described by a generalized order parameter S^2 and an exponential decay with a constant of τ_e . The transformation from the principal coordinate frame to the local coordinate frame is assumed to be independent of time. Hence, Ω^t_{LF} in Eq. 7 are decomposed into three rotational transformations: Ω^t_{LP} describes the motion of the principal frame in the laboratory

frame, Ω_{PD} represents the time-independent orientation of local coordinate frame in the principal frame, and Ω^t_{DF} describes the orientation of the spin-spin vector in the local coordinate frame. Due to these assumptions, the model-free approach is valid when the overall rotations are much slower than the internal motions ($\tau_M \gg \tau_e$), and there are no large conformation changes modifying the global shape of the molecule and the diffusion tensor.

For the model-free approach, the correlation function Eq. 7 is reduced to

$$\begin{aligned} C_q^{(2)}(t) &= < D_{q0}^{(2)}(\Omega^t_{LF}) D_{q0}^{(2)*}(\Omega^0_{LF}) > \\ &=< \sum_{m=-2}^{2} \sum_{n=-2}^{2} D_{qm}^{(2)}(\Omega^t_{LP}) D_{qn}^{(2)*}(\Omega^0_{LP}) > \\ &\sum_{k=-2}^{2} \sum_{l=-2}^{2} D_{mk}^{(2)}(\Omega_{PD}) D_{nl}^{(2)*}(\Omega_{PD}) < D_{k0}^{(2)}(\Omega^t_{DF}) D_{l0}^{(2)*}(\Omega^0_{DF}) >, \end{aligned} \tag{11}$$

where $< D_{k0}^{(2)}(\Omega^t_{DF}) D_{l0}^{(2)*}(\Omega^0_{DF}) >$ describes the internal motions of a spin vector in the local coordinate frame. For the model-free approach there is no dependence on k or $l \neq 0$ and the internal motion reduces to:

$$C_I(t) =< D_{00}^{(2)}(\Omega^t_{DF}) D_{00}^{(2)*}(\Omega^0_{DF}) >= S^2 + (1 - S^2) e^{-t/\tau_e}. \tag{12}$$

S^2 is called the square of the generalized parameter S and can be calculated by

$$\begin{aligned} S^2 = C_I(\infty) &=< D_{00}^{(2)}(\Omega^\infty_{DF}) D_{00}^{(2)*}(\Omega^0_{DF}) > \\ &= \quad < D_{00}^{(2)}(\Omega^0_{DF}) >< D_{00}^{(2)*}(\Omega^0_{DF}) >=< P^{(2)}(\cos\theta) >< P^{(2)*}(\cos\theta) >, \\ S &= < D_{00}^{(2)}(\Omega^0_{DF}) >=< P^{(2)}(\cos\theta) >, \end{aligned} \tag{13}$$

where θ is the angle between the spin-spin vector and the Z axis of the local coordinates frame if the internal motion is azimuthally symmetric.[18]

The total correlation function in an isotropic solution will become

$$C^{(2)}(t) = \frac{1}{5}\sum_{q=-2}^{q=2} C_q^{(2)}(t) = C_O(t)C_I(t), \tag{14}$$

$$C_O(t) = \frac{1}{5}\sum_{q=-2}^{q=2}\sum_{m=-2}^{2}\sum_{n=-2}^{2} < D_{qm}^{(2)}(\Omega_{LP}^{t}) D_{qn}^{(2)*}(\Omega_{LP}^{0}) > D_{m0}^{(2)}(\Omega_{PD}) D_{n0}^{(2)*}(\Omega_{PD}). \tag{15}$$

The overall rotation function $C_O(t)$ depends on the global shape of molecule and can be evaluated by finding the Green's function (conditional probability) from the free-diffusion equation for a rigid-body rotor[14–17,23]

$$\frac{\partial}{\partial t}P(\Omega_{LP},t) = -\sum_{i,j=1}^{3} L_i D_{ij} L_j P(\Omega_{LP},t), \tag{16}$$

where L_i represents the angular momentum operator about the ith axis and D_{ij} denotes the ijth component of the second-order rotational diffusion tensor for the molecule. Subjected to the symmetry of the molecule, the solution of Eq. 16 and the correlation function $C_O(t)$ have been discussed in several places.[15,23]

The model-free approach assumes that the internal motions are well represented by one correlation time and has been applied to account for relaxation data on many small proteins and simple polymers with less flexibility. For many complex biomolecules, the time evolution of the internal correlation function will be multi-exponential or highly nonexponential with slow components. The extended model-free approach[24] introduced two distinct correlation times to describe the internal motions.

$$C_I(t) = S^2 + A_f e^{-t/\tau_f} + A_s e^{-t/\tau_s} = S^2 + (1-S_f^2)e^{-t/\tau_f} + (S_f^2 - S^2)e^{-t/\tau_s}, \tag{17}$$

where $S^2 + A_f + A_s = 1$. τ_f and τ_s describe the fast and slow decay components respectively. When τ_s and τ_f differ by at least 1 order of magnitude, $C_I(t)$ reaches an intermediate plateau value of $1 - A_f = S_f^2$, besides $C_I(\infty) = S^2$.

2.3 Conformational Exchange between Two Macrostates

The model-free approach[18,19] is quite successful for applications to molecules with well-defined mostly rigid structures and fast local motions. However, for many cases, such as HIV-1 Protease, the dynamics often involves slower conformational changes including loop motions and domain reorientations. For example, there exist two well-defined macrostates with substantial populations for HIV-1 protease: the semiopen and the closed states. Here we derive the analytical forms for the correlation functions and the effective order parameter S^2 and the relaxation time τ_e for the two macrostates with independent internal motions specified by S_A^2, S_B^2, τ_{eA}, and τ_{eB} but sharing the same diffusion tensor as shown in Fig. SS1.

