HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations

Viktor Hornak*, Asim Okur†, Robert C. Rizzo§, and Carlos Simmerling*‡¶

Departments of *Chemistry and †Chemistry and ‡Applied Mathematics and Statistics and *Center for Structural Biology, Stony Brook University, Stony Brook, NY 11794; and §Computational Science Center, Brookhaven National Laboratory, Upton, NY 11973

Edited by Janet M. Thornton, European Bioinformatics Institute, Cambridge, United Kingdom, and approved November 18, 2005 (received for review September 27, 2005)

We report unrestrained, all-atom molecular dynamics simulations of HIV-1 protease that sample large conformational changes of the active site flaps. In particular, the unliganded protease undergoes multiple conversions between the “closed” and “semiopen” forms observed in crystal structures of inhibitor-bound and unliganded protease, respectively, including reversal of flap “handedness.” Simulations in the presence of a cyclic urea inhibitor yield stable closed flaps. Furthermore, we observe several events in which the flaps of the unliganded protease open to a much greater degree than observed in crystal structures and subsequently return to the semiopen state. Our data strongly support the hypothesis that the unliganded protease predominantly populates the semiopen conformation, with closed and fully open structures being a minor component of the overall ensemble. The results also provide a model for the flap opening and closing that is considered to be essential to enzyme function.

AIDS | HIV protease | molecular dynamics simulations | protein dynamics

Due to its central role in processing viral polypeptide precursors, HIV-1 protease (HIV-PR) continues to be one of the primary targets of anti-AIDS drug discovery. A greater understanding of the mechanistic events associated with binding of HIV-PR substrates and inhibitors is critical for the design of more potent inhibitors of the enzyme. An extensive set of x-ray crystal structures of HIV-1 protease have been solved (1), revealing a C2 symmetric homodimer with a large substrate-binding pocket covered by two glycine-rich β-hairpins, or flaps (2–4). Consistent structural differences are present between the bound and free states of the protein (Fig. 1). In all of the liganded forms, the flaps are pulled in toward the bottom of the active site (“closed” form), whereas the structures for the unbound enzyme all adopt a “semiopen” conformation with the flaps shifted away from the dual Asp-25-Thr-26-Gly-27 catalytic triads but still substantially closed over the active site and in contact with each other. A more striking difference between the two forms is that the relative orientation (the “handedness”) of the β-hairpin flaps is reversed (Fig. 1). The nonflap residues show only slight variation. Although large-scale flap opening is presumably required for normal substrate access to the active site, no crystallographic structures representing such an open conformation have been reported. NMR experiments (5, 6) have established the flexibility of the flap region, suggesting that closed, semiopen, and fully open conformations of the protease are in dynamic equilibrium, with the semiopen form being prevalent for the free protease. The computational results presented here strongly support this hypothesis; the correct semiopen structure dominates our computational ensemble of unbound HIV-PR, but all three structurally distinct forms are observed to interconvert during the simulations.

Understanding the issues that govern HIV-PR flap mobility has profound implications for elucidating the detailed mechanism of this enzyme and in the design of new therapeutic agents, such as allostERIC inhibitors intended to interfere with flap opening and thereby with enzymatic function. Thus, numerous prior computational studies have aimed at understanding flap opening dynamics. Collins et al. (7) reported results from activated molecular dynamics (MD) simulations in the gas phase that involved forcing the atomic coordinates for nonflap regions of a closed structure to the semiopen state. These simulations produced flap opening, but the use of restraints prevented investigation of coupling between flap and core domains (8). Scott and Schiffer (9) also observed irreversible flap opening after 3 ns of a 10-ns MD simulation in explicit water, but the extent of flap opening was not quantitatively described. Instead they focused on the flap tip regions, which “curled” back into the protein structure during the opening event, burying several hydrophobic residues. This flap curling was hypothesized to provide a key conformational trigger necessary for subsequent large-scale flap opening and therefore HIV-PR function. More recently, Hamelberg and McCammon (10) used activated dynamics to produce flap opening in HIV-PR. Perryman et al. (11) reported dynamics of unbound V82F/I84V mutant in which the closed form opened somewhat, but detailed analysis of flap conformation and handedness was not reported.

One drawback to prior computational studies is that none have reported observing the return of these open states back to either of the crystallographic forms, and thus their relevance to the true dynamics of HIV-PR is unclear; the simulations may have been too short, or perhaps the observed opening was an artifact of force field inaccuracies or the impulse forces some studies used to aid sampling. A recent study by Meagher and Carlson (12) suggested that large conformational changes within the first few nanoseconds of HIV-PR simulation may simply be artifacts of improper equilibration. Likewise, the transition between closed and semiopen flap handedness that has been experimentally characterized has not been reported in unrestrained simulations.

These uncertainties reinforce the need for simulations of flap opening under reversible conditions that reproduce experimental structural data without imposing restraints or unphysical forces that may obscure dynamic correlations involved in true opening events. We report here a series of fully unrestrained standard molecular dynamics simulations that were initiated from both closed and semiopen forms of HIV-PR. NMR data suggest that these large-scale flap motions occur on the microsecond–millisecond timescale (13); previous simulations typically modeled only several nanoseconds of dynamics. To facilitate dynamic events on a computationally feasible timescale, we used a continuum solvent representation with a reduced viscosity. It has been shown previously that peptide folding rates are accelerated in low viscosity implicit solvent simulations (14).

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: HIV-PR, HIV-1 protease; RMSD, rms deviation; MD, molecular dynamics; PDB, Protein Data Bank.

†To whom correspondence should be sent at the † address. E-mail: carlos.simmerling@stonybrook.edu.

© 2006 by The National Academy of Sciences of the USA

www.pnas.org/cgi/doi/10.1073/pnas.0508452103

PNAS | January 24, 2006 | vol. 103 | no. 4 | 915–920
of the ligand. We performed an analysis of that structure set along with crystal structures of unbound HIV-PR to investigate specific differences between closed and semioopen forms. Even though approximately one-third of the closed structures have a hydrogen bond between the flap tips, nearly the same number form a different flap tip hydrogen bond and the remainder, like the semioopen structures, have none. Likewise, the flap tip β-turn type is not a unique characteristic of the closed conformation. It is likely that the resolution of the crystallographic data, coupled with intrinsic flap tip flexibility, is insufficient to unambiguously define this region. We found that the most reliable measure of closed vs. semioopen form exploits the consistent difference in the handedness of the flaps (Fig. 1), which results in large rms deviation (RMSD) values when flap residue α-carbons are overlapped; this RMSD value between the closed and semioopen forms is 3.3 Å.

Unlike the closed and semioopen forms, a structure of the protease with fully open flaps has not yet been observed experimentally; thus, it is somewhat arbitrary to define which structures qualify as open. The open ensemble most likely consists of many structures with different relative positioning of both flaps. However, because the proposed mechanism of ligand binding requires extensive flap opening, the distance between flap tips (Ile-50C- Cile-50’C) is a reasonable metric to gauge the extent of opening. This distance is ≈4.3 Å in semioopen and ≈5.8 Å in closed forms (somewhat longer in the closed form, see Fig. 1).

MD Simulations of Inhibitor Bound HIV-PR. A 28-ns simulation of closed, inhibitor-bound HIV-PR at 300 K was initiated from the crystal structure of the complex and produced a stable trajectory. No large structural changes were observed; the flaps remained in their closed conformation throughout the entire run (see Fig. 3A). The hydrogen bond between the cyclic urea inhibitor’s carbonyl oxygen and at least one of the Ile-50/50’ amides as well as the hydrogen bond between flap tips (Gly-51N-Ile-50’O) remained in place, likely contributing to the stability of the otherwise flexible glycine-rich flap tips. This behavior is typical of protein simulations on this timescale.

Whereas the global structure of the protease remained stable (Cα RMSD plateau < 2.0 Å) (Fig. 7, which is published as supporting information on the PNAS website), certain regions exhibited higher flexibility (Fig. 2), including the termini, flap elbows (residues 37–42), cantilever β-turn region (residues 66–69), and two narrow regions around Gly-17 and Thr-80. Importantly, the flaps and flap tips did not show higher fluctuations than the core during simulation of the complex.

MD Simulations of Ligand-Free HIV-PR. A 42-ns simulation of ligand-free HIV-PR at 300 K was also initiated from the crystallographic coordinates of the complex, that is, after removal of the inhibitor. This procedure permits use of the bound simulations as a direct control. The flaps spontaneously rearranged (Fig. 3B) from the closed to semioopen form near the beginning of the simulation (including reversal of flap handedness), and exhibited significantly increased fluctuations (Fig. 2, orange line). This finding is consistent with experimental data that indicate higher flexibility and predominance of the semioopen form in the absence of ligands. Of particular importance is the revisiting of closed flap conformations (Fig. 3B) before returning to the semioopen form and transient sampling of flap configurations that are neither semioopen nor closed. Thus a ligand-free HIV-PR simulation initiated from the closed crystal structure spontaneously converts to a more flexible ensemble that is dominated by the semioopen form yet also visits structures that are closed and others that are neither semioopen nor closed. This behavior is in sharp contrast with the stability of the same closed state under equivalent conditions in the presence of the inhibitor.

Consistent with experiment, only the closed form is observed when the inhibitor is present. When the inhibitor is removed from this structure, the protease spontaneously converts to a more flexible and semioopen state with appropriate reversal of flap handedness. Similar to previous studies, we also observe spontaneous and large-scale opening of flaps in multiple simulations of unbound HIV-PR. Importantly, these open states are transient and subsequently convert back to the semioopen state. Our results are consistent with available x-ray and NMR experimental data and provide a model for the open state of the enzyme that is required for ligand entry to the active site and for the dynamic interconversion between this open state and the closed and semioopen forms that have been observed experimentally.

Results and Discussion

Difference Between Closed, Semioopen, and Open States of HIV-PR. Although the conformation of the flap region clearly differs in the closed and semioopen forms (Fig. 1), we desired a simple metric to monitor the structure during simulation. A comparative study of 73 HIV-PR complexes by Zoete and Karplus (15) showed that the structure is relatively insensitive to the identity of the ligand. We performed an analysis of that structure set along with crystal structures of unbound HIV-PR to investigate specific differences between closed and semioopen forms. Even though approximately one-third of the closed structures have a hydrogen bond between the flap tips, nearly the same number form a different flap tip hydrogen bond and the remainder, like the semioopen structures, have none. Likewise, the flap tip β-turn type is not a unique characteristic of the closed conformation. It is likely that the resolution of the crystallographic data, coupled with intrinsic flap tip flexibility, is insufficient to unambiguously define this region. We found that the most reliable measure of closed vs. semioopen form exploits the consistent difference in the handedness of the flaps (Fig. 1), which results in large rms deviation (RMSD) values when flap residue α-carbons are overlapped; this RMSD value between the closed and semioopen forms is 3.3 Å.

Unlike the closed and semioopen forms, a structure of the protease with fully open flaps has not yet been observed experimentally; thus, it is somewhat arbitrary to define which structures qualify as open. The open ensemble most likely consists of many structures with different relative positioning of both flaps. However, because the proposed mechanism of ligand binding requires extensive flap opening, the distance between flap tips (Ile-50C- Cile-50’C) is a reasonable metric to gauge the extent of opening. This distance is ≈4.3 Å in semioopen and ≈5.8 Å in closed forms (somewhat longer in the closed form, see Fig. 1).

MD Simulations of Inhibitor Bound HIV-PR. A 28-ns simulation of closed, inhibitor-bound HIV-PR at 300 K was initiated from the crystal structure of the complex and produced a stable trajectory. No large structural changes were observed; the flaps remained in their closed conformation throughout the entire run (see Fig. 3A). The hydrogen bond between the cyclic urea inhibitor’s carbonyl oxygen and at least one of the Ile-50/50’ amides as well as the hydrogen bond between flap tips (Gly-51N-Ile-50’O) remained in place, likely contributing to the stability of the otherwise flexible glycine-rich flap tips. This behavior is typical of protein simulations on this timescale.

Whereas the global structure of the protease remained stable (Cα RMSD plateau < 2.0 Å) (Fig. 7, which is published as supporting information on the PNAS website), certain regions exhibited higher flexibility (Fig. 2), including the termini, flap elbows (residues 37–42), cantilever β-turn region (residues 66–69), and two narrow regions around Gly-17 and Thr-80. Importantly, the flaps and flap tips did not show higher fluctuations than the core during simulation of the complex.

MD Simulations of Ligand-Free HIV-PR. A 42-ns simulation of ligand-free HIV-PR at 300 K was also initiated from the crystallographic coordinates of the complex, that is, after removal of the inhibitor. This procedure permits use of the bound simulations as a direct control. The flaps spontaneously rearranged (Fig. 3B) from the closed to semioopen form near the beginning of the simulation (including reversal of flap handedness), and exhibited significantly increased fluctuations (Fig. 2, orange line). This finding is consistent with experimental data that indicate higher flexibility and predominance of the semioopen form in the absence of ligands. Of particular importance is the revisiting of closed flap conformations (Fig. 3B) before returning to the semioopen form and transient sampling of flap configurations that are neither semioopen nor closed. Thus a ligand-free HIV-PR simulation initiated from the closed crystal structure spontaneously converts to a more flexible ensemble that is dominated by the semioopen form yet also visits structures that are closed and others that are neither semioopen nor closed. This behavior is in sharp contrast with the stability of the same closed state under equivalent conditions in the presence of the inhibitor.
Unbound HIV-PR flap flexibility was mostly localized in the tips (Fig. 2); the flap β-hairpins themselves were not disrupted even during the rearrangement events (RMSD of residues 43–48 and 53–58 remained near 0.5 Å at all times). Intraflap hydrogen bonds were always present (Fig. 8, which is published as supporting information on the PNAS web site), with the exception of the tip hydrogen bonds between Gly-49 and Gly-52. This observation is in good agreement with cross-strand amide nuclear Overhauser effect data indicating that the flap β-hairpins are present in solution and are well ordered except at the tips (6). Backbone torsions underwent substantial changes in the tips and elbows (Figs. 9 and 10, which is published as supporting information on the PNAS web site). These torsion data imply that the flap β-hairpins each move as a rigid body with highly flexible tips, with the flap elbows acting as “hinges.” This flap motion generates a wide range of structures; yet, as can be seen from Fig. 3B, the overall conformation of the flap pair typically resembles either the semiopen or closed form, with other conformations forming a group of more open structures with flap tip separations as large as 30 Å. In each case, the protease returned to the semiopen form after the opening event. The large flap movements were primarily absorbed by the flap elbow, cantilever, and fulcrum regions.

Scott and Schiffer (9) proposed flap curling as a “trigger” for HIV-PR opening based on their observation of a hydrophobic cluster involving Ile-50 at the flap tip and the nearby Val-32, Pro-79, Thr-80, and Pro-81. We calculated the distance of Ile-50 Cβ to the center of mass of residues 32 and 79–81. Although some flap movements are accompanied by formation of this cluster, the more significant opening events are not; instead, they are associated with greater departure of flap tips away from the binding site (Fig. 11, which is published as supporting information on the PNAS web site). NMR relaxation experiments (6) indicate that flexibility of the flap tips (“flap curling”) is related to rapid (nanosecond timescale) interconversion between the members of the semiopen ensemble, whereas the actual opening events happen on a much slower (microseconds to milliseconds) timescale, and thus are likely not directly related to rapid flap tip fluctuations.

Comparison of Simulations with NMR Data on HIV-PR Dynamics. Torchia et al. (6, 13, 16) have proposed a model for flap dynamics in which the predominant structure in solution is semiopen with fast (≈1 ns) conformational changes in flap tips. The semiopen ensemble is in slower (~100 μs) dynamic equilibrium with closed conformations as well as with less ordered open flap conformations (6, 13), which permit access to the active site as an outcome of a rare event (16). This picture is consistent with our simulations, in which flaps rapidly interconvert between a variety of flexible semiopen conformations, with infrequent opening events when flaps separate and become less ordered. The presence of a low population of closed flap conformations in the solution ensemble also is supported by our data; simulations of unbound protease transiently converted from the predominant semiopen form to the closed one, as indicated by the reversal of flap handedness. This model also is in accord with free energy calculations of Rick et al. (17), which indicated that the closed state is less favorable than the semiopen one in the absence of a ligand.

