J|A|C|S ARTICLES

Published on Web 07/31/2009

Substrate Binding Mechanism of HIV-1 Protease from Explicit-Solvent Atomistic Simulations

Fabio Pietrucci, Fabrizio Marinelli, Paolo Carloni,* and Alessandro Laio

International School for Advanced Studies (SISSA-ISAS), via Beirut 2-4, I-34014 Trieste, Italy

Received April 16, 2009; E-mail: carloni@sissa.it

Abstract: The binding mechanism of a peptide substrate (Thr-Ile-Met-Met-Gln-Arg, cleavage site p2-NC of the viral polyprotein) to wild-type HIV-1 protease has been investigated by 1.6 μs biased all-atom molecular dynamics simulations in explicit water. The configuration space has been explored biasing seven reaction coordinates by the bias-exchange metadynamics technique. The structure of the Michaelis complex is obtained starting from the substrate outside the enzyme within a backbone rmsd of 0.9 Å. The calculated free energy of binding is -6 kcal/mol, and the kinetic constants for association and dissociation are 1.3×10^6 M^{-1} s^{-1} and 57 s^{-1}, respectively, consistent with experiments. In the main binding pathway, the flaps of the protease do not open sizably. The substrate slides inside the enzyme cavity from the tight lateral channel. This may contrast with the natural polyprotein substrate which is expected to bind by opening the flaps. Thus, mutations might influence differently the binding kinetics of peptidomimetic ligands and of the natural substrate.

1. Introduction

For more than 20 years, the Human Immunodeficiency Virus type-1 Protease (HIV-1 PR) has been one of the main targets of anti-AIDS drug design. HIV-1 PR cuts polyproteins to smaller fragments and is essential for the virus life cycle. Its inactivation leads to noninfectious viral particles.[1] Many inhibitors have been developed to block the aspartic protease active site, and several of them are used in medical treatment. Most FDA-approved HIV-1 PR inhibitors are peptidomimetic; i.e., they mimic the structure of a fragment of the natural substrate, competing with its binding.[2,3]

The experimental literature about the structure and function of HIV-1 PR is vast. X-ray crystal structures of HIV-1 PR in bound and unbound forms have been reported, revealing a C_2 symmetric homodimer with a large binding pocket covered by two Gly-rich β-hairpins (flaps). The bound and unbound forms of the enzyme show sizable structural differences: the complex has "closed" flaps, i.e. in contact with the ligand.[2] On the other hand, the free enzyme can also adopt a "semiopen" conformation with the flaps shifted up from the active site and only partially in contact with each other.[2] The residues that do not belong to the flaps show smaller displacements upon binding. The flexibility of the flaps is confirmed by solution NMR[4,5] and fluorescence experiments.[6,7] In the unliganded protease, the semiopen form is considered predominant, but there are also indications of the presence of a scarcely populated truly open form with separated flap tips.[4,5] Arguably, an extended opening of the flaps is necessary to allow binding of the viral polyproteins due to the large size of the substrate. It is not clear if the binding of peptidomimetic inhibitors is a one-step or a two-step process:[6,7] in the latter case, binding would proceed through the fast formation of a collision complex, followed by a slower conformational change to a tighter complex.[6]

HIV-1 PR has also been largely investigated by computational methods. In several works, the binding affinity of different inhibitors has been computed by molecular docking based on scoring functions[8,9] or on the more predictive Molecular Mechanics Poisson−Boltzmann Surface Area model (MM/PBSA).[10,11] Although computationally expedient and very useful for drug design, these approaches might be too approximate to capture the correct physical chemistry of the binding process. In particular, they typically do not account for the flexibility of the whole protein and for hydration. Water molecules can make hydrogen bonds simultaneously with both the enzyme and the ligand. This is believed to lower sizably both the enthalpy and entropy of binding.[12] In particular, a water molecule (usually named W301) is hydrogen-bonded between the flaps tips and the ligand in most experimental complexes with peptidomimetic inhibitors.[2] To address the role of water and the flexibility of

(1) Pettit, S. C.; Moody, M. D.; Wehbie, R. S.; Kaplan, A. H.; Nantermet, P. V.; Klein, C. A.; Swanstrom, R. *J. Virol.* **1994**, *68*, 8017–8027.
(2) Wlodaver, A.; Vondrasek, J. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 249.
(3) Ohtaka, H.; Freire, E. *Prog. Biophys. Mol. Biol.* **2005**, *88*, 193.
(4) Ishima, R.; Freedberg, D. I.; Wang, Y. X.; Louis, J. M.; Torchia, D. A. *Struct. Fold. Des.* **1999**, *7*, 1047.
(5) Ishima, R.; Louis, J. M. *Proteins* **2008**, *70*, 1408.
(6) Furfine, E. S.; Dsouza, E.; Ingold, K. J.; Leban, J. J.; Spector, T.; Porter, D. J. T. *Biochemistry* **1992**, *31*, 7886.
(7) Rodriguez, E. J.; Debouck, C.; Deckman, I. C.; Abusoud, H.; Raushel, F. M.; Meek, T. D. *Biochemistry* **1993**, *32*, 3557.
(8) Lee, C. Y.; Yang, P. K.; Tzou, W. S.; Hwang, M. J. *Protein Eng.* **1998**, *11*, 429.
(9) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W. M.; Taylor, R. D.; Taylor, R. *J. Med. Chem.* **2005**, *48*, 6504.
(10) Bartels, C.; Widmer, A.; Ehrhardt, C. *J. Comput. Chem.* **2005**, *26*, 1294.
(11) Stoica, I.; Sadiq, S. K.; Coveney, P. V. *J. Am. Chem. Soc.* **2008**, *130*, 2639.
(12) Li, Z.; Lazaridis, T. *Phys. Chem. Chem. Phys.* **2007**, *9*, 573.

the system appropriately,[13] one can use explicit-solvent molecular dynamics (MD). The flap dynamics of the free enzyme has been simulated by MD observing spontaneous opening and closing events.[14,15] Interestingly, full opening of the flaps has been suggested to be unnecessary for dissociation of Saquinavir from protease mutants.[16] Coarse grained[17,18] and implicit solvent MD[17,19] have further provided insight on the binding of ligands to the proteases. In particular, in ref 17 it is suggested that a cyclic-urea inhibitor can bind without full opening of the flaps.

Unfortunately, to date the high computational cost has limited the time scale of explicit-solvent MD to 100 ns. This is not sufficient to observe binding and unbinding processes. Several powerful approaches have been designed to overcome such limitation of time scale, suitably enhancing the capability of MD to explore new configurations.[20] Among these, the metadynamics technique[21] accelerates rare events by biasing selected reaction coordinates. This technique has been applied to study protein− and DNA−ligand binding with all-atom explicit-solvent force fields.[22−24] The latter applications showed that it is possible, in specific cases, to observe the binding and dissociation pathways, to predict the geometry of the complex, and to reconstruct the free-energy surface as a function of a few (up to three) reaction coordinates. In these systems, the calculated binding affinity turns out to be in fair agreement with the experimental data.[22−24] However, similarly to other methods, metadynamics is limited by the need to identify in advance a small number (two or three) of reaction coordinates. Strategies have been developed to address this limitation, e.g., by combining metadynamics with parallel tempering,[25] by iteratively constructing an optimal reaction coordinate,[26] or, in the bias-exchange metadynamics approach,[27−30] by treating simultaneously a large number of collective variables. For systems of the complexity of HIV-1 PR, several variables may simultaneously play an important role for binding, e.g., the opening of the flaps, the distance between the ligand and the cavity, the number of hydrogen bonds or hydrophobic contacts with the flaps or with the cavity, the number of interfacial water molecules, etc.[13] Here we adopted the latter method[27] to obtain a comprehensive picture of the substrate binding and unbinding mechanism of wild-type HIV-1 PR. The calculations are validated against experimental thermodynamic and kinetic data. We focus on the substrate Thr-Ile-Met-Met-Gln-Arg (p2-NC cleavage site of the gag-pol viral polyprotein). The choice of this ligand is motivated by the fact that most of the inhibitors in clinical use are peptidomimetic; i.e., they are based on such a natural substrate, and they might share with it several features.

Our simulation shows several binding events, starting from the ligand outside the enzyme and ending in the Michaelis complex, which is predicted to be the lowest free-energy minimum. A quantitative kinetic model of the association and dissociation process obtained here allows calculating binding free energies and rate constants that turn out to be in agreement with available experimental data. Hydration and flap fluctuations turn out to play a key role for the binding. Remarkably, the main binding pathway of the small substrate does not involve full opening of the flaps, which is instead expected to occur, for topological reasons, when the enzyme binds the long viral polyprotein in vivo. Several mutations that have been reported to bring drug resistance involve residues forming important contacts with the substrate in the molecular recognition process. Because of the similarity between the substrate and some FDA-approved drugs, it can be speculated that some of these mutations would affect to a smaller extent the binding of the polyprotein expressed by the virus, as this binding would take place through a different pathway. Thus, mutations have high chances of influencing differently the binding kinetics of small ligands like drugs and of the natural substrate.

2. Methods

2.1. Bias-Exchange Metadynamics Simulations. MD simulations have been performed with the Amber03 force field[31] for the enzyme and substrate (hereafter termed SUB) and the TIP3P[32] model for water. We modified a version of the GROMACS 3.3.1 simulation program[33] to run bias-exchange metadynamics (BE-META).[27] The initial atomic model has been obtained starting from the experimental coordinates of the HIV-1 PR/MVT-101 complex (pdb code 4HVP), as described in a previous work.[34] HIV-1 PR and SUB have been solvated by 7710 water molecules in a 269 nm^3 orthorhombic periodic box. Six Cl^- ions were added to neutralize the net positive charge. The particle-mesh Ewald method[35,36] was used for long-range electrostatic with a short-range cutoff of 0.8 nm. A cutoff of 0.8 nm was used for the Lennard-Jones interactions. All bond lengths were constrained to their equilibrium length with the LINCS[37] algorithm. Also the C_γ(Asp25)−C_γ(Asp25′) distance was constrained to 0.34 nm. The time step for the MD simulation was 2.0 fs. NPT simulations at 300 K and 1 atm were performed by coupling the system to a

(13) Gilson, M.; Given, J.; Bush, B.; McCammon, J. *Biophys. J.* **1997**, *72*, 1047–1069.
(14) Meagher, K.; Carlson, H. *Proteins* **2005**, *58*, 119.
(15) Hornak, V.; Okur, A.; Rizzo, R.; Simmerling, C. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 915.
(16) Sadiq, A. K.; Wan, S.; Coveney, P. V. *Biochemistry* **2007**, *46*, 14865–14877.
(17) Chang, C. E.; Trylska, J.; Tozzini, V.; McCammon, J. A. *Chem. Biol. Drug Des.* **2007**, *69*, 5–13.
(18) Trylska, J.; Tozzini, V.; Chang, C. E.; McCammon, J. A. *Biophys. J.* **2007**, *92*, 4179–4187.
(19) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. *J. Am. Chem. Soc.* **2006**, *128*, 2812–2813.
(20) Dellago, C.; Bolhuis, P. G. *Advanced Computer Simulation Approaches for Soft Matter Sciences III*; Springer Berlin/Heidelberg, 2008; Vol. 221, p 167.
(21) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
(22) Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2005**, *127*, 2600.
(23) Fiorin, G.; Pastore, A.; Carloni, P.; Parrinello, M. *Biophys. J.* **2006**, *91*, 2768–2777.
(24) Vargiu, A. V.; Ruggerone, P.; Magistrato, A.; Carloni, P. *Nucleic Acids Res.* **2008**, *36*, 5910–5921.
(25) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 13435–13441.
(26) Branduardi, D.; Gervasio, F. L.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 054103.
(27) Piana, S.; Laio, A. *J. Phys. Chem. B* **2007**, *111*, 4553–4559.
(28) Piana, S.; Laio, A.; Marinelli, F.; Troys, M. V.; Bourry, D.; Ampe, C.; Martins, J. C. *J. Mol. Biol.* **2008**, *375*, 460–470.
(29) Todorova, N.; Marinelli, F.; Piana, S.; Yarovsky, I. *J. Phys. Chem. B* **2009**, *113*, 3556–3564.
(30) Leone, V.; Lattanzi, G.; Molteni, C.; Carloni, P. *PLoS Comput. Biol.* **2009**, *5*, e1000309.
(31) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
(32) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(33) Lindahl, E.; b., Hess,; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.
(34) Piana, S.; Carloni, P.; Parrinello, M. *J. Mol. Biol.* **2002**, *319*, 567.
(35) Darden, T. A.; York, D. *J. Chem. Phys.* **1993**, *98*, 10089.
(36) Essman, U.; Perera, L.; Berkowitz, M. L.; Darden, T. A.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.
(37) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, G. E. M. J. *J. Comput. Chem.* **1997**, *18*, 1463.

Nose−Hoover thermostat[38,39] and a Berendsen barostat,[40] both with a relaxation time of 1 ps. After 1.4 ns of equilibration, the barostat was removed and the BE-META simulation was started. The atomic coordinates were saved every 5 ps and the energy every 0.1 ps.