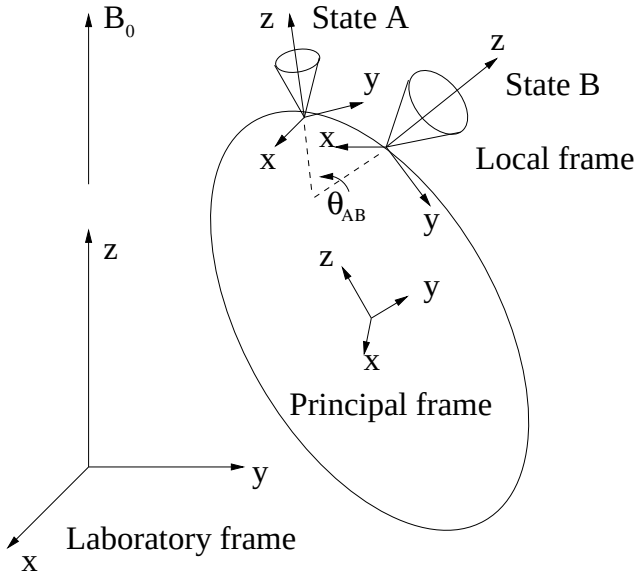


Figure S1: Two-macrostate conformational exchange model. The macromolecule has two distinct conformational states A and B which share the same principal coordinate frame (diffusion tensor).

If conformational exchange is much slower than overall rotations and internal motions, ($\tau_{EX} \gg \tau_M \gg \tau_e$), the cross terms between two different states are very small and can be neglected when evaluating the correlation function using conditional probability. Hence the total correlation function in Eq. 7 is reduced to

$$C_T(t) = C_O(t)C_I(t) = P_A C_A(t) + P_B C_B(t). \tag{18}$$

When the two states share the same diffusion tensor it is easy to show that

$$C_I(t) = P_A C_{IA}(t) + P_B C_{IB}(t) = P_A[S_A^2 + (1-S_A^2)e^{-t/\tau_{eA}}] + P_B[S_B^2 + (1-S_B^2)e^{-t/\tau_{eB}}]$$

$$C_I(t) = S_T^2 + P_A(1-S_A^2)e^{-t/\tau_{eA}} + P_B(1-S_B^2)e^{-t/\tau_{eB}} \tag{19}$$

$$S_T^2 = C_I(\infty) = P_A S_A^2 + P_B S_B^2. \tag{20}$$

For the alternative case as shown in Fig.S1, namely the overall rotation is much slower than conformation exchange and internal motions of both states ($\tau_M \gg \tau_{EX} \gg \tau_e$), the correlation function Eq. 7 can be written as

$$\begin{aligned} C_q^{(2)}(t) &= < D_{q0}^{(2)}(\Omega_{LF}^t) D_{q0}^{(2)*}(\Omega_{LF}^0) > \\ &= \sum_{m=-2}^{2} \sum_{n=-2}^{2} < D_{qm}^{(2)}(\Omega_{LP}^t) D_{qn}^{(2)*}(\Omega_{LP}^0) >< D_{m0}^{(2)}(\Omega_{PF}^t) D_{n0}^{(2)*}(\Omega_{PF}^0) > \qquad (21) \\ &= \sum_{m=-2}^{2} \sum_{n=-2}^{2} C_{qmqn}^{(2)LP}(t) C_{m0n0}^{(2)PF}(t), \qquad (22) \end{aligned}$$

Where $C_{qmqn}^{(2)LP}(t) =< D_{qm}^{(2)}(\Omega_{LP}^t) D_{qn}^{(2)*}(\Omega_{LP}^0) >$ describes the overall rotation part and $C_{m0n0}^{(2)PF}(t) =< D_{m0}^{(2)}(\Omega_{PF}^t) D_{n0}^{(2)*}(\Omega_{PF}^0) >$ corresponds to the internal motion part for conformational exchange between and local internal motions within the macrostates. Since the local internal motions are very fast and they can be assumed to only depend on conformation states within each macrostate, the internal part of Eq. 22 can be described as

$$\begin{aligned} C_{m0n0}^{(2)PF}(t) &= < D_{m0}^{(2)}(\Omega_{PF}^t) D_{n0}^{(2)*}(\Omega_{PF}^0) > \\ &=< \sum_{\Omega_{PF}^t = \Omega_{PF}^A, \Omega_{PF}^B} \sum_{\Omega_{PF}^0 = \Omega_{PF}^A, \Omega_{PF}^B} D_{m0}^{(2)}(\Omega_{PF}^t) D_{n0}^{(2)*}(\Omega_{PF}^0) P(\Omega_{PF}^t, t | \Omega_{PF}^0, 0) P_{eq}(\Omega_{PF}^0) >_{Local} \\ &=< \sum_{\beta=A,B} \sum_{\alpha=A,B} D_{m0}^{(2)}(\beta, t) D_{n0}^{(2)*}(\alpha, 0) P(\beta, t | \alpha, 0) P_{eq}(\alpha) >_{Local}, \qquad (23) \end{aligned}$$

where $< \cdots >_{Local}$ represents the ensemble average in the local coordinates frame for each state.

