Supporting Information

Sadiq et al. 10.1073/pnas.1210983109

SI Materials and Methods

Molecular Simulation Protocol. The exact protocol for preparation depended on the specific requirements of each system and are described in the following sections. The general methodology is described here. Initial atomic coordinates were extracted from corresponding crystal structures in the Protein Data Bank (1). The standard AMBER force field (ff03) with standard ions (2) was used to describe all protease parameters. Each system was solvated using atomistic TIP3P water (3) and then electrically neutralized with an ionic concentration of 0.15 M NaCl, resulting in fully atomistic, explicit solvent systems each of ∼40,000 atoms. Minimization and equilibration simulations were carried out on a local cluster. The SHAKE algorithm (4) was used on all atoms covalently bonded to a hydrogen atom for stages of each simulation whose time step was greater than 1 fs. The long-range Coulomb interaction was handled using a GPU implementation (5) of the particle mesh Ewald summation method (PME) (6). A nonbonded cutoff distance of 9 Å was used with a switching distance of 7.5 Å for Van der Waals (VdW) interactions. A time step of 4 fs was made possible in all production simulations via the use of the hydrogen mass repartitioning scheme (7) implemented in ACEMD. This scheme takes advantage of the fact that individual atom masses do not appear explicitly in the equilibrium distribution; therefore, changing them only affects the dynamic properties of a system marginally (7) but not the equilibrium distribution. The change in the diffusion coefficient is minimal (10%) and small relative to the approximation that the TIP3 water model makes compared with real water (7).

Experimental accuracy of the molecular simulation protocol for the HIV-1 protease has been validated previously using NMR S^2 order parameters (8). All production simulations were carried out using ACEMD (9). Production-ensemble simulations were deployed for systems E1, E2, C1, and C2 on the GPUGRID compute infrastructure (10). Coordinate snapshots from all production simulations were generated every 100 ps. Initially, 500 runs each for systems E1 and E2 were submitted with a 500-ns limit per trajectory. Because several trajectories were not returned by the server, and some were more advanced than others at the time of analysis, a subset of 417 × 400 ns and 416 × 400 ns was used for the analysis of E1 and E2 sets, respectively. Similarly, a subset of 100 × 100 ns each was used for the analysis of the C1 and C2 control sets.

N-Terminal Extension of HIV-1 Protease. Initial structures were prepared corresponding to a dimeric immature protease with a single 5-aa (residues −5 to −1) N-terminal extension, with a sequence corresponding to the wild-type GagPol p6-PR cleavage site (VSFNF-PQITL), termed N-ter. The following structures were prepared: (*i*) N-ter disassociated with semiopen protease flap conformation (E1); (*ii*) N-ter self-associated in the active site with semiopen-flap conformation (E2); and (*iii*) a self-associated N-ter in the active site with closed-flap conformation (R1).

E1 was prepared as follows: atomic coordinates for wild-type dimeric HIV-1 protease were extracted from the crystal structure of Protein Data Bank (PDB) ID code 1HHP (1). For N-terminal self-association to occur, the immature protease cannot have the native-state conformation, as in the mature protease the N-terminal strand is distal from the active site. Furthermore, N-terminal association implies substantial conformational flexibility of a hinge region (residues 6–8) that connects the N-ter to the first region of downstream secondary structure, beginning at residue P9, which would be present in a folded precursor. Therefore, to allow conformational sampling of the putative N-terminal disassociated conformation, residues 1–8 in the first monomer of the 1HHP crystal structure were deleted, the residue builder in VMD (11) was used to add a preliminary chain from residue 5–8 with the corresponding sequence (VSFNF-PQITLWQR).

E2 and R1 were prepared as follows: The 1HHP structure was aligned (by backbone atoms of the protease excluding the flap residues 43–58 of each monomer) to the 1KJ4 structure containing the MA-CA cleavage site peptide bound to HIV-1 protease in the closed flap conformation. Using VMD, the first monomer N terminus chain (PDB ID code 1HHP for E2; PDB ID code 1KJ4 for R1) was dihedrally rotated by π and −0.1 π radians around the 8–9 and 6–7 C-N peptide bond, respectively, and residues 1–2 were deleted. This allowed the remaining N terminus to be rotated into the active site while avoiding steric clashes. Atomic coordinates for residues 1–7 (relabeled 5–2) of the MA-CA (VSQNY-PIVQ) peptide were connected to residues 3–99 of the first monomer of the respective systems and the MA-CA mutated into the p6-PR cleavage site. A monoprotonated (D25) state was assigned to the catalytic dyad for all three systems (12, 13). Crystallographic water molecules in 1KJ4 were preserved for R1. Water molecules are not present in the 1HHP structure. The final size of the E1, E2, and R1 systems was 39,369, 39,935, and 37,653 atoms, respectively.

Conjugate-gradient minimization was performed for 2,000 steps. During equilibration the position of all heavy protein atoms were restrained by a 10 $kcal \cdot mol^{-1} \cdot Å^{-2}$ spring constant. For systems E2 and R1 in which the connecting residues between the bound peptide and the protease were arbitrarily constructed, no restraints were applied to residues 3–10 from the outset of equilibration to allow for correct conformational reorientation of the linkage region. For all three systems, the hydrogen atoms and water molecules were then allowed to evolve for a total of 500 ps at 300 K to ensure thorough solvation of the system (14). The magnitude of the restraining spring constant was then set to 1 $kcal \cdot mol^{-1} \cdot Å^{-2}$ for 500 ps, then to 0.05 $kcal \cdot mol^{-1} \cdot Å^{-2}$ for another 500 ps, and then, finally, to zero for 500 ps. The temperature was maintained at 300 K using a Langevin thermostat with a low damping constant of 0.1/ps and the pressure maintained at 1 atm. An integration time step of 1 fs was used. The system was finally equilibrated for 6 ns of unrestrained simulation in the isothermal–isobaric ensemble (NPT) with an integration time step of 2 fs. All subsequent simulations were carried out in the canonical ensemble (NVT) with an integration time step of 4 fs. The final coordinates of E1 were used as input for production simulations. Because systems E2 and R1 were deliberately and arbitrarily built into self-associated conformations, further conformational relaxation was necessary. E2 was extended to 100 ns and R1 to 1 μs on a local GPU cluster; the output of these served as the input of subsequent production simulations; the R1 output additionally was the reference structure for subsequent analyses.

Production ensemble simulations were deployed for systems E1 and E2 on the GPUGRID compute infrastructure (10). Initially, 500 runs each for systems E1 and E2 were submitted with a 500-ns limit per trajectory. Because several trajectories were not returned by the server, and some were more advanced than others at the time of analysis, a subset of 417 × 400 ns and 416 × 400 ns was used for the analysis of E1 and E2 sets, respectively.

Control Octapeptide Systems. Initial structures were prepared for octapeptide ligand bound HIV-1 protease complexes for the (*i*) MA-CA (termed C1) and (*ii*) p6-PR (termed C2) cleavage sites.

Atomic coordinates for wild-type dimeric HIV-1 protease were extracted from the crystal structure of PDB ID code 1KJ4 (1). The first residue of the ligand was deleted to produce an octapeptide MA-CA sequence centered on the lytic peptide bond (SQNY-PIVQ) for C1. For C2, the corresponding residues were mutated to produce the p6-PR sequence (SFNF-PQIT). The inactive catalytic dyad D25N was converted into catalytically active D25 form with a monoprotonated state for both controls (12, 13). Crystallographic water molecules in PDB ID code 1KJ4 were preserved for C1 and C2. An additional water molecule was inserted between the lytic peptide bond and the catalytic dyad, as is expected in the general acid/general base (GA/GB) cleavage mechanism (15). The final size of the C1 and C2 was 37,647 and 37,695 atoms, respectively.

The equilibration procedure for C1 and C2 required more stringent considerations because these structures were derived closely from the crystal structure. Conjugate-gradient minimization was performed for 2,000 steps. During equilibration the position of all heavy protein atoms were restrained by a 1 kcal·mol^{-1}·Å^{-2} spring constant. All protein hydrogen atoms and all water molecules (except the catalytic one) were allowed to evolve for 1 ns at 300 K. Then, the restraining constant was set to zero for all atoms except the C_γ atoms of the catalytic dyad, the lytic bond atoms, and the catalytic water oxygen to preserve the geometry of the cleavage region, and the systems evolved for a further 1 ns. The restraining constant was then set to 0.1 kcal·mol^{-1}·Å^{-2} for the catalytic water oxygen and set to zero for all other atoms, and the systems evolved for a further 1 ns. The temperature was maintained at 300 K using a Langevin thermostat with a low damping constant of 0.1/ps, and the pressure was maintained at 1 atm. An integration time step of 2 fs was used. The systems were finally equilibrated for 10 ns of unrestrained simulation in the canonical ensemble (NVT) with an integration time step of 4 fs. The final coordinates of C1 and C2 were used as input for production simulations. All subsequent simulations were carried out in the NVT ensemble.

Production-ensemble simulations were deployed for systems C1 and C2 on the GPUGRID compute infrastructure (10). Initially, 150 runs each for systems C1 and C2 were submitted with a 100-ns limit per trajectory. Coordinate snapshots were generated every 25,000 time steps (100 ps). Because several trajectories were not returned by the server and some were more advanced than others at the time of analysis, a subset of 100 × 100 ns each was used for the analysis of C1 and C2 sets.

SI Conformational Relaxation of Preconstructed N-Terminal Self-Associated Protease

In addition to investigating the self-association process, we also tested the null hypothesis. That is, if autocatalysis of HIV-1 protease does not occur through intramolecular N-terminal self-association, then it follows that a constructed self-associated state with the flaps in a closed conformation and derived from existing crystal structures of octapeptide-bound cleavage complexes should be stoichiometrically unstable. This motivated the construction and relaxation of a closed-flap N-terminal self-associated structure, used as a reference system (termed R1).

Cross-Comparison of Octapeptide-Complexed Control Systems. No atomic resolution structure exists of an N-terminal self-associated HIV-1 protease, nor of the p6-PR octapeptide complex, which shares sequence identity to the N-terminal region. Therefore, R1 was derived from an existing crystal structure of MA-CA cleavage region peptide-bound HIV-1 protease, because of its partial sequence and structural overlap to the p6-PR cleavage region (Fig. S1*A*). It was then necessary to compare the flexibility of the R1 construct with respect to a p6-PR complexed HIV-1 protease which (as it was itself derived from MA-CA), in turn, needed to be compared against an MA-CA complexed system.