To compare the flexibility of HIV-PR inferred from MD simulations with experimental data (5, 6, 18), generalized order parameters $S^2$ were computed for free and ligand-bound protease. The calculated results show very good qualitative accord with experiment; trends characterizing regions of high and low flexibility are reproduced quite well (Fig. 5). Flexible regions in the unliganded protease include residues 37–41 (flap elbows),...
The most pronounced difference between order parameters of free and bound protease is the loss of flexibility in the flap tip region upon formation of the complex (Fig. 5). Our simulations and experimental data match these trends very well. The fulcrum loop around Gly-17 also exhibits higher flexibility in our simulations. A protease complex with DMP323 and P9941 inhibitors does show low \( S^2 \) values in this region (5), but more recent NMR data on the complex with DMP323 did not reproduce this trend (18). This discrepancy might be caused by the elevated temperature (34°C vs. 27°C) used in the earlier Nicholson et al. experiments. Increased flexibility in this region was also reported by others using simulations with quite different solvent models and protein force fields (9, 12, 19). The source of the discrepancy between simulations and experiment for Gly-17 (fulcrum loop) flexibility is thus unclear.

Absolute \( S^2 \) values, however, are in general lower in the calculated values than in the experiment. It has been previously reported that quantitative comparison of calculated and experimental order parameters can be problematic, and higher flexibility from molecular dynamics simulations is often reported. Several issues have been discussed in detail (20); uncertainties in experimental values can arise from the assumptions applied in “model free” analysis and differences in procedures. Simulated values also depend on the force field used and convergence of the ensemble of structures sampled, leading to values that can be either too small or too large.

**Simulations with an Explicit Water Model.** Another possible explanation for lack of quantitative agreement of \( S^2 \) values is that our use of low viscosity increased the amplitude of conformational fluctuations, in addition to the more rapid structural rearrangements that we desired. To evaluate this possibility, we performed additional HIV-PR simulations using an explicit solvent model. For the inhibitor-bound complex in explicit solvent (30 ns), the \( C^\alpha \) atomic fluctuations were nearly identical to those obtained with the implicit water model (Fig. 12, which is published as supporting information on the PNAS web site), strongly suggesting that the range of motion sampled in our simulations is unlikely to be an artifact of the low viscosity, and the lack of quantitative agreement between simulated and experimental \( S^2 \)
arises from other sources. Consistent with timescales inferred from experiment, the closed unbound protease in a 45-ns explicit solvent simulation was unable to convert to semiopen or open states.

Conclusions

Using all-atom simulations with continuum solvation, we observed conversion between closed, semiopen, and fully open forms of HIV-PR, including the reversal of flap handedness between semiopen and closed forms as observed in all crystal structures. The preferred structure, as well as the extent of flexibility, depended strongly on the presence of an inhibitor. These results are in good agreement with existing NMR and crystallographic data, but the use of an artificially low solvent viscosity precludes the comparison of absolute timescales from our simulations to NMR relaxation experiments. However, this low viscosity permitted the observation of structural transitions on a computationally tractable timescale. Moreover, we do reproduce the trend of flap tip flexibility occurring on a much more rapid timescale than full opening. Future exploration of the events that we report but with potentially more accurate explicit solvent models would be highly desirable. Such studies would not only facilitate quantitative comparison of timescales but also provide useful insight into a possible structural role for water during HIV-PR dynamics. For example, a well ordered water molecule mediating interaction between the flap tips and peptide-based inhibitors is observed in multiple HIV-PR crystal structures (21, 22). Such interactions transcend bulk viscosity effects and may further reduce the likelihood of observing flap transitions in explicit solvent simulations.