The conditional probability in Eq. 23 can be solved from the master equations as below,

$$\begin{aligned}\frac{dP(A,t|A,0)}{dt} &= -\frac{1}{\tau_{AB}}P(A,t|A,0)+\frac{1}{\tau_{BA}}P(B,t|A,0)\\ \frac{dP(B,t|B,0)}{dt} &= +\frac{1}{\tau_{AB}}P(A,t|B,0)-\frac{1}{\tau_{BA}}P(B,t|B,0)\end{aligned} \tag{24}$$

with the initial conditions at $t \to 0$ and normalized conditions at any t.

$$\begin{aligned}&P(A,t|A,0)|_{t\to 0}=1, \qquad P(A,t|A,0)+P(B,t|A,0)=1\\ &P(B,t|B,0)|_{t\to 0}=1, \qquad P(A,t|B,0)+P(B,t|B,0)=1\end{aligned} \tag{25}$$

The solutions are:

$$\begin{aligned}P(A,t|A,0) &= \frac{\tau_{AB}}{\tau_{BA}+\tau_{AB}}+\frac{\tau_{BA}}{\tau_{BA}+\tau_{AB}}e^{-t/\tau_{EX}} = P_A+P_Be^{-t/\tau_{EX}}\\ P(B,t|A,0) &= \frac{\tau_{BA}}{\tau_{BA}+\tau_{AB}}(1-e^{-t/\tau_{EX}}) = P_B(1-e^{-t/\tau_{EX}})\\ P(A,t|B,0) &= \frac{\tau_{AB}}{\tau_{BA}+\tau_{AB}}(1-e^{-t/\tau_{EX}}) = P_A(1-e^{-t/\tau_{EX}})\\ P(B,t|B,0) &= \frac{\tau_{BA}}{\tau_{BA}+\tau_{AB}}+\frac{\tau_{AB}}{\tau_{BA}+\tau_{AB}}e^{-t/\tau_{EX}} = P_B+P_Ae^{-t/\tau_{EX}}\end{aligned} \tag{26}$$

where

$$\begin{aligned}P_A = P_{eq}(A) = P(A,t|A,0)|_{t\to\infty} = P(A,t|B,0)|_{t\to\infty} &= \frac{\tau_{AB}}{\tau_{BA}+\tau_{AB}}\\ P_B = P_{eq}(B) = P(B,t|A,0)|_{t\to\infty} = P(B,t|B,0)|_{t\to\infty} &= \frac{\tau_{BA}}{\tau_{BA}+\tau_{AB}}\\ \frac{1}{\tau_{EX}} = k_{AB}+k_{BA} &= \frac{1}{\tau_{AB}}+\frac{1}{\tau_{BA}}\end{aligned} \tag{27}$$

Substituting the solutions of Eq. 26 to 27 into the correlation function Eq. 23 we obtain that

$$
\begin{aligned}
C^{(2)PF}_{m0n0}(t) &= < D^{(2)}_{m0}(\Omega^{t}_{PF}) D^{(2)*}_{n0}(\Omega^{0}_{PF}) > \\
&= < D^{(2)}_{m0}(A,t) D^{(2)*}_{n0}(A,0) P(A,t|A,0) P_A + D^{(2)}_{m0}(A,t) D^{(2)*}_{n0}(B,0) P(A,t|B,0) P_B \\
&+ D^{(2)}_{m0}(B,t) D^{(2)*}_{n0}(A,0) P(B,t|A,0) P_A + D^{(2)}_{m0}(B,t) D^{(2)*}_{n0}(B,0) P(B,t|B,0) P_B >_{Local} \\
&= < D^{(2)}_{m0}(A,t) D^{(2)*}_{n0}(A,0) (P_A + P_B e^{-t/\tau_{EX}}) P_A + D^{(2)}_{m0}(A,t) D^{(2)*}_{n0}(B,0) P_A (1 - e^{-t/\tau_{EX}}) P_B \\
&+ D^{(2)}_{m0}(B,t) D^{(2)*}_{n0}(A,0) P_B (1 - e^{-t/\tau_{EX}}) P_A + D^{(2)}_{m0}(B,t) D^{(2)*}_{n0}(B,0) (P_B + P_A e^{-t/\tau_{EX}}) P_B >_{Local} \\
&= S^2_{m0n0} + [F_{m0n0} - S^2_{m0n0}] e^{-t/\tau_{EX}}, \qquad (28)
\end{aligned}
$$

where

$$
\begin{aligned}
S^2_{m0n0} = \quad & < [P_A D^{(2)}_{m0}(A,t) + P_B D^{(2)}_{m0}(B,t)][P_A D^{(2)*}_{n0}(A,0) + P_B D^{(2)*}_{n0}(B,0)] >_{Local}, \qquad (29) \\
F_{m0n0} = \quad & < P_A D^{(2)}_{m0}(A,t) D^{(2)*}_{n0}(A,0) + P_B D^{(2)}_{m0}(B,t) D^{(2)*}_{n0}(B,0) >_{Local} . \qquad (30)
\end{aligned}
$$

For the model-free approach, there is no m or n dependence and $m = 0$ and $n = 0$. The internal correlation function can be written:

$$
C_I(t) = C^{(2)PF}_{0000}(t) = S^2_{0000} + (F_{0000} - S^2_{0000}) e^{-t/\tau_{EX}}, \qquad (31)
$$

where

$$
\begin{aligned}
S^2_{0000} = & \\
&= < [P_A D^{(2)}_{00}(A,t) + P_B D^{(2)}_{00}(B,t)][P_A D^{(2)*}_{00}(A,0) + P_B D^{(2)*}_{00}(B,0)] >_{Local} \\
&= P^2_A < D^{(2)}_{00}(A,t) D^{(2)*}_{00}(A,0) >_{Local} + P_A P_B < D^{(2)}_{00}(A,t) D^{(2)*}_{00}(B,0) >_{Local} \\
&+ P_B P_A < D^{(2)}_{00}(B,t) D^{(2)*}_{00}(A,0) >_{Local} + P^2_B < D^{(2)}_{00}(B,t) D^{(2)*}_{00}(B,0) >_{Local} . \qquad (32)
\end{aligned}
$$

From the model-free approach in the last section, we have

$$< D_{00}^{(2)}(A,t)D_{00}^{(2)*}(A,0) >_{Local} = S_A^2 + (1-S_A^2)e^{-t/\tau_{eA}},$$
$$< D_{00}^{(2)}(B,t)D_{00}^{(2)*}(B,0) >_{Local} = S_B^2 + (1-S_B^2)e^{-t/\tau_{eB}}. \quad (33)$$

To evaluate the cross terms $< D_{00}^{(2)}(A,t)D_{00}^{(2)*}(B,0) >_{Local}$ and $< D_{00}^{(2)}(B,t)D_{00}^{(2)*}(A,0) >_{Local}$, we need to decompose the Ω_{PF}^{t} of each state into two components: one is the time-independent term Ω_{PF}^{0} describing the equilibrium position, and the other is the time-dependent term $\Omega_{PF}^{\Delta t}$ representing the deviation from the equilibrium position. From the properties of the Wigner D-matrix, we have

$$\begin{aligned}
&< D_{00}^{(2)}(A,t)D_{00}^{(2)*}(B,0) >_{Local} = \\
&\quad =< D_{00}^{(2)}(\Omega(A_0,t))D_{00}^{(2)}(\Omega(\Delta A,t))D_{00}^{(2)*}(\Omega(B_0,0))D_{00}^{(2)*}(\Omega(\Delta B,0)) >_{Local} \\
&\quad =< D_{00}^{(2)}(\Omega(\Delta A,t))D_{00}^{(2)*}(\Omega(\Delta B,0)) >_{Local} D_{00}^{(2)}(\Omega(A_0))D_{00}^{(2)*}(\Omega(B_0)) \\
&\quad =< D_{00}^{(2)}(\Omega(\Delta A,t)) >_{Local} < D_{00}^{(2)*}(\Omega(\Delta B,0)) >_{Local} P^{(2)}(\cos\theta_{AB}) \\
&\quad = S_A S_B P^{(2)}(\cos\theta_{AB}). \quad (34)
\end{aligned}$$

Similarly we also have $< D_{00}^{(2)}(B)D_{00}^{(2)*}(A) >_{Local} = S_B S_A P^{(2)}(\cos\theta_{AB})$.

Finally we have

$$\begin{aligned}
S_{0000}^2(t) &= P_A^2(S_A^2 + (1-S_A^2)e^{-t/\tau_{eA}}) + P_B^2(S_B^2 + (1-S_B^2)e^{-t/\tau_{eB}}) + 2P_A S_A P_B S_B P^{(2)}(\cos\theta_{AB}) \\
&= P_A^2 S_A^2 + 2P_A S_A P_B S_B P^{(2)}(\cos\theta_{AB}) + P_B^2 S_B^2 + P_A^2(1-S_A^2)e^{-t/\tau_{eA}} + P_B^2(1-S_B^2)e^{-t/\tau_{eB}} \quad (35)
\end{aligned}$$

and

$$\begin{aligned} F_{0000}(t) &= < P_A D_{00}^{(2)}(A,t) D_{00}^{(2)*}(A,0) + P_B D_{00}^{(2)}(B,t) D_{00}^{(2)*}(B,0) >_{Local} \\ &= P_A(S_A^2 + (1-S_A^2)e^{-t/\tau_{eA}}) + P_B(S_B^2 + (1-S_B^2)e^{-t/\tau_{eB}}) \\ &= P_A S_A^2 + P_B S_B^2 + P_A(1-S_A^2)e^{-t/\tau_{eA}} + P_B(1-S_B^2)e^{-t/\tau_{eB}} \end{aligned} \quad (36)$$

and

$$C_I(t) = S_{0000}^2(t) + (F_{0000}(t) - S_{0000}^2(t))e^{-t/\tau_{EX}}. \quad (37)$$

Comparing the t with τ_{eA}, τ_{eB} and $\tau_{EX}(>\tau_{eB}(\tau_{eA}))$, the internal correlation function Eq. 37 becomes

$$C_I(t) = F_{0000}(t) \qquad t \ll \tau_{EX} \quad (38)$$

$$C_I(t) = S_{0000}^2(t) \qquad t \gg \tau_{EX} \quad (39)$$

and converges to three important values at different time scales as shown in Fig. SS2

$$C_I(0) = F_{0000}(0) = P_A + P_B = 1 \qquad t \to 0 \quad (40)$$

$$S_{P1}^2 = C_I(t) = F_{0000}(t) = P_A S_A^2 + P_B S_B^2 \quad \tau_{eB}(\tau_{eA}) \ll t \ll \tau_{EX} \quad (41)$$

$$S_{P2}^2 = C_I(\infty) = S_{0000}^2(\infty) = P_A^2 S_A^2 + 2P_A S_A P_B S_B P^{(2)}(\cos\theta_{AB}) + P_B^2 S_B^2 \qquad t \to \infty \quad (42)$$

If the local motions of both states are much faster than the exchange rate between macrostates and $\tau_{eB}(\tau_{eA}) \to 0$, the effective internal correlation function is

$$C_I(t) = S_{P2}^2 + (S_{P1}^2 - S_{P2}^2)e^{-t/\tau_{EX}}, \quad (43)$$

$$\tau_{EX} = \frac{1}{k_{AB} + k_{BA}}.$$

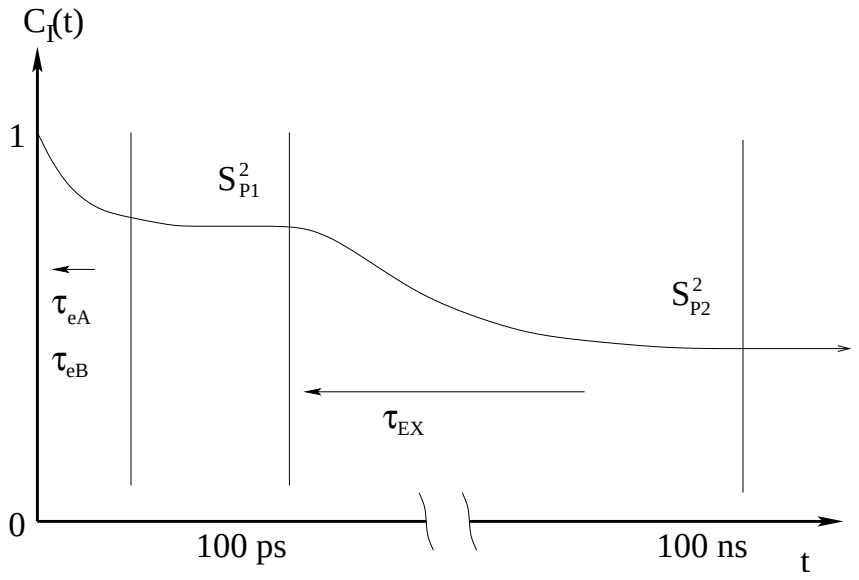


Figure S2: Schematic graph for a rotational correlation function of internal motions predicted by the two-macrostate conformational exchange model with fast internal motions.The exchange time scale τ_{EX} is much faster than the overall tumbling time τ_M but slower than the time scale of local motions τ_e. For the slow exchange case ($\tau_{EX} \gg \tau_M$), S^2_{P2} is averaged out by the overall tumbling and not observable.

Furthermore, if $S^2_A = S^2_B = 1$ the effective internal function can be simplified further to

$$C_I(t) = S^2_T + (1 - S^2_T)e^{-t/\tau_{EX}}, \tag{44}$$

$$S^2_T = P^2_A + 2P_A P_B P^{(2)}(\cos\theta_{AB}) + P^2_B.$$

This is consistent with the recent derivation for the exchange of two rigid rotors (no internal motions) with the same diffusion tensor.[21]

When $\theta_{AB} = 0$, namely the z directions for both states are parallel to each other, $S^2_{P2} = (P_A S_A + P_B S_B)^2$.

In summary, we obtained the total internal correlation functions for two limiting cases:

1. Slow exchange limit ($\tau_{eA}(\tau_{eB}) \ll \tau_M \ll \tau_{EX}$) when the decay of overall rotation is much quicker than the rate of exchange between the two macrostates. The total internal correlation function is simply a population average of two states as shown in Eq. 19 or 38. There is only one plateau value which is also the ensemble-averaged value of the generalized order

parameters of the two states given by Eq. 20 or 41.

2. Fast exchange limit ($\tau_{eA}(\tau_{eB}) \ll \tau_{EX} \ll \tau_M$) when the decay of overall rotation is much slower than the rate of exchange between the two states. The total internal correlation function involves the exchange term as displayed in Eq. 37. There are two plateau values, an ensemble averaged plateau value at short time by Eq. 41 (the same as slow exchange case) and a long-time combination as in Eq. 42 including a cross term of a second-order Legendre polynomial.

Assuming that the rotational diffusion tensor for overall motions is a symmetric top ($D_x = D_y$), we can multiply the overall rotation correlation function by the internal part and obtain the total function as below:

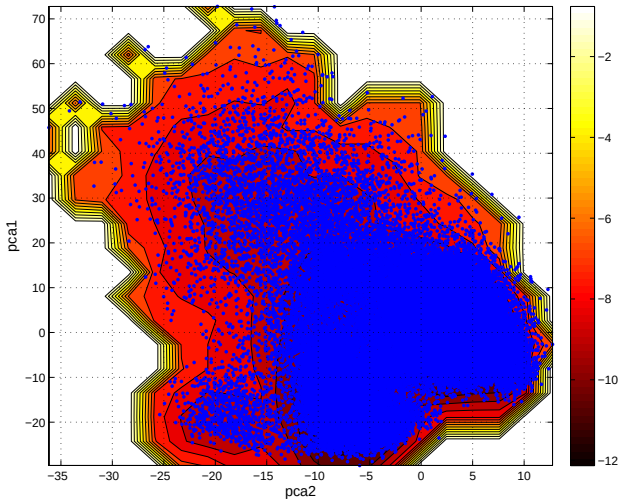
$$\text{Slow exchange:} \quad C_T(t) = \frac{1}{5}\{Ae^{-t/\tau_1} + Be^{-t/\tau_2} + Ce^{-t/\tau_3}\}(P_A C_{IA}(t) + P_B C_{IB}(t)) \tag{45}$$

$$C_T(t) = C_T(t|\theta, \tau_1, \tau_2, \tau_3, P_A, P_B, S_A^2, S_B^2, \tau_{eA}, \tau_{eB})$$

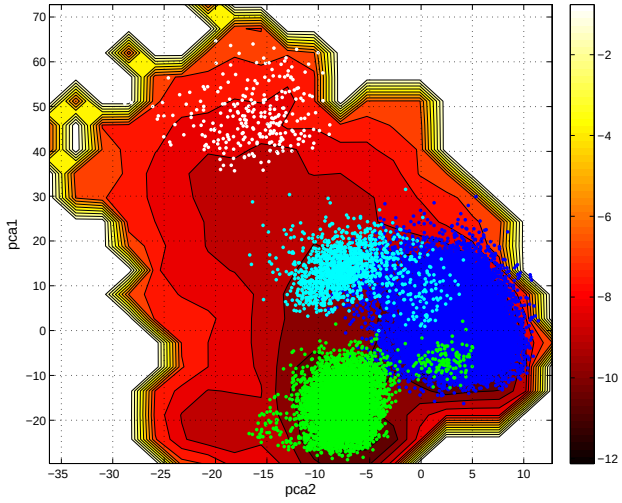
$$\text{Fast exchange:} \quad C_T(t) = \frac{1}{5}\{Ae^{-t/\tau_1} + Be^{-t/\tau_2} + Ce^{-t/\tau_3}\}(S_{0000}^2(t) + (F_{0000}(t) - S_{0000}^2(t))e^{-t/\tau_{EX}}) \tag{46}$$

$$C_T(t) = C_T(t|\theta, \tau_1, \tau_2, \tau_3, \tau_{EX}, P_A, P_B, S_A^2, S_B^2, \theta_{AB}, \tau_{eA}, \tau_{eB})$$

3 Supporting Data



(a)



(b)

Figure S3: Potential of mean force and clustered nodes projected onto the first two principal components. a) all nodes (82291 in blue) and b) semiopen (52204 in blue), closed (6554 in green), transit (1293 in cyan), and expanded (262 in white) state nodes in different colors. All nodes were classified by the RMSDs (< 2.2 Å) to the four individual macrostates.

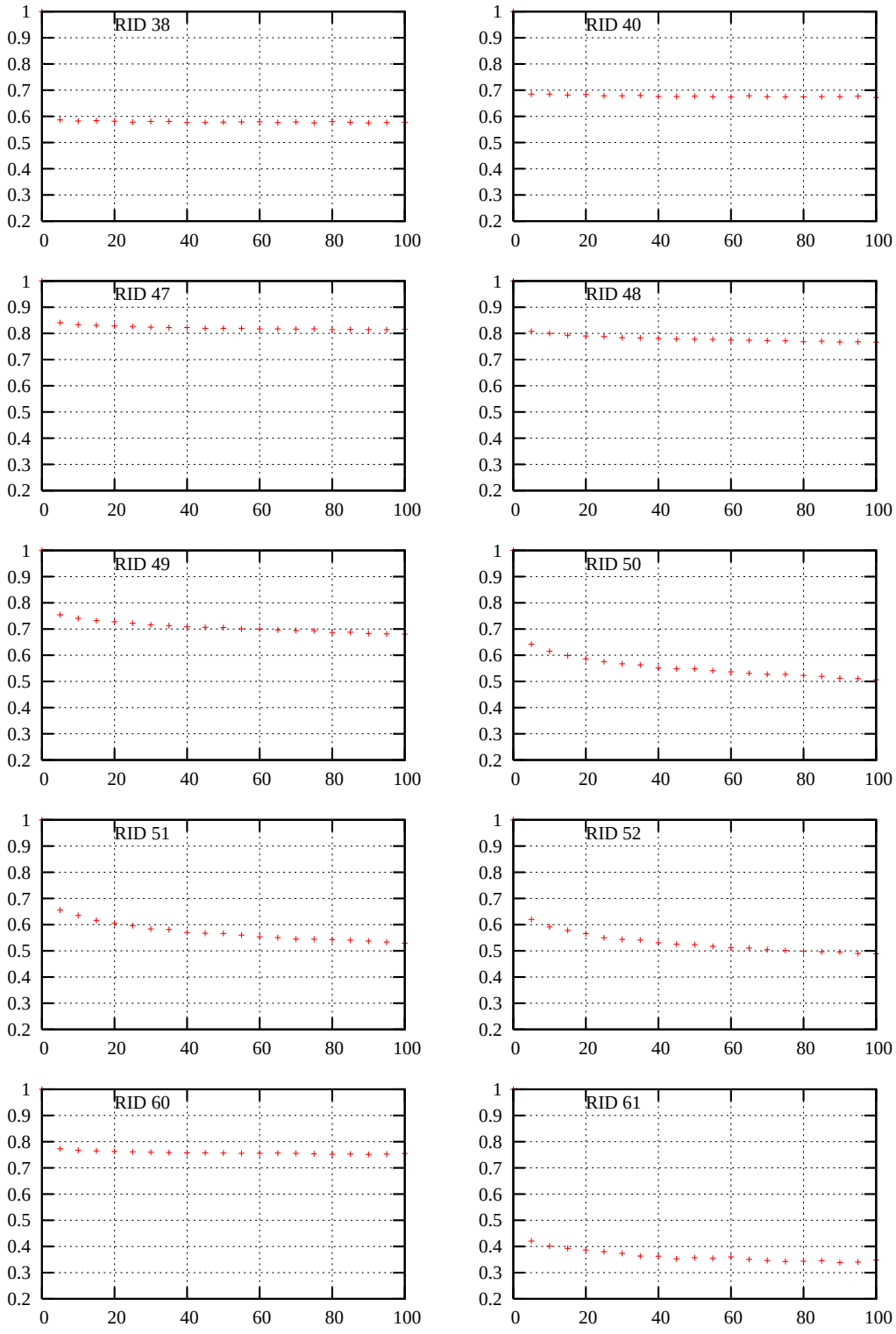


Figure S4: The first 100ps parts of the internal correlation functions of N-H vectors of residues related to the floppy region, calculated from the 10μs network MSM trajectory.

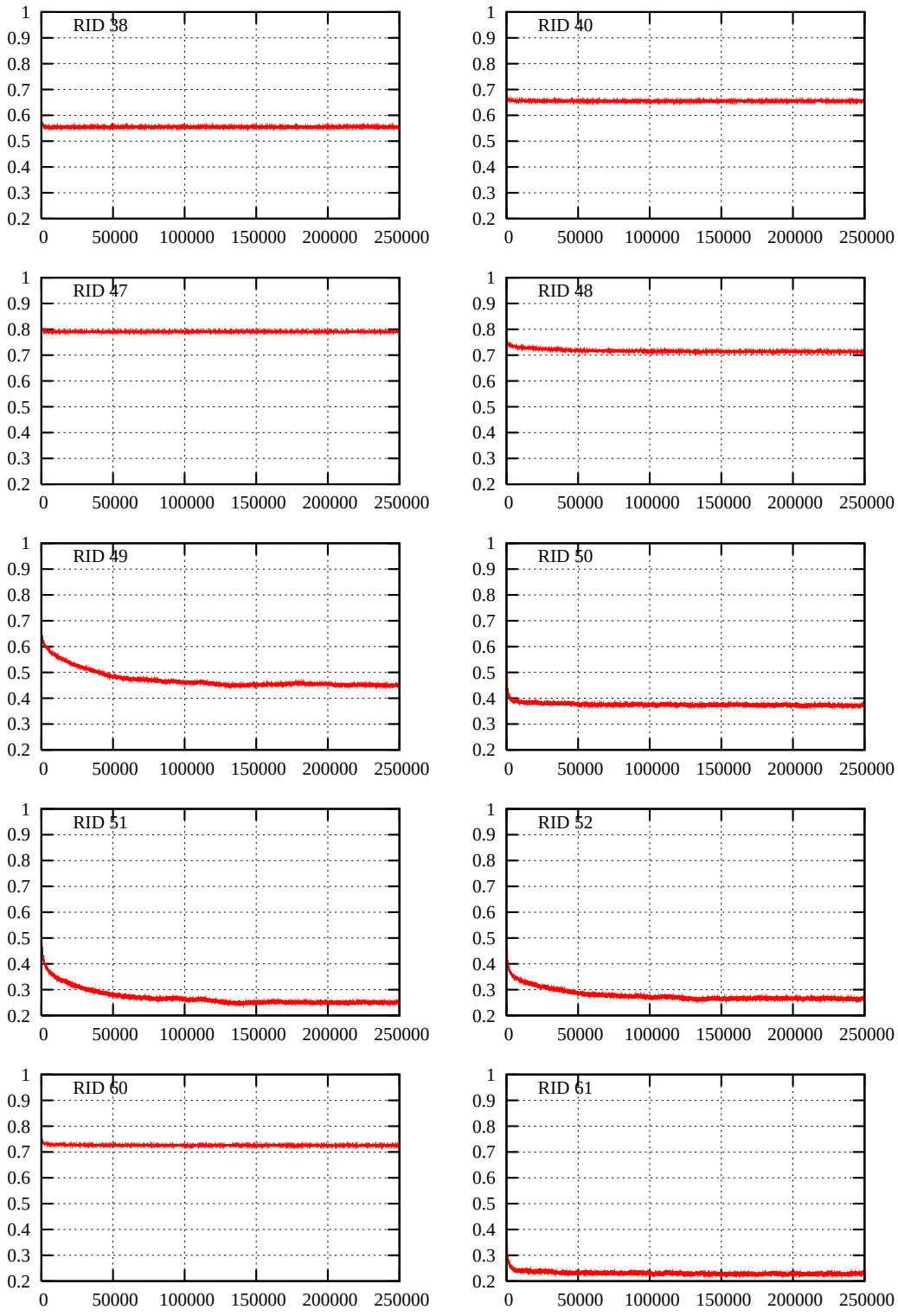


Figure S5: Same as Fig. S4 but on a much longer time (250ns).

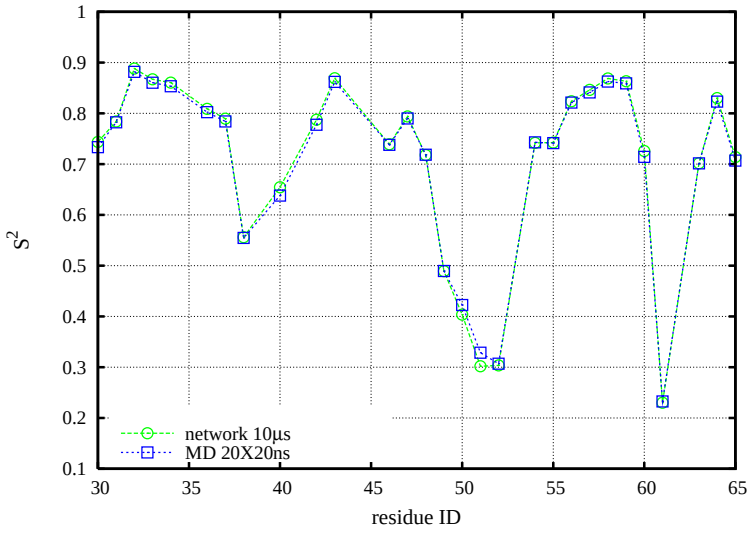


Figure S6: Comparison between the equilibrium S^2_{P2} ($C_I(\infty)$) calculated from the 10 μs network MSM trajectory with that from 20 $\times$ 20 ns MD trajectories.