The indexing 1−99 and 1′−99′ is adopted for the two dimers forming HIV-1 PR. On the basis of experimental[41] and theoretical evidence,[42] Asp25 has been taken deprotonated and Asp25′ monoprotonated. The substrate is indexed as Thr(P3)-Ile(P2)-Met(P1)-Met(P1′)-Gln(P2′)-Arg(P3′), the scissile bond being P1−P1′. As a reference experimental structure for the HIV-1 PR/SUB complex, the crystallographic positions of the complex between the inactive D25N protease and ATIMMQRG substrate[43] (pdb code 1KJ7) are considered. The crystallographic water molecule located in the cavity under the flap tips and above SUB is called W301 following the usual terminology.

The BE-META approach allows biasing simultaneously several collective variables (CVs). The following set of 7 CVs has been selected as putative reaction coordinates to explore the binding mechanism. The CVs are explicit functions of the atomic coordinates:

• S_A is the sum of hydrophobic side chain carbon contacts between ligand and flaps. S_B is identical but counts contacts between the ligand and the part of cavity not belonging to flaps. They are defined as

$$S_{A,B} = \sum_i \sum_j C(R_{ij})$$

with the sums running on the appropriate sets of atoms. Contacts are defined by the following function, which switches smoothly from 0 to 1 for distances below a threshold R_0

$$C(R) = [1 - (R/R_0)^n]/[1 - (R/R_0)^m]$$

where R is the distance between two atoms, $R_0 = 0.3$ nm, $n = 8$, and $m = 12$.

• S_C is defined like S_A and S_B, with the sum running over the possible H-bonds among O and H atoms in the ligand and in the cavity.

• S_D is the distance between the center of the scissile peptidic bond of the ligand and the center of the C_γ(Asp25)−C_γ(Asp25′) atom pair in the catalytic dyad.

• S_E is the distance between the center of the C_α's in the left flap tip (residues 48−53) and the center of those in the right flap tip (residues 48′−53′).

• S_F counts the number of water molecules bridging between the ligand and the cavity

$$S_F = \sum_i \sum_j \sum_w C(R_{iw}) \cdot C(R_{wj})$$

where the sums over i and j run over O/H atoms (corresponding to native H bonds) in the ligand and in the enzyme, respectively, while the sum over w runs over all O atoms belonging to water molecules.

• S_G is the distance between the center of the C_α's of the ligand and the center of the C_α's of residues 24, 26, 27, 24′, 26′, and 27′, located in the middle of the enzyme and close to the Asp25-Asp25′ catalytic dyad.

It has to be stressed that variable S_C provides only an approximate count of the H bonds since a more rigorous definition would include the angles formed by atoms and a more sharp switching function $C(R)$. Similarly, the number of water molecules bridging through H bonds between ligand and cavity is only approximately proportional to S_F. However, the present definitions are more suitable to be used as differentiable collective variables. The CVs were saved every 0.1 ps.

To reduce the computational cost, walls have been put on variables S_D and S_E preventing them to reach values larger than 1.9 nm. Moreover, the ligand was restricted to a cone of angle 45° with axis equal to the C_2 symmetry axis of HIV-1 PR. Control simulations have also been performed beyond the restrictions above, to check the convergence of the results.

The parameters adopted in the BE-META simulations are the following: Gaussian height 0.05 kcal/mol, Gaussian widths equal to 0.5 for $S_{A,B,C,F}$ and 0.02 nm for $S_{D,E,G}$, deposition of a Gaussian every 1 ps, exchanges of bias attempted every 2 ps. These parameters have been optimized to obtain a fast exploration of the conformations while still retaining a good accuracy in the reconstructed free energy.

BE-META simulations were performed biasing each of the seven CVs on a different replica (plus one replica without any bias) for a total of 45 ns per replica. We initialized the simulation with the substrate completely outside the cavity and misoriented with respect to the complex, namely, with Arg(P3′) pointing toward the cavity. After a few nanoseconds, in several replicas the substrate approaches the cavity and starts to thread inside the pocket starting from Thr(P3). After 10 ns in one replica, SUB entered the binding pocket, and after 35 ns, four out of eight replicas fully accomplished the binding process by reaching the structure of the Michaelis complex (Figure 2 and state B1 in Figure 1). During the simulation, a total of four and ten independent binding and unbinding events were observed. This is sufficient to ensure an accurate and reproducible description of the binding/unbinding process and allows harvesting enough statistics for constructing an accurate rate model.

To improve the statistics on the explored states, a simulation has also been performed replicating four times each replica, for a total of 32 replicas and 40 ns each. This redundancy speeds up the convergence of the reconstructed free-energy profiles. The trajectories from both the 8-replica and 32-replica simulations, for a total simulation time of 1.6 μs, have been employed to construct a kinetic model, as described in the following section.

2.2. Construction of the Kinetic Model. Starting from the BE-META simulation data, a thermodynamic and kinetic model of the binding process has been constructed, applying the methodology of ref 44. In short, the CV space is divided in a grid of bins; the free energy of each bin is estimated by a weighted-hystogram approach; and the matrix of transition rates is computed assuming a Markovian diffusive behavior among the bins. A careful analysis shows that to describe accurately the thermodynamics and kinetics of the binding process it is sufficient to consider variables C, D, E, and F, as the others are correlated to these. First, the BE-META trajectories have been analyzed by subdividing the CV space of these four variables in a hypercubic grid of 3208 bins with sides in the four directions $R_C = 1$, $R_D = 0.1$ Å, $R_E = 0.1$ Å, $R_F = 1$. Molecular structures within each bin differ by C_α rmsd < 2 Å (substrate plus enzyme cavity), indicating that this choice of variables appropriately discriminates among all the relevant structures. The equilibrium free energy of each bin has been computed by a weighted-histogram technique[45] in which the effect of the bias is removed from the populations.[44] It has been verified that multiplying by 1.5 or dividing by 1.5 the side of the hypercubes has no relevant qualitative effect on the description of the system,

(38) Nose, S. *Mol. Phys.* **1984**, *52*, 255.
(39) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.
(40) Berendsen, H. J. C.; Postma, J. P. M.; Gusteren, W. F. V.; Nola, A. D.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(41) Hyland, L. J.; Tomaszek, T. A.; Roberts, G. D.; Carr, S. A.; Magaard, V. W.; Bryan, H. L.; Fakhoury, S. A.; Moore, M. L.; Minnich, M. D.; Culp, J. S.; Desjarlais, R. L.; Meek, T. D. *Biochemistry* **1991**, *30*, 8441.
(42) Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. *J. Am. Chem. Soc.* **2001**, *123*, 8730.
(43) Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. *Structure* **2002**, *10*, 369–381.
(44) Marinelli, F.; Pietrucci, F.; Laio, A.; Piana, S. A kinetic model of Trp-cage folding from multiple biased molecular dynamics simulations. *PLoS Comput. Biol.* **2009**, in press.
(45) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339–1350.