Although the four available crystal structures of unbound HIV-PR all show the semiopen conformation, it has not been entirely clear whether this reflects the preferred flap conformation or crystal packing effects (21, 23). For the free protease, early MD simulations suggested that the closed form was predominant in solution (24), whereas several more recent computational studies showed irreversible opening of the flaps. Based on solution NMR data for the free protease, Torchia et al. have suggested (6, 13) that the ensemble of unbound structures is dominated by the semiopen family with subnanosecond timescale fluctuation in the flap tips and with closed structures being a minor component of the ensemble. The semiopen form is in slow equilibrium (~100 μs) with a less structured open form that exposes the binding site cavity. Our simulations of unliganded HIV-PR strongly support this latter model both in the types and relative populations of different structures.

Full flap opening in our simulations occurs through a concerted downward motion of the cantilever, fulcrum, and flap elbows that results in upwards motion of the flaps and, consequently, domain rotation around the region near the dimer interface and a shift of the catalytic Asp-25/25' up from the bottom of the active site, which could be speculated to facilitate ligand binding (Fig. 6). This mechanism is further supported by detailed comparison of the crystal structures of inhibitor-bound and free protease (Fig. 6). Although the differences in the two structures are dominated by the flap rearrangement, concerted motions of flap elbows, fulcrum, and cantilever that are consistent with our model are indeed apparent. Interestingly, early dynamic analysis by Beveridge and coworkers (8, 25) also showed correlation between the flap and cantilever domains as well as between the two cantilever domains. The change in flap handedness upon ligand binding is clearly reproduced in our simulations, but its possible role in HIV-PR function remains unexplained.

Methods

Simulations in AMBER (26) were initiated from crystallographic coordinates. The closed form used a structure of wild-type sequence in complex with the cyclic urea inhibitor XK263 (22) [Protein Data Bank (PDB) ID code 1HVR]. Semiopen coordinates were obtained from 1HHF (27), after changing two side chains (I3V and C95A) to match the sequence of 1HVR. Both catalytic Asp side chains were modeled in the protonated state. Hydrogen atoms were added by using the LEAP module in AMBER. FF99 (28, 29) was used for the protein, with backbone parameters (Table 1, which is published as supporting information on the PNAS web site) refit to ab initio energies for Ala-3 and Gly-3. The ANTECHAMBER module and GAFF (30) with AM1-BCC charges (31) were used to obtain force field parameters for XK263 (Table 2, which is published as supporting information on the PNAS web site). Order parameters (S2) were calculated from a plateau region of the N–H internuclear vector autocorrelation functions (32), which were obtained with the PTRAJ module.

Aqueous solvation was modeled implicitly by using the Generalized Born approach (33) with modification for more accurate solvation of large proteins (34). Continuum solvent simulations of HIV-PR have been reported in the past (35). Temperature was maintained by using Langevin dynamics with a collision frequency of 1 ps−1, the timestep was 1 fs, and all bonds to hydrogen were constrained. All nonbonded interactions were calculated at each time step (i.e., no cutoff was used). The system was minimized and then gradually heated from 200K to 300K. Positional restraints were used first on all heavy atoms and then only for backbone atoms. Restraint force constants were decreased from 2 to 0.1 kcal/mol Å2 (1 kcal = 4.18 kJ) in several stages. During the last step of equilibration, restraints were removed entirely, and all production simulations were at 300K and fully unrestrained. These simulations required ~400 core processing unit hours on 1.5-GHz Itanium processors for each nanosecond of dynamics. Details for explicit solvent simulations are provided in the legend to Fig. 12.

Nomenclature for HIV-PR regions adopts the convention proposed by Harte and Beveridge (8). We thank Ad Bax and Dennis Torchia for providing NMR data and Roberto Gomperts (Silicon Graphics, Mountain View, CA) for providing important Amber optimizations and Altix resources. This work was supported by the National Institutes of Health (grants GM59875, GM51350, and GM61347 to W.H.W. and R.M.K., respectively). We thank Ad Bax and Dennis Torchia for providing NMR data and Roberto Gomperts (Silicon Graphics, Mountain View, CA) for providing important Amber optimizations and Altix resources.
was supported by National Institutes of Health Grant GM6167803 and Department of Energy Grant DE-AC02–98CH10886 and by additional supercomputer time at the National Center for Supercomputing Applications under National Partnership for Advanced Computational Infrastructure MCA02N028. C.S. is a Cottrell Scholar of Research Corporation.