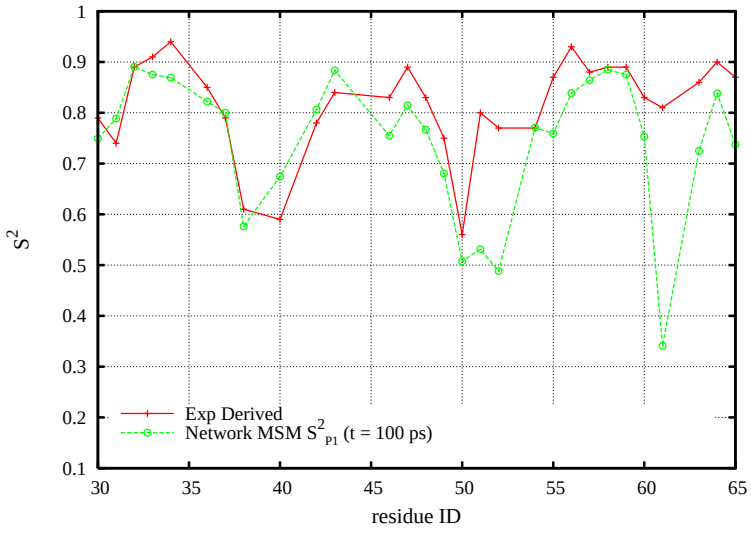


Figure S7: Comparison between the S^2_{P1} ($C_I(t = 100ps)$) calculated from the 10 μs network MSM trajectory with that experimentally derived.[25]

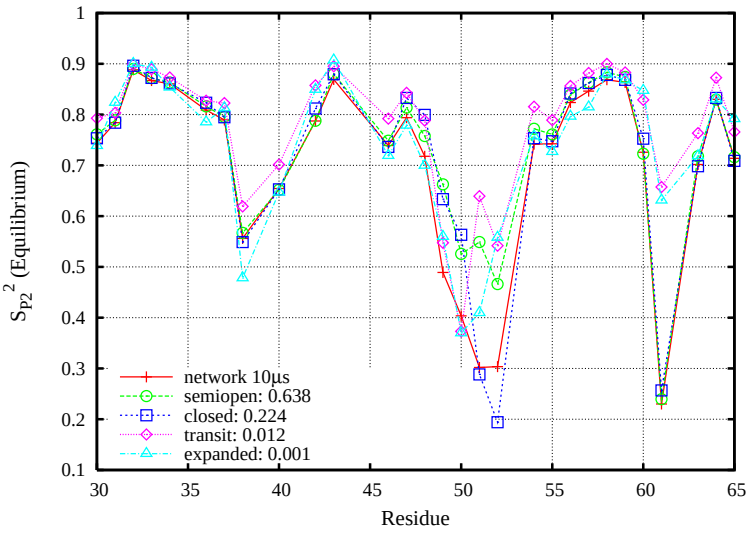


Figure S8: S^2_{P2} ($C_I(\infty)$) for four different macrostates (semiopen, closed, transit, and expanded) of HIV-1 PR, calculated from the 10 μs network MSM trajectory. S^2_{P2} is calculated using the equilibrium average of all snapshots belonging to the individual states. The total number of snapshots is 3123612. The legends in the figure also show the statistical populations (0.638, 0.224, 0.012, 0.001) of four different states respectively. There are other states not belonging to any four states with a total population of 0.14 roughly.

References

(1) Banks, J. L. et al. *J. Comput. Chem.* **2005**, *26*, 1752–1780.

(2) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(3) Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2004**, *25*, 479–499.

(4) Gallicchio, E.; Paris, K.; Levy, R. M. *J. Chem. Theory Comput.* **2009**, *5*, 2544–2564.

(5) Deng, N.-j.; Zheng, W.; Gallicchio, E.; Levy, R. M. *J. Am. Chem. Soc.* **2011**, *133*, 9387–9394.

(6) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schuette, C.; Noe, F. *J. Chem. Phys.* **2011**, *134*, 174105–174127.

(7) Bowman, G. R.; Beauchamp, K. A.; Boxer, G.; Pande, V. S. *J. Chem. Phys.* **2009**, *131*, 124101–124111.

(8) Deng, N.-J.; Dai, W.; Levy, R. M. submitted to J. Phys. Chem. B.

(9) Gillespie, D. T. *J. Phys. Chem.* **1977**, *81*, 2340–2361.

(10) Gillespie, D. T. *Annu. Rev. Phys. Chem.* **2007**, *58*, 35–55.

(11) Gillespie, D. T. *Markov Processes: an Introduction for Physical Scientists*; Academic Press: Boston, 1992.

(12) Andrec, M.; Felts, A. K.; Gallicchio, E.; Levy, R. M. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6801–6806.

(13) Zheng, W.; Gallicchio, E.; Deng, N.; Andrec, M.; Levy, R. M. *J. Phys. Chem. B* **2011**, *115*, 1512–1523.

(14) Favro, L. D. *Phys. Rev.* **1960**, *119*, 53–62.

(15) Woessner, D. E. *J. Chem. Phys.* **1962**, *37*, 647–654.

(16) Woessner, D. E.; Snowden, B. S.; Meyer, G. H. *J. Chem. Phys.* **1969**, *50*, 719–721.

(17) Wallach, D. *J. Chem. Phys.* **1967**, *47*, 5258–5268.

(18) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546–4559.

(19) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4559–4570.

(20) Levy, R. M.; Karplus, M.; McCammon, J. A. *J. Am. Chem. Soc.* **1981**, *103*, 994–996.

(21) Wong, V.; Case, D. A.; Szabo, A. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 11016–11021.

(22) Levy, R. M.; Karplus, M.; Wolynes, P. G. *J. Am. Chem. Soc.* **1981**, *103*, 5998–6011.

(23) Woessner, D. E. *J. Chem. Phys.* **1962**, *36*, 1–4.

(24) Clore, G. M.; Szabo, A.; Bax, A.; Kay, L. E.; Driscoll, P. C.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1990**, *112*, 4989–4991.

(25) Freedberg, D. I.; Ishima, R.; Jacob, J.; Wang, Y. X.; Kustanovich, I.; Louis, J. M.; Torchia, D. A. *Protein Sci.* **2002**, *11*, 221–232.