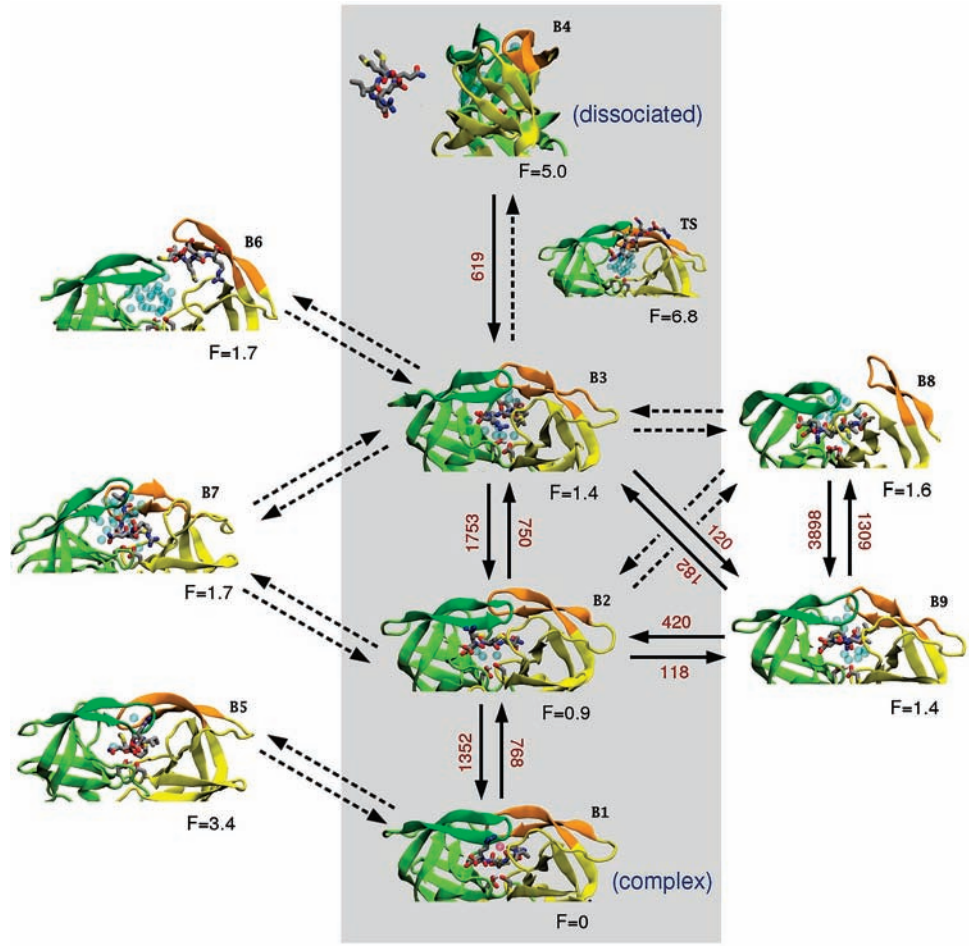

Figure 1. Representative structures of the free-energy clusters (B1−B4) involved in the main binding/unbinding mechanism (shaded box), plus the transition state (TS). The clusters B6−B9, which are not involved in the main binding pathway, are also displayed. Water molecules inside the enzyme cavity are displayed as blue spheres, except for W301 which is a red sphere. The free energy of each state, in kcal/mol, is reported below the corresponding structure. Arrows are labeled with the corresponding transition rates (ms^{-1}) when larger than 100 ms^{-1} (solid line). Transitions with a smaller associated rate are depicted as dashed arrows.

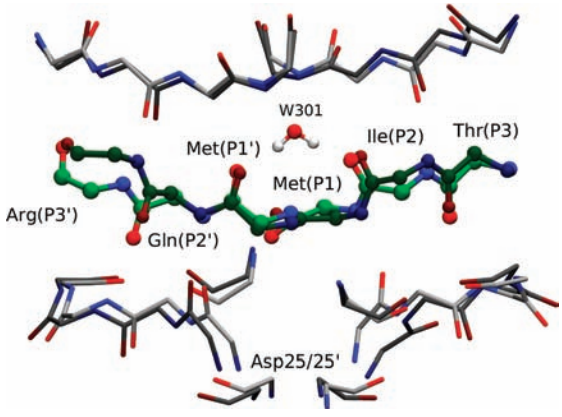

Figure 2. Comparison of the lowest free-energy structure from simulations (B1 in Figure 1 and Table 1, dark colors) with the experimental structure of the complex between the inactive D25N protease and substrate ATIM-MQRG (pdb code 1KJ7, light colors).[43] Backbone atoms are displayed. The substrate is shown in green. W301 is the crystallographic water molecule bridging between flap tips and the substrate.

while it deteriorates the accuracy of the thermodynamic and kinetic models. Similarly, the analysis has been repeated adding other variables and deriving the kinetic model in five or six dimensions. Also, this leads to larger errors but no qualitative changes in the overall picture. The error on the bin free energies, estimated according to standard weighted-histogram analysis as detailed in ref 44, is at most 0.5 kcal/mol. The reliability of the error estimate is verified by comparing the bin free energies obtained from BE-META simulations with those obtained from an ergodic equilibrium MD simulation on trialanine in ref 44.

The kinetic model has been constructed assigning the rates for transitions between each bin and its first and second nearest neighbors, following the procedure in ref 44. The transition rate between two bins α and β is assumed to depend only on the free-energy difference $\Delta G_{\alpha\beta}$ and on the diffusion matrix D_{ij} in the space of CVs C, D, E, and F. D is estimated from a 10 ns MD simulation started in the HIV-1 PR/SUB complex by maximizing the likelihood of the trajectory within the kinetic model.[46] For a time lag of 100 ps, the following values are obtained: $D_{CC} = 0.27$, $D_{DD} = 0.0011$, $D_{EE} = 0.0011$, $D_{FF} = 0.35$, $D_{CD} = -0.0033$, $D_{CE} = -0.0013$, $D_{CF} = -0.073$, $D_{DE} \approx 0$, $D_{DF} = 0.0034$, and $D_{EF} = -0.0021$. By enlarging the time lag to 200 ps, the coefficients vary by less than 16%, indicating that on this time scale a Markovian behavior is attained.[44] The position dependence of D has been investigated by computing D also from 10 ns MD trajectories started in each of the other metastable states along the main binding pathway (states B2, B3, and B4 in Figure 1). The maximum variation in the diagonal elements of the diffusion matrix is only 30%, which leads to similar variations in the relaxation times of the system. By comparison, the maximum estimated error on the bin free energies is 0.5 kcal/

(46) Hummer, G. *New J. Phys.* **2005**, *7*, 34.