Therefore, two sets of control simulations were performed. The first control (C1) was a set of 100 × 100 ns explicit solvent simulations of the MA-CA octapeptide substrate complexed to HIV-1 protease and prepared from the 1KJ4 crystal structure. The second control (C2) was a set of 100 × 100 ns explicit solvent simulations of the p6-PR octapeptide HIV-1 protease complex derived from the crystal structure of the MA-CA complex.

The normalized frequency distribution of the rmsd relative to the C_α atoms of the equilibrated MA-CA ligand is shown for both C1 and C2 (Fig. S1*B*). Both systems exhibit a small rmsd for both the flaps (red) and the octapeptide ligands (cyan). The flap rmsd distribution peaks at 2 and 1.5 Å for the MA-CA (Fig. S1 *B, i*) and p6-PR systems (Fig. S1 *B, ii*), respectively, the latter being slightly sharper. Similarly, ligand rmsd peaks at 2 and 1.8 Å, respectively (Fig. S1 *B, iii* and *iv*), for the two systems, and, again, the p6-PR distribution is sharper.

Analysis of individual residue rmsds relative to the MA-CA structure for both octapeptide systems (Fig. S1*C*) was also performed. Sequence differences in p6-PR compared with MA-CA are highlighted in blue. The peak of all distributions was less than 3 Å, indicating stable fluctuations of each ligand amino acid within the active site. The distribution was not always identical for corresponding amino acids at a given sequence position (P_{-4} and P_1) and, indeed, was sometimes similar for nonidentical amino acids (P_2). Because the basis of differential enzymatic specificity for the different cleavage regions has partial mechanistic roots in the different flexibility of each ligand, it is not surprising that there is both overlap between nonidentical residues and heterogeneity between identical residues. The lack of large scale conformational fluctuations indicate that the p6-PR system sampled an equilibrium distribution over the 10-μs aggregate simulation time.

Comparison of N-Terminal Self-Associated Construct Against Octapeptide-Complexed Control Systems. A single production simulation of R1 was performed on a local GPU cluster for 1 μs to validate the stoichiometric stability of the closed bound state, hypothesized in our study. The flexibility of R1 was then compared against the intrinsic flexibility of the control system C2, consisting of the p6-PR octapeptide, which shares sequence identity with the N-terminal region (Fig. S2*A*).

The normalized frequency distribution of the rmsd relative to the C_α atoms of the equilibrated p6-PR ligand is shown for both C2 and S3 (Fig. S2*B*). The C2 system exhibits peak flap (red) and ligand (cyan) rmsds compared with its own equilibrated structure of 1.5 and 2 Å, respectively. The R1 system exhibits marginally increased rmsds relative to p6-PR, with peak values of 2 and 2.2 Å, respectively, for flaps and the octapeptide cleavage region. Thus, on average, the R1 construct exhibits very similar flexibility to C2.

Analysis of individual residue rmsds relative to the p6-PR equilibrated structure (Fig. S2*C*) reveals almost identical rmsd distributions near the N-terminal positions of the cleavage region (P_{-4} to P_{-2}). The rmsd increases by 2 Å at the P_{-1} and P_1 positions, is very similar again at P_2 and P_3, and differs substantially at P_4. These results are explained by the fact that R1 consists of a cleavage peptide region that is effectively tethered to the frame of the protease via the hinge region, whereas C2 is free. Thus, residues further from the hinge region should exhibit more similar distributions than those that are nearer.

Overall, the analysis of the N-terminal self-association construct (R1) with respect to control systems (C1 and C2) shows that the cleavage peptide region of the construct maintains a stable conformation with cleavable geometry. Based on our analyses, the final structure of the R1 simulation was selected as a reference structure to compare unbiased self-association simulations of systems E1 and E2.

SI MSMs

A discretized MSM of the complete N-ter self-association process was built. The 2D λ_x N-ter rmsd space within the range of access of the simulation data were discretized into 35 × 40 states based on a 1.0-Å^2 bin size. Variation of the bin size did not qualitatively change the outcome of the model. Each data point from the aggregate of ∼335-μs simulation trajectories was assigned to a corresponding state. A reversible transition matrix $T(\tau)$ was then constructed by counting transitions between these states at varying lag times, from which a plot of the implied timescale as a function of the lag time is obtained (Fig. S3). Convergence of the slowest mode of motion occurs at $\tau = 50$ ns and the corresponding transition matrix is used in the subsequent energetic and kinetic analyses. For example, the PMF of the conformational space is obtained from the first eigenvector of the transition matrix $T(\tau)$. This is further validated by performing a Chapman–Kolmogorov test as described in ref. 16 (Fig. S4).

The discretized model is further clustered into a smaller set of coarse-grained metastable states using the PCCA$^+$ method (17). This procedure clusters states into coarse-grained sets for which all constitutive states are more kinetically similar than other sets (i.e., have faster timescales for interconversion). The number of coarse-grained sets was set to seven based on the number of observable metastable states. The relative free energy of each metastable state was determined by integrating the probability distribution across all of the discretized microstates that compose it; the minimum free energy was arbitrarily set to zero.

The PCCA$^+$ method using a seven-state definition does not kinetically distinguish between semiopen and open conformations in the self-associating intermediate structure and, therefore, treats the observed modes of flap-opened association, N-terminal hairpin formation, and lateral threading as the same. This suggests fast interconversion between the two flap states; however, in vivo, the flaps need to open to permit traversal of the GagPol chain into the active site. Lateral threading in our simulations is observed because of the size of the much smaller size of the N-terminal system construct in comparison with a full-length GagPol. Therefore, to compute the error associated with the system setup, we exclude from the analysis all trajectories for which the initial mode of self-association is lateral threading and reconstruct the MSM using the reduced set of trajectories (Fig. S5). Excluding these trajectories makes very little difference to the PMF or to the segregation of states, both properties being qualitatively similar. The total flux computed for the S7→S2 transition is $k_{on} = 11.78 \times 10^{-6}$ ns^{-1}, which is a negligible difference to the original result. This confirms that interconversion between semiopen and open in the intermediate occurs rapidly and suggests that, in vivo, flap opening from an intermediate HIV-1 protease-GagPol structure is rapidly followed by GagPol entry and reformation of the semiopen state.

The total flux was decomposed into pathways along sets of states between S7 to S2. These, together with the percentage contribution of each pathway to the total flux, is listed in Table S1.

We applied a bootstrapping method to analyze the error in both the PMF and the flux calculations. Ten subsets, each containing 400 randomly selected trajectories from of the overall set of 833 trajectories in the E1 and E2 ensemble, were selected. The transition matrix and subsequently the PMF was calculated for each subset. The mean and SD of the PMF is shown in Fig. S6.

We applied the same bootstrapping method to calculate the mean and SD for each of the fluxes from S7 to S2, the total flux and the forward self-association constant. The transition matrices from each of the 10 subsets were used in conjunction with kinetic clustering that had been applied to the entire dataset. The fluxes for the overall dataset, each subset, together with the mean and SD are shown in Tables S2 and S3.

SI Movies

Supporting movies are provided that display various aspects of the HIV-1 protease N-terminal self-association process. Each movie displays a single trajectory from the E1 or E2 run sets and has a duration of between 400 and 500 ns of simulation time. The tertiary structure of the dimeric HIV-1 protease is depicted in white ribbon and white surface, the β-hairpin flaps are in red ribbon, the N-terminal extended region is in cyan surface, and the molecular structure representation and the hinge region are in cyan ribbon.

1. Berman HM, et al. (2000) The protein data bank. *Nucleic Acids Res* 28(1):235–242.
2. Duan Y, et al. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem* 24(16):1999–2012.
3. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
4. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J Comput Phys* 23:327–341.
5. Harvey M, De Fabritiis G (2009) An implementation of the smooth particle mesh Ewald method on gpu hardware. *J Chem Theory Comput* 5(9):2371–2377.
6. Essmann U, Perera L, Berkowitz ML, Darden T (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8577–9593.
7. Feenstra K, Hess B, Berendsen H (1999) Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J Comput Chem* 20: 786–798.
8. Sadiq SK, De Fabritiis G (2010) Explicit solvent dynamics and energetics of HIV-1 protease flap opening and closing. *Proteins* 78(14):2873–2885.
9. Harvey MJ, Giupponi G, Fabritiis GD (2009) Acemd: Accelerating biomolecular dynamics in the microsecond time scale. *J Chem Theory Comput* 5(6):1632–1639.
10. Buch I, Harvey MJ, Giorgino T, Anderson DP, De Fabritiis G (2010) High-throughput all-atom molecular dynamics simulations using distributed computing. *J Chem Inf Model* 50(3):397–403.
11. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14(1):33–38, 27–28.
12. Kovalskyy D, Dubyna V, Mark AE, Kornelyuk A (2005) A molecular dynamics study of the structural stability of HIV-1 protease under physiological conditions: The role of Na+ ions in stabilizing the active site. *Proteins* 58(2):450–458.
13. Wittayanarakul K, Hannongbua S, Feig M (2008) Accurate prediction of protonation state as a prerequisite for reliable MM-PB(GB)SA binding free energy calculations of HIV-1 protease inhibitors. *J Comput Chem* 29(5):673–685.
14. Meagher KL, Carlson HA (2005) Solvation influences flap collapse in HIV-1 protease. *Proteins* 58(1):119–125.
15. Park H, Suh J, Lee S (2000) Ab initio studies on the catalytic mechanism of aspartic proteinases: Nucleophilic versus general acid/general base mechanism. *J Am Chem Soc* 122(16):3901–3908.
16. Prinz JH, et al. (2011) Markov models of molecular kinetics: Generation and validation. *J Chem Phys* 134(17):174105.
17. Noé F, Horenko I, Schütte C, Smith JC (2007) Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states. *J Chem Phys* 126 (15):155102.

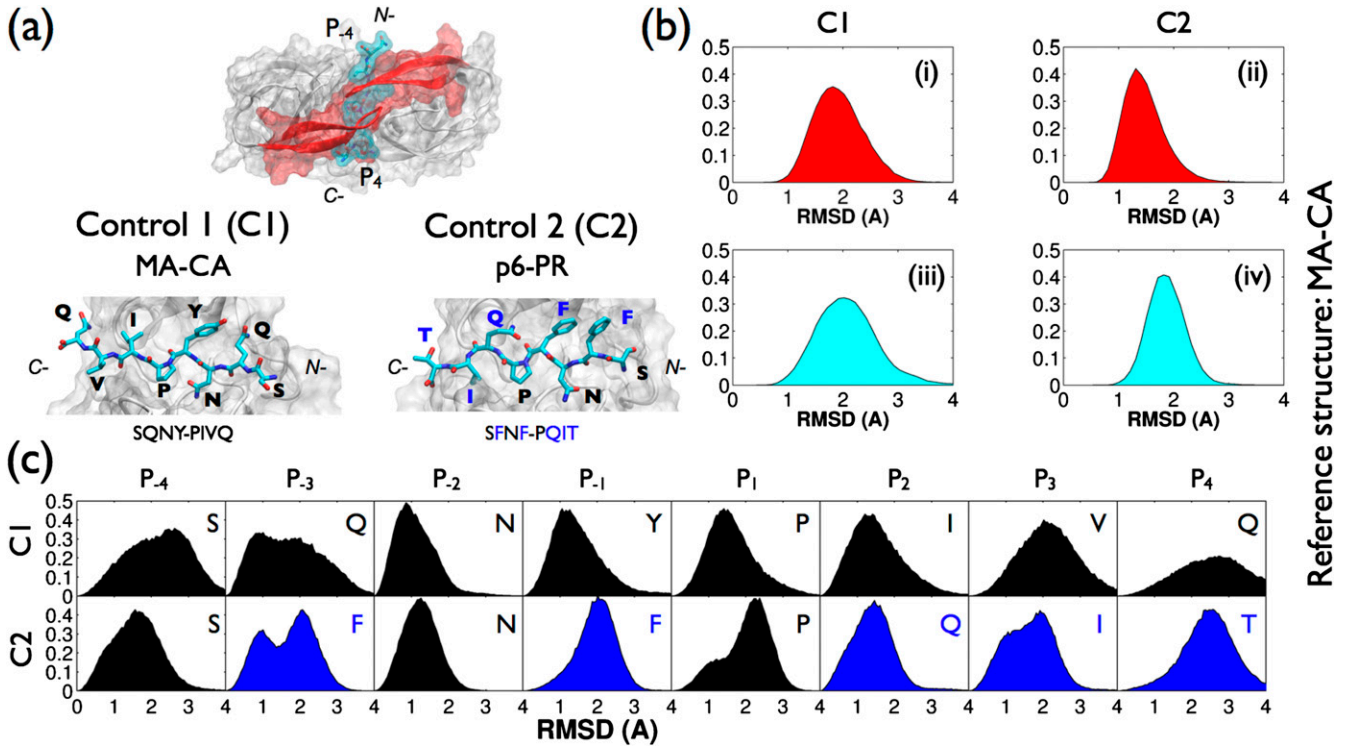


Fig. S1. Comparison of the flexibility of ligand-bound HIV-1 protease for the case of the MA-CA and p6-PR octapeptides. (*A*) Surface representation of ligand-bound HIV-1 protease from an equilibrated structure. The flaps (red) are in a closed conformation over the peptide (cyan). The p6-PR peptide exhibits partial sequence and structural similarity with MA-CA. (*B*) Normalized frequency distribution of C_α rmsd of C1, *i* and *iii*, and C2, *ii* and *iv*, relative to the equilibrated MA-CA–complexed structure for the flaps of the protease (red) and the ligand (cyan). (*C*) Normalized frequency distribution of C_α rmsd for individual residues in each octapeptide ligand in C1 and C2 relative to corresponding residues in the MA-CA–equilibrated structure.

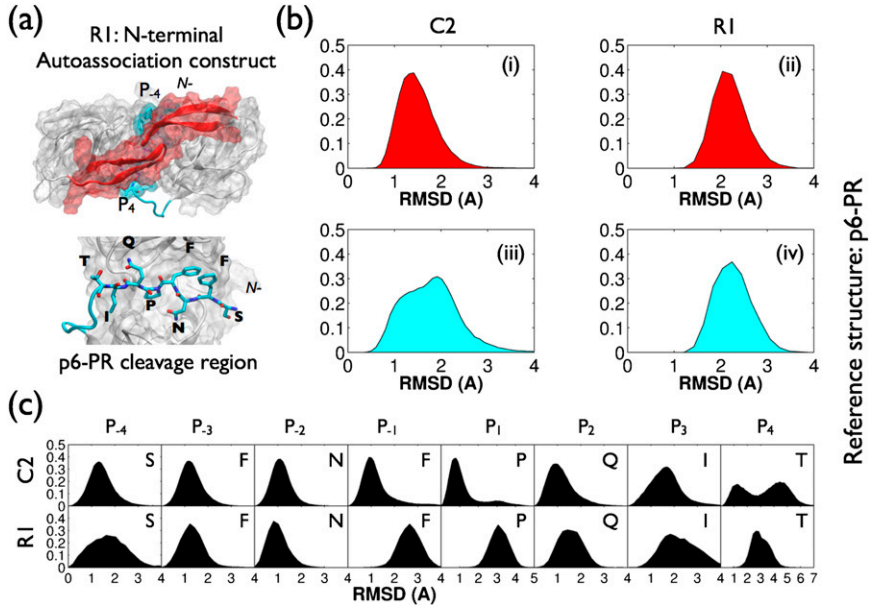


Fig. S2. Comparison of flexibility of N-terminal self-association construct with p6-PR cleavage peptide-bound HIV-1 protease. (*A*) Surface representation of S3. The flaps (red) are in a closed conformation over the N-terminal region (cyan). The N-terminal cleavage region sequence is by definition identical to p6-PR. (*B*) Normalized frequency distribution of C_α rmsd of C2, *i* and *iii*, and S3, *ii* and *iv*, relative to the equilibrated p6-PR–complexed structure for the flaps of the protease (red) and the cleavage region (cyan). (*C*) Normalized frequency distribution of C_α rmsd for individual residues in the cleavage peptide region for C2 and R1 relative to corresponding residues in the p6-PR–equilibrated structure.

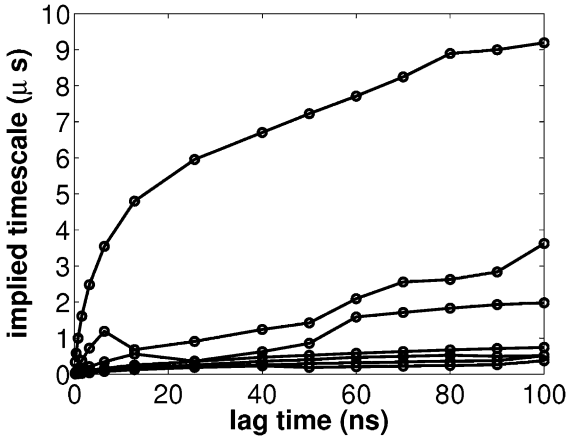


Fig. S3. Plot of implied timescale vs. lag time (τ) for 1,400-state λ_x N-ter rmsd MSM. The implied timescale of the slowest mode shows convergence by 50 ns.

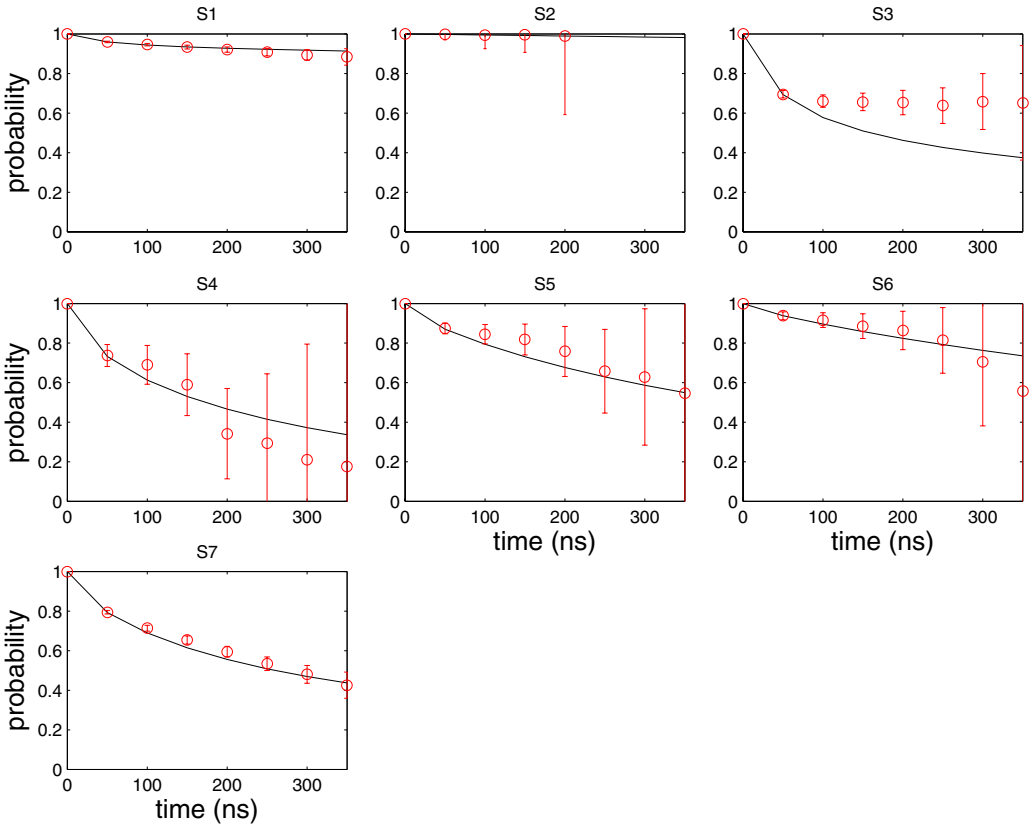


Fig. S4. Chapman–Kolmogorov test validating the Markov model. Each graph shows the evolution of the population of a given metastable state (S1 through S7) when initializing the population in that state. The evolution predicted by the Markov model estimated at a lag time of 50 ns (black solid lines) are compared with the directly observed in trajectory data (red bullets with error bars) over the accessible range of 0–350 ns. Error bars represent the 1-σ statistical uncertainty (68% confidence interval) of the simulation data. The curves are generally in good agreement within 2 σ, indicating that the Markov model is a good approximation of the underlying simulations.

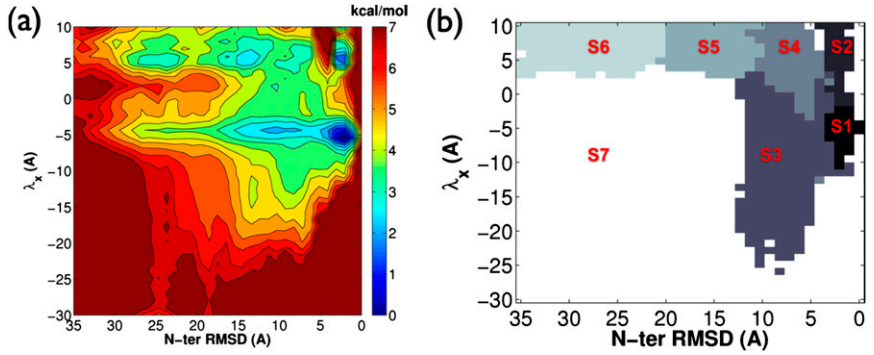


Fig. S5. (*A*) PMF from a MSM in the discretized λ_x N-terminal rmsd projection, excluding trajectories that exhibit threading as the predominant mode of self-association. (*B*) Kinetic clustering of the corresponding discretized λ_x N-terminal rmsd space into seven distinct metastable states, S1 to S7.

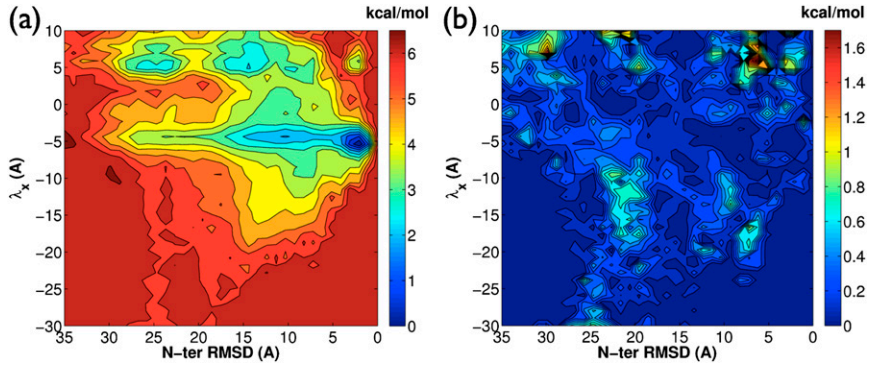


Fig. S6. (*A*) Mean of 10 PMF calculations using randomly selected subsets of the entire dataset. (*B*) SD of the 10 PMF calculations.

Table S1. Flux pathway decomposition for sets of states between S7 and S2

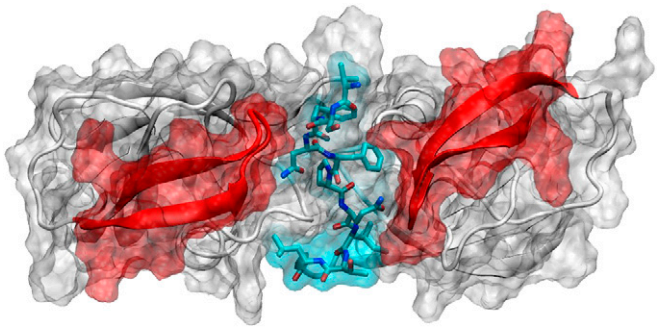
Pathway	Flux, ms^{-1}	Cumulative flux, ms^{-1}	Percentage of total flux	Cumulative percentage of total flux
S7→S3→S1→S2	5.734	5.734	53.59	53.59
S7→S3→S2	1.664	7.398	15.55	69.14
S7→S2	1.022	8.420	9.55	78.69
S7→S1→S2	0.478	8.898	4.47	83.16
S7→S5→S4→S3→S2	0.454	9.352	4.24	87.40
S7→S5→S3→S2	0.382	9.734	3.57	90.97
S7→S4→S2	0.294	10.028	2.75	93.72
S7→S5→S2	0.252	10.280	2.36	96.07
S7→S6→S2	0.228	10.508	2.13	98.21
S7→S6→S3→S2	0.098	10.606	0.92	99.12
S7→S5→S4→S1→S2	0.036	10.642	0.34	99.46
S7→S6→S4→S2	0.022	10.664	0.21	99.66
S7→S6→S5→S4→S2	0.012	10.676	0.11	99.78
S7→S6→S1→S2	0.010	10.686	0.09	99.87
S7→S5→S1→S2	0.008	10.694	0.07	99.94
S7→S5→S4→S2	0.006	10.700	0.06	100.00

Table S2. Total flux decomposition from S7 to S2 for all data and subsets 1–5 used in bootstrapping method

	Flux, ms^{-1}					
Pathway	All data	Subset 1	Subset 2	Subset 3	Subset 4	Subset 5
S3→S1	5.728	9.448	5.058	0.489	6.066	5.782
S4→S1	0.036	0.080	0.046	0.005	0.046	0.048
S5→S1	0.008	0.030	0.012	0.003	0.020	0.022
S6→S1	0.010	0.032	0.016	0.003	0.028	0.028
S7→S1	0.478	0.838	0.434	0.071	0.634	0.698
S1→S2	6.256	10.422	5.564	1.480	6.734	6.572
S3→S2	2.598	3.272	2.438	0.269	3.432	3.594
S4→S2	0.334	0.494	0.538	0.317	0.360	0.352
S5→S2	0.252	0.540	0.458	0.240	0.332	0.440
S6→S2	0.238	0.416	0.566	1.753	0.464	0.500
S7→S2	1.022	1.852	2.756	0.478	2.040	2.388
S4→S3	1.018	0.886	0.666	0.020	1.130	0.848
S5→S3	0.382	0.710	0.376	0.077	0.454	0.524
S6→S3	0.098	0.240	0.144	1.693	0.232	0.214
S7→S3	7.398	11.334	6.890	0.314	8.354	8.386
S6→S5	0.012	0.032	1.174	0.204	1.124	1.250

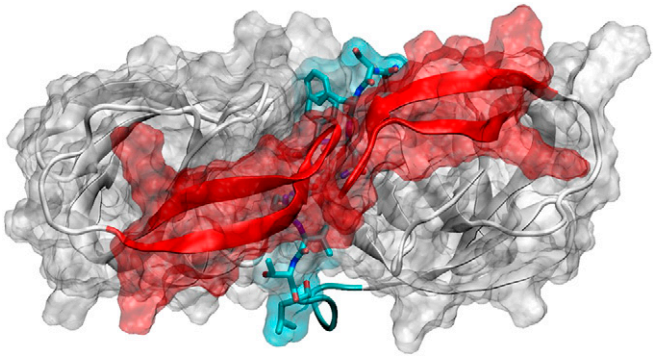
Table S3. Total flux decomposition from S7 to S2 for subsets 6–10 together with mean and SDs of all 10 subsets obtained from bootstrapping method

	Flux, ms^{-1}					
Pathway	Subset 6	Subset 7	Subset 8	Subset 9	Subset 10	Mean (SD)
S3→S1	7.620	6.232	5.968	0.651	5.034	5.235 (2.780)
S4→S1	0.046	0.060	0.042	0.006	0.042	0.042 (0.022)
S5→S1	0.016	0.020	0.016	0.004	0.014	0.016 (0.008)
S6→S1	0.020	0.030	0.018	0.005	0.016	0.020 (0.010)
S7→S1	0.412	0.666	0.230	0.114	0.460	0.456 (0.258)
S1→S2	8.114	7.006	6.264	0.541	5.562	5.826 (2.911)
S3→S2	2.598	3.198	2.822	1.607	2.294	2.552 (1.003)
S4→S2	0.784	0.624	0.442	0.194	0.414	0.452 (0.168)
S5→S2	0.550	0.664	0.396	0.247	0.436	0.430 (0.134)
S6→S2	0.610	0.922	0.394	0.323	0.452	0.640 (0.424)
S7→S2	2.758	3.344	1.726	1.512	2.112	2.097 (0.794)
S4→S3	1.054	1.000	0.982	0.160	0.676	0.742 (0.376)
S5→S3	0.434	0.626	0.280	0.059	0.422	0.396 (0.211)
S6→S3	0.194	0.252	0.158	0.044	0.112	0.328 (0.484)
S7→S3	9.266	8.386	7.870	1.904	6.542	6.925 (3.353)
S6→S5	0.030	0.040	0.042	0.010	0.016	0.066 (0.124)
S7→S5	0.556	0.442	0.446	0.117	0.312	0.333 (0.179)
S5→S4	0.054	0.122	0.014	0.056	0.100	0.119 (0.163)
S6→S4	1.622	1.800	1.166	0.451	1.354	1.018 (0.592)
S7→S4	0.622	0.490	0.502	0.107	0.410	0.473 (0.449)
S7→S6	0.798	1.122	0.608	0.325	0.490	0.644 (0.246)
Total flux	15.414	15.759	12.045	4.423	11.270	11.981 (4.375)
$k_{S7→S2}$	17.568	17.915	13.767	4.555	12.303	13.787 (5.568)



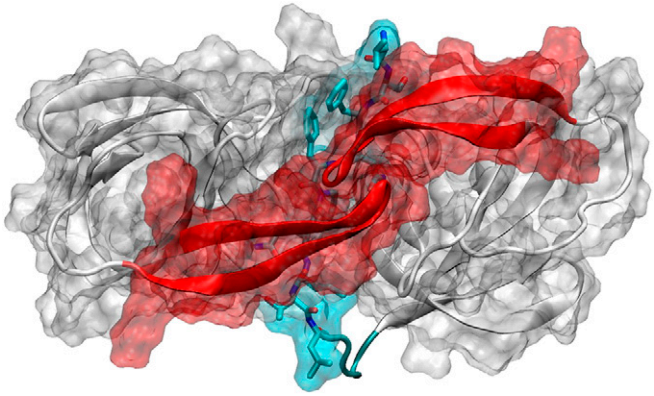
Movie S1. N-terminal self-association to the HIV-1 protease active site via an open-flap mechanism (E1).

Movie S1



Movie S2. N-terminal self-association to the HIV-1 protease active site via a threading mechanism (E1).

Movie S2



Movie S3. Conformational change of an N-terminal self-associated HIV-1 protease into a catalytically viable closed form (E2).

Movie S3