JCTC Journal of Chemical Theory and Computation

Conformational Motions of HIV-1 Protease Identified Using Reversible Digitally Filtered Molecular Dynamics

Adrian P. Wiley,† Sarah L. Williams,† and Jonathan W. Essex*

School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, U.K.

Received May 7, 2008

Abstract: HIV-1 protease performs a vital step in the propagation of the HIV virus and is therefore an important drug target in the treatment of AIDS. It consists of a homodimer, with access to the active site limited by two protein flaps. NMR studies have identified two time scales of motions that occur in these flaps, and it is thought that the slower of these is responsible for a conformational change that makes the protein ligand-accessible. This motion occurs on a time scale outside that achievable using traditional molecular dynamics simulations. Reversible Digitally Filtered Molecular Dynamics (RDFMD) is a method that amplifies low frequency motions associated with conformational change and has recently been applied to, among others, *E. coli* dihydrofolate reductase, inducing a conformational change between known crystal structures. In this paper, the conformational motions of HIV-1 protease produced during MD and RDFMD simulations are presented, including movement between the known semiopen and closed conformations, and the opening and closing of the protein flaps.

1. Introduction

Human immunodeficiency virus-1 protease (HIV-1 PR) performs a vital step in the life cycle of HIV and is therefore an important drug target for the treatment of acquired immunodeficiency syndrome (AIDS).[1] The enzyme is responsible for the cleavage of the viral polyproteins to yield their functional constituent proteins, a crucial process for the correct assembly and maturation of the HIV virions.

HIV-1 PR is a homodimeric enzyme of 198 residues, containing two mobile flaps that cover the active site. For convenience, the residues of the two monomers are numbered 1 to 99, and 101 to 199. The HIV-1 PR active site consists of catalytically important, spatially neighboring aspartic acid residues, 25 and 125, at the base of a hydrophobic cleft. Each flap contains two β-sheets (residues 43 to 49 and 52 to 66) connected by a flexible Gly-Gly-Ile-Gly-Gly sequence (residues 48 to 52). Access to the active site requires substantial conformational change in the flap residues, and the nature of this motion has been the topic of much research.

In this paper, an investigation into the flap motions of HIV-1 PR is presented, including molecular dynamics trajectories obtained at a range of temperatures, and reversible digitally filtered molecular dynamics (RDFMD) simulations performed using a variety of protocols. Conformational motions are discussed according to relevant experimental data, and the accessibility of the active site is reported following different opening events.

1.1. Experimental Studies of HIV-1 PR. The HIV-1 PR system has been extensively studied by experimental methods. Abundant structural data of the apoenzyme and the complex formed with a number of substrate analogues and inhibitors is available from the protein data bank (pdb)[2] and HIV structural reference database (HIVSDB).[3,4] A review of the X-ray data was performed using 73 differing structures,[5] locating rigid regions of the protein (such as the residues that surround the active site) and more flexible areas (such as the "flap elbows" around residue 40).

The flap region is seen to exhibit a semiopen conformation in the majority of experimentally determined structures of the free protein, with one structure showing the flaps in an open arrangement allowing access to the active site, obtained with the use of mutations. The flaps have been shown to be in a more ordered, closed conformation in the presence of a ligand. The semiopen (pdb structure 1HHP[6]), open (pdb structure 1TW7[7]) and closed (pdb structure 1G6L[8] observed

* Corresponding author e-mail: J.W.Essex@soton.ac.uk.
† These authors contributed equally to this work.