Table 1. Description of the Kinetic Clusters (Local Free-Energy Minima) Explored by Bias-Exchange Metadynamics Simulations

basin	ΔG (kcal/mol)	SUB	flaps	water in cavity	comments
B1	0.0	in	closed	1 above SUB (W301)	crystallographic structure of MVT-101
B2	0.9	in	closed	3−4 under SUB and laterally	
B3	1.4	in	closed	7−10 all around SUB	
B4	5.0	out	closed	SUB fully solvated	
(TS)	6.8	out	closed	SUB almost fully solvated	H-bonds Arg(P3′)-Glu35, Thr(P3)-Gly48′, Met(P1′)-Gly49′, Met(P1′)-Gly51′
B5	3.4	in	closed	1 between flap tips	similar to B1, but W301 moved between flap tips
B6	1.7	half in	one up	SUB almost fully solvated	SUB between flaps and loop P79-T80−P81-V82
B7	1.7	in	open laterally	>10 above SUB and laterally	
B8	1.6	in	one up	>10 above SUB and laterally	SUB conformation similar to B1
B9	1.4	in	one up	SUB almost fully solvated	

mol, which leads to almost 1 order of magnitude larger variations on the computed transition rates. Therefore, the position dependence of D is neglected.

The metastable states (local minima) of the network of bins have been found by the Markov cluster analysis (MCL) algorithm[47] using $p = 1.2$. This allows identifying nine kinetic clusters (free-energy basins), corresponding to metastable states of the enzyme−ligand system, which together include 99.1% of the equilibrium population. These states have been characterized by the atomic structures corresponding to the kinetic attractors (Figure 1, Table 1). Within each of the clusters B1, B2, and B3, atomic structures differ by C_α rmsd of maximum 2.5 Å. Using a smaller MCL parameter $p = 1.15$ brings into coalescence the structurally distinct clusters B1, B2, and B9 in Figure 1 (which differ for the amount of water in the enzyme cavity or for the opening of the flaps). A larger $p = 1.3$ leads to the fragmentation of cluster B2, which is structurally homogeneous (C_α rmsd $\leq$ 2 Å).

To allow comparison with kinetic experiments, the long-time scale dynamics of the system has been modeled on the network of bins by generating a kinetic Monte Carlo (KMC)[48] trajectory of 100 s. The trajectory, starting from the complex, performs $\approx$ 6000 transitions between the complex and the dissociated state. From the analysis of the trajectory, the most probable binding/unbinding pathways between selected states are identified as the lowest free-energy paths. Transition rate constants between a pair a and b of bins are computed as $k_{a\to b} = N_{a\to b}/P_a t_{\mathrm{tot}}$, where $N_{a\to b}$ is the number of transitions; P_a is the probability of state a; and t_{tot} is the duration of the trajectory. This procedure gives directly the dissociation rate constant (units s^{-1}) taking a as the bin corresponding to the experimental complex (lowest free energy in cluster B1) and b as the bin corresponding to the dissociated pair (lowest free energy in cluster B4). To compare with experimental data at standard conditions, the association rate constant (units $M^{-1}\ s^{-1}$) is obtained by scaling $k_{b\to a}$ with the ratio between the standard concentration (1 M) and the simulated one (0.2 M), which is equivalent to scaling the probability of the dissociated state by the inverse of the ratio. The transition rates have been reduced by a factor 2.26 to correct for the self-diffusion coefficient of TIP3P water which is larger than the experimental one.[49]

2.3. Poisson−Boltzmann Calculations. To check the effect of the finite simulation box on the free energies computed from BE-META, (linearized) Poisson−Boltzmann calculations have been performed on several HIV-1 PR/SUB structures using the program apbs.[50] The following parameters have been used: grid spacing 0.45 Å, ion exclusion radius 2.2 Å, solute dielectric constant 2, continuum solvent dielectric constant 78.5, boundary conditions based on focusing, solute surface defined by a probe sphere of radius 1.4 Å.

The free-energy correction (at zero ionic strength) to bring the ligand from state B4 (see Figure 1) to infinity is $<k_BT$, which confirms that SUB does not sizably interact with HIV-1 PR in this state.

In the same manner, it has been checked by a Poisson−Boltzmann calculation that insertion of one Cl^- ion in the cavity in the absence of SUB requires 4.0−12.0 kcal/mol at ionic strength in the range 0−0.1 M, due to the negative polarization of the catalytic cavity. Indeed, the Cl^- counterions never enter the enzyme cavity during the BE-META simulations.

3. Results

3.1. Binding and Unbinding Processes. Bias-exchange metadynamics (1.6 μs) (BE-META) simulations have been performed to investigate the free-energy landscape and the mechanism of substrate binding to and dissociation from HIV-1 PR. We focused on the substrate Thr-Ile-Met-Met-Gln-Arg (hereafter SUB). The free energy has been computed as a function of seven collective variables (CVs) (see Methods). Such a large number of CVs is employed in an attempt at including all degrees of freedom which may be important for the process. The CVs describe the geometrical approach of the substrate, the opening of the flaps, and the physical interactions that might play a role in stabilizing the complex and the transition states.

From the BE-META trajectories, a kinetic model has been constructed based on the weighted-histogram approach described in the Methods section. The calculated lowest free-energy path passes through the following states (Figure 1, Table 1):

• **B4**. SUB is solvated without any contact with the enzyme; the flaps are quite closed (see below).

• **TS**. SUB outside the cavity, perpendicular with respect to the orientation in the complex, with Thr(P3) close to the cavity. H-bonds are formed first with Asp30′, then with Gly48′, Gly49′, and Gly51′ in one flap and with Glu35 (salt bridge with Arg(P3′)) on the loop in front of the cavity (Figure 3).

• **B3**. The cavity, enlarged by a moderate displacement of the flaps, allows SUB to enter (starting from Thr(P3)) together with a solvation shell.

• **B2**. The water molecules bridging between SUB and the flap tips are expelled, tightening the cavity.

• **B1**. The water molecules bridging between SUB and the catalytic dyad are expelled. A water molecule (W301) is introduced between SUB and Ile50-Ile50′ on the flap tips. The experimental complex is formed (SUB + cavity all-atom root-mean-square deviation (rmsd) 1.6 Å compared to experimental structure,[43] backbone rmsd = 0.9 Å. Cavity is defined as residues within 4.5 Å from SUB. See Figure 2).

The most populated state (46.9%) is the HIV-1 PR/SUB complex B1 (Figure 1). This result is a prediction of the experimental structure and stability of the complex. The equi-

(47) Gfeller, D.; Rios, P. D. L.; Caflisch, A.; Rao, F. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1817–1822.
(48) Voter, A. F. Introduction to the Kinetic Monte Carlo Method. In *Radiation Effects in Solids*; Sickafus, K. E., Kotominv, E. A., Eds.; Springer. NATO Publishing Unit: Dordrecht, The Netherlands, 2005.
(49) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2001**, *114*, 363.
(50) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037–10041.

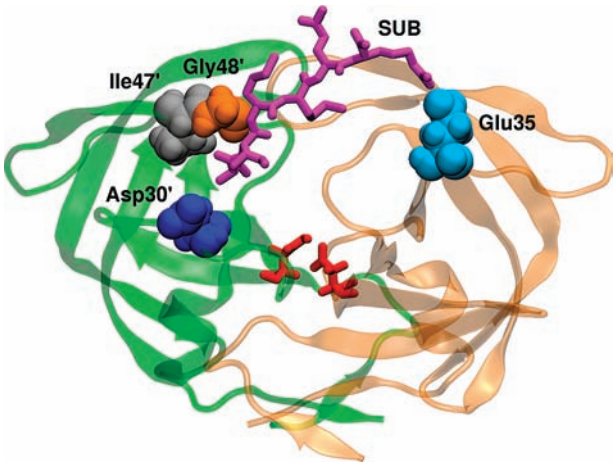

Figure 3. Protease residues forming H-bonds with the substrate in either the transition state or the earliest stage of binding (between B4 and TS in Figure 1) are labeled. The catalytic dyad Asp25-Asp25′ is shown in red.

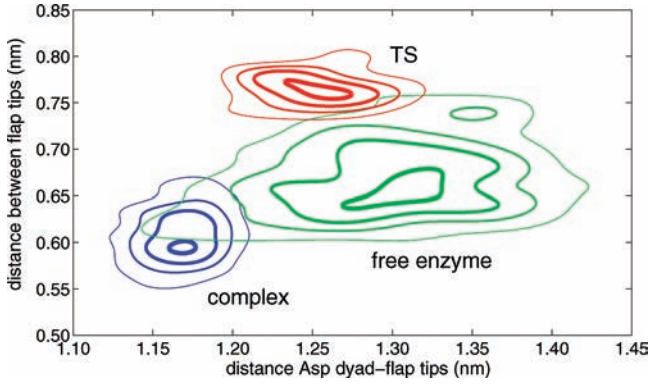

Figure 4. Probability distribution of HIV-1 PR conformations in the absence of SUB (green), in complex with SUB (blue), and in the transition state (TS) for binding (red). The two variables employed are the distance between Asp dyad and flap tips (C_α Asp25−C_α Gly49) and the distance between flap tips (variable S_E, see Methods). Equally spaced isoprobability lines are shown.

librium population of the cluster corresponding to the enzyme separated from the ligand (B4 in Figure 1) is instead almost negligible.

The values taken by the reaction coordinates along the pathway are reported in Figure 5. During the binding process, the flap tips undergo moderate displacements, without fully opening: the distance between their tips (S_E) fluctuates between 0.55 nm (bound complex B1) and 0.83 nm (transition state TS, Figure 5 and Figure S1 in Supporting Information), and the distance between the Asp dyad and flap tips varies between 1.1 and 1.8 nm (all distances are referred to C_α's). The displacements are asymmetric within the flap pair: one flap lifts above the cavity more than the other while approaching the transition state, interacting directly with the ligand. An asymmetric role of the flap tips upon binding is also suggested from crystallographic data of the complex with the NC-p1 substrate.[51] The most probable binding pathway is reversible: the unbinding pathway does not show sizable differences.

3.2. Thermodynamics and Kinetics of the Binding Process. Our model of the binding and unbinding processes allows computing all relevant thermodynamic and kinetic parameters, which can be compared with experimental data.

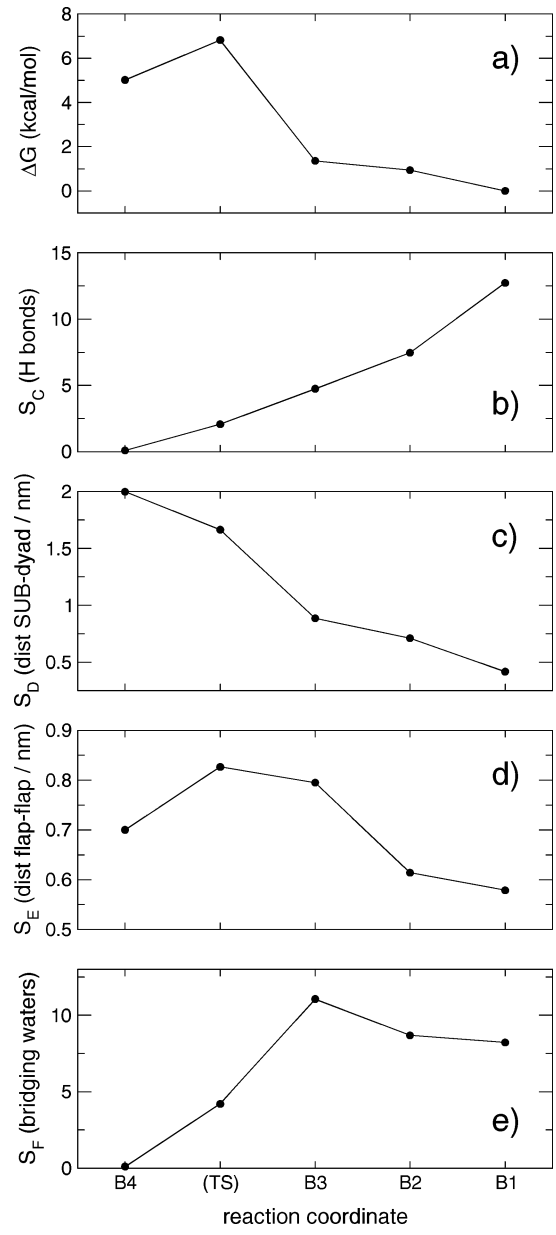

Figure 5. Variation of the free energy (panel a) and of selected CVs (panels b−e) along the main pathway for binding and unbinding.

The predicted binding free energy is $\Delta G_b \sim F_{B1} - F_{B4} = -5.0$ kcal/mol, with a statistical error of 0.5 kcal/mol (see Methods). The value of ΔG_b has been obtained for a simulated molarity of 0.2 M. Correcting for normal conditions (1 M) gives $\Delta G_b = -6(1)$ kcal/mol. Corrections due to the finite size of the simulation box are instead small: indeed by a Poisson−Boltzmann calculation it can be shown that in state B4 SUB is practically not interacting with the enzyme (see Methods). The final estimate $\Delta G_b = -6(1)$ kcal/mol compares well with $\Delta G_b^{exp} = -8.1$ kcal/mol measured on the peptidomimetic inhibitor MVT-101[52] which is similar to SUB. Comparison with other inhibitors whose thermodynamic data are available[3] is not reported here as their Tanimoto shape and electrostatic similarity[53] with SUB are significantly smaller.

(51) Prabu-Jeyabalan, M.; Nalivaika, E.; Romano, K.; Schiffer, C. *J. Virol.* **2006**, *80*, 3607.
(52) Verkhivker, G.; Appelt, K.; Freer, S. T.; Villafranca, J. E. *Protein Eng.* **1995**, *8*, 677.
(53) *OEChem, version 1.3.4*, OpenEye Scientific Software, Inc.: Santa Fe, NM, USA, www.eyesopen.com, 2005.

Also, the kinetics predicted by the model can be compared with experiments. Fluorometric assay data are available for peptide substrates similar to SUB (= TIM*MQR): values of $k_{cat} = 41\ s^{-1}$ and $K_M \equiv (k_{off} + k_{cat})/k_{on} = 3.0 \times 10^{-3}$ M have been reported for the substrate ATIM*MQRG (pH = 5.5, ionic strength 0.4 M),[54] while $k_{cat} = 6.99\ s^{-1}$ and $K_M = 2.5 \times 10^{-4}$ M have been reported for TNSATIM*MQRGNF (pH = 5.5).[55] The rate constants for binding and unbinding transitions, k_{on} (B4 → B1) and k_{off} (B1 → B4), have been computed from a 100 s kinetic Monte Carlo trajectory (see Methods) and corrected for a 1 M concentration. The result is $k_{on} = 1.26 \times 10^6\ M^{-1}\ s^{-1}$ and $k_{off} = 57.1\ s^{-1}$. Employing the experimental values of k_{cat}, a theoretical K_M in the range $1.4-2.0 \times 10^{-5}$ M is obtained. The kinetic rate constants k_{on} and k_{off} are directly accessible from biosensor and fluorescence experiments on inhibitors, which are not cleaved by HIV-1 PR. For the inhibitor MVT-101, analogue to SUB, $k_{on} = 1.6 \times 10^5\ M^{-1}\ s^{-1}$ and $k_{off} = 0.2-0.4\ s^{-1}$ have been reported (pH = 5.5).[6] These values are compatible with our theoretical estimates for SUB, considering the different ΔG_b (−8.1 kcal/mol for MVT-101 and −6.0 kcal/mol for SUB). In fact, FDA-approved peptidomimetic inhibitors of similar size as SUB, binding more effectively than MVT-101, have ΔG_b between −12 and −15 kcal/mol[3] and consistently display lower k_{off} values, in the range $10^{-4}-10^{-3}\ s^{-1}$, whereas the k_{on} values are similar to MVT-101, in the range $10^5-10^6\ M^{-1}\ s^{-1}$.[56] It must also be considered that the kinetic constants display a strong dependence on the experimental conditions like pH and ionic stregth: in our simulation, the protonation of residues corresponds to pH = 7 and the ionic strength is zero.

3.3. States with Extensive Flap Opening. The dynamics of flap opening in HIV-1 PR has been extensively studied, due to its possible functional role.[4−7,14,15] Indeed, substrates of the protease with open flaps are crucial to allow binding of the long viral polyproteins, for simple topological reasons connected to the large size of the substrate. However, our results show that full opening of the flaps is not necessary to bind the smaller SUB ligand. Our model includes also several states with open flaps (B6−B9 in Figure 1 and Table 1), which are not part of the most probable binding/unbinding pathway of SUB but which may be relevant for the viral polyprotein; e.g., state B9 is similar to B2, but it has a larger distance between flap tips (1.1−1.8 nm, compared to 0.4−1.1 nm in B2), and it has a free energy 0.7 kcal/mol higher. In states B6 and B7, the flaps act as "tweezers" which trap the ligand in between. The binding rate associated to the pathway in which SUB approaches the enzyme cavity through wide-open flaps is at least 2 orders of magnitude smaller than that associated with the lowest free-energy pathway, in which the flaps are closed.

The free energy of HIV-1 PR as a function of the distance between flap tips (CV S_E), restricted to configurations with the ligand far from the cavity, shows a preferred distance of 0.6−0.75 nm. A free energy of 1 kcal/mol is required for opening to more than 1 nm, consistent with refs 5, 14, and 15. Instead, the complex shows tighter flap tips with a distance of 0.55−0.65 nm (Figure 4 and Figure S1 in Supporting Information). The associated and dissociated states are also distinguished by a different elevation of the flap tips (C_α of Gly49-Gly49′) above the catalytic dyad (C_α of Asp25-Asp25′): 1.1−1.2 nm for the complex and 1.2−1.4 nm for the free enzyme, with the latter displaying also a larger flexibility of the flaps (see Figure 4). These two different conformations compare well, respectively, to the closed and semiopen structures observed in liganded (e.g., 4HVP) and unliganded (e.g., 1HHP) crystal structures. In agreement with recent solution NMR studies,[5] fully open flaps have a negligible population.

3.4. Hydration Inside the Enzyme Cavity. The crystallographic structure of HIV-1 PR in complex with most peptidomimetic inhibitors shows a water molecule (usually named W301) tetrahedrally hydrogen-bonded between the flap tips and the ligand.[2] W301 is also part of the calculated Michaelis complex (Figure 2 and state B1 in Figure 1). Relocation of this water to a position between the flap tips (stabilized by H-bonds with Gly51, Gly52, Gly49′, Ile50′, Gly52′) is associated with an increase of free energy of 3.4 kcal/mol. This state (B5 in Figure 1) is not involved in the main binding pathway. A previous Free Energy Perturbation (FEP) calculation[57] predicted a cost of 3.1 ± 0.6 kcal/mol for the displacement of W301 in the HIV-1 PR/KNI-272 complex, while a different FEP calculation on HIV-1 PR/MVT-101[58] displayed a negligible free-energy difference. Care should be taken in comparing the free-energy value in the latter calculations with our results due to differences in the setup and in the reference states.

In summary, the present approach allows identifying also states that are not involved in the most likely binding/unbinding pathway. Most of these states are already described in the literature and may play a role for binding the long polyprotein cleaved by the enzyme.

4. Discussion

The binding and unbinding mechanism of wild-type HIV-1 PR to a model peptide substrate (Thr-Ile-Met-Met-Gln-Arg) has been investigated by molecular dynamics simulations using an all-atom force field with explicit water. To enhance the sampling of the configuration space, the bias-exchange metadynamics technique[27] has been employed. The free energy has been computed as a function of seven reaction coordinates. These have been chosen in an attempt at capturing the most important degrees of freedom associated with binding: the approach of the ligand, the opening of the flaps, the hydrophobic and electrostatic interactions between ligand and enzyme, and the amount of interfacial water molecules. Each one of the seven reaction coordinates has been biased on a different replica of the system, until convergence in the reconstructed free-energy projections has been achieved. Starting with the ligand far from the enzyme, several binding events have been observed leading to the Michaelis complex. The substrate slides inside the cavity starting with the Thr head and without sizably opening the flaps. The dissociation and association pathways turn out to be very similar.

The simulation data have been used to build a detailed thermodynamic and kinetic scheme of the binding and unbinding processes. These feature several intermediate states (Figure 1 and Table 1). The estimated statistical error on the calculated free energies is about 0.5 kcal/mol. The binding free energy has been estimated as −6.0 kcal/mol. This compares well with the experimental value of −8.1 kcal/mol measured with the

(54) Maschera, B.; Darby, G.; Palu, G.; Wright, L. L.; Tisdale, M.; Myers, R.; Blair, E. D.; Furfine, E. S. *J. Biol. Chem.* **1996**, *271*, 33231.
(55) Schock, H. B.; Garsky, V. M.; Kuo, L. C. *J. Biol. Chem.* **1996**, *271*, 31957.
(56) Shuman, C. F.; Markgren, P. O.; Hamalainen, M.; Danielson, U. H. *Antiviral Res.* **2003**, *58*, 235.
(57) Hamelberg, D.; McCammon, J. A. *J. Am. Chem. Soc.* **2004**, *126*, 7683.
(58) Johnson, E. C. B.; Malito, E.; Shen, Y.; Pentelute, B.; Rich, D.; Florian, J.; Tang, W.; Kent, S. B. H. *J. Mol. Biol.* **2007**, *373*, 573.

structurally similar peptidomimetic inhibitor MVT-101. The calculated absolute rate constant for binding is 1.3×10^6 M^{-1} s^{-1} and for unbinding is 57 s^{-1}, consistent with experiments on MVT-101 and on peptide substrates similar to SUB.

Flap opening is usually considered to play an important role in the binding of the natural polyprotein substrate. Instead, the extent to which flaps open upon the binding of a small substrate is not clear.[59] Here we found that in the most probable association and dissociation pathways the flaps of the protease do not open sizably, and the substrate threads inside the enzyme cavity from the tight lateral channel. When the ligand approaches the enzyme in the early stages of binding, and in the transition state, it forms hydrogen bonds with residues Asp30′, Ile47′, Gly48′, and Glu35, which all lie around the opening of the channel leading to the catalytic cavity (Figure 3). It is remarkable that mutation of residues in this region causes resistance to some FDA-approved peptidomimetic drugs.[60] We here propose an argument which might help rationalize the origin of such drug resistance, by making the plausible assumption that the binding process observed for SUB is similar to that of other peptidomimetic inhibitors. In fact, the viral polyprotein cleaved in the biological function of the enzyme is expected to bind through a different pathway. It may approach the cavity from above, after the flaps have opened completely, due to the very large substrate size. Thus, mutations are likely to affect in a different manner the barrier along the two pathways and may change the relative binding efficiency of drugs and substrate even without affecting their relative affinity. A possible confirmation to our hypothesis would be provided by kinetic experiments measuring the association rate of noncleavable peptides of increasing length. For longer peptides, the preferred binding pathway should switch from the closed flaps- to the open flaps-one, with a resulting significant decrease of the association rate.

A crude estimate of the effect of mutations from a polar residue to an apolar one is provided by alanine scanning.[61,62] We applied this method on the reference structures of TS, B1, B6, and B8 (the latter two having open flaps, Figure 1). Interestingly, Glu35, whose mutation to Gly is indicated as generating resistance to a few FDA-approved drugs, is found to be a "hot spot"[61,62] only in TS. This suggests, at a speculative level, that this mutation may play a role in the binding pathway. To quantitatively assess this scenario, free-energy calculations of the binding mechanism of inhibitors to drug-resistant variants would be required.

Our results show that, due to sizably different conformations of the flaps, the conformations of HIV-1 PR in the absence of the ligand, in the transition state for binding, and in the complex have little overlap (see Figure 4). This confirms the appropriateness of docking protocols which account for the flexibility of the receptor by using an ensemble of target enzyme conformations (e.g., relaxed complex scheme).[63]

Individual water molecules at the interface between ligand and enzyme play a pivotal role throughout the binding process, and they constitute a relevant reaction coordinate which enables differentiating the intermediate states explored by the system. Therefore we stress the importance of treating explicitly and accurately the solvent molecules in computational studies of protein−ligand binding.

In conclusion, in this work we have provided the detailed binding process of a substrate to HIV-1 PR by atomistic simulations with explicit solvent. Alternative pathways are analyzed, and the most probable is found. The knowledge of the full binding pathway of HIV-1 PR, together with the associated energetics, may help rationalize the mechanism by which mutations located at the entrance of the lateral channel induce resistance to drugs without destroying the biological function.

Acknowledgment. The authors acknowledge Vincenzo Carnevale and Stefano Piana for kindly providing the atomic coordinates of the HIV-1 PR/SUB complex. We also acknowledge the grant MIUR PRIN-2006025255 and the Australian Partnership for Advanced Computing for providing computational resources. The authors are grateful to Vanessa Leone for performing the alanine scanning analysis of mutations. FP. thanks Xevi Biarnes, Attilio V. Vargiu, Agata Kranjc, and Fernando Herrera for useful discussions and suggestions.

Supporting Information Available: Figure S1. This material is available free of charge via the Internet at http://pubs.acs.org.

JA903045Y

(59) Hornak, V.; Simmerling, C. *Drug Discovery Today* **2007**, *12*, 132–138.
(60) Shafer, R. W. *J. Infect. Dis.* **2006**, *194*, S51–S58.
(61) Kortemme, T.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14116–14121.
(62) Kortemme, T.; Kim, D. E.; Baker, D. *Sci. STKE* **2004**, *219*, pl2.
(63) Amaro, R. E.; Baron, R.; McCammon, J. A. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 693–705.