10.1021/ct800152d CCC: $40.75
Published on Web 03/12/2009

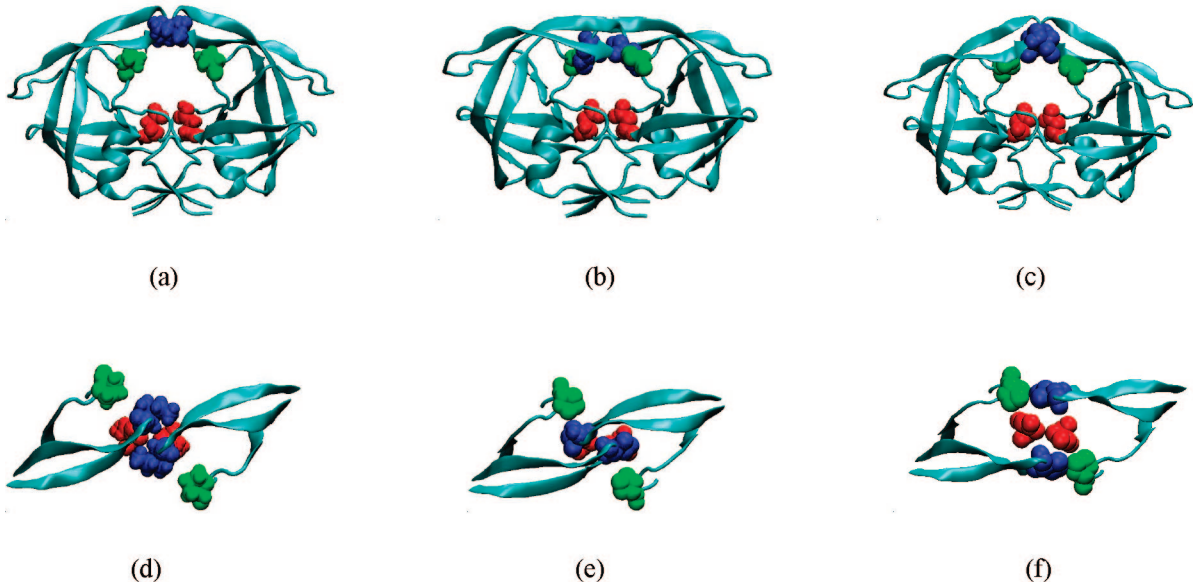

Figure 1. Cartoon representation of HIV-1 PR of semiopen (pdb 1HHP) (a and d), closed (pdb 1G6L) (b and e), open (pdb 1TW7) (c and f)) conformations. For the top view (d, e, and f), only flap and P1 loop residues are shown for clarity (residues 44 to 59, 76 to 83, 144 to 159, and 176 to 183). The van der Waals radii of particularly significant residues are displayed; the active site, residues 25 and 125, are shown in red, the flap tips, residues 50 and 150, are shown in blue, and residues 81 and 181 of the P1 loop are shown in green.

in a tethered dimer) conformations are shown in Figure 1. There is a significant change in the packing of the flap tips, which reach past each other in the closed form but cross in front of the other in the semiopen state. Also shown in Figure 1 are residues 81 and 181 which belong to the hydrophobic P1 loops (residues 79 to 81 and 179 to 181[9]) and are known to be important in flap motions.

In the inhibitor-bound state, all of the available X-ray structures show the flaps to persist in the closed conformation, with one exception, where the structure possesses a metal carborane inhibitor bound to part of the flaps, causing them to be held in an open-type conformation (pdb structure 1ZTZ[10]).

Of particular interest to this study are two NMR studies which shall be referred to in later sections of this paper. In 1999, a model of the flap mobility of HIV-1 PR was proposed by Ishima et al.[11] This was based on a detailed NMR investigation of a mutant that exhibits the same catalytic activity as the wild type protein but is stable in the concentrations required for NMR. The free enzyme was proposed to exist mainly in a semiopen state in solution, and two flap motions were detected. One occurs on a time scale well within 10 ns and involves the flap tip residues, 48 to 52. The other is a larger movement involving more of the flap residues and exists on a time scale of approximately 100 μs. The less ordered protein flaps are proposed to be in an open conformation that allows access to the active site. It was also suggested that an approaching substrate could facilitate the opening of the flaps via interaction with phenylalanine 53.

An extension of the study by Ishima et al., that includes the effects of a bound substrate, was published in 2003.[12] It was proposed that the association rate of substrates was not controlled by diffusion, but by a rare event, such as a flap opening motion. It is noted that the motions previously observed in the free protein are of limited amplitude in the substrate complex.

1.2. Theoretical Studies of Flap Mobility in HIV-1 PR. A number of computational studies of the apoenzyme have sampled a variety of flap conformations, the most successful being those which employed approaches which enhance conformational sampling.

Of studies which used conventional MD and explicit solvent, only one reports significant motion of the flaps. Scott et al. report a large opening event within 3 ns which persists until the end of the 10 ns simulation.[9] In this conformation, the flaps are lying in an asymmetric arrangement with isoleucine 50 interacting with the P1 loop, residues 79 to 81, and with isoleucine 47 and 54. This motion is stabilized by the clustering of hydrophobic sidechains, including those of valine 32, proline 79, and proline 81. Within this hydrophobic cluster is a hydroxyl group of threonine 80, proposed to keep the flap regions mobile by destabilizing the open conformation. The motion is not sampled in the presence of an inhibitor. Scott et al. claim that the simulation results are in agreement with the presence of a motion occurring in a time scale less than 10 ns, reported in the NMR study by Ishima et al. However, the motion described by Ishima et al. should be reversibly sampled well within 10 ns and should include only residues 49 to 52. It seems more likely that a motion more similar to the slower event described by Ishima et al. has been sampled in this study. This motion exists on the 100 μs time scale, in which the protein moves to a less ordered, more open state which is considered to be ligand-accessible. It is possible that Scott et al. have simulated this rare event, but the lack of a reverse path, which has not been previously reported, limits the conclusions that can be taken from this study. Other possible causes for sampling the opening motion have been suggested to be the use of the GROMOS forcefield[13] or issues with

inadequate solvation of the protein cavity.[14] A longer 22 ns MD simulation using explicit solvent, carried out by Perryman et al., does not observe this large conformational change in the WT apoenzyme and only sees one flap of a more flexible mutant move from the intial closed conformation to one similar to the semiopen state.[13]

Several simulations have been carried out using techniques which enhance sampling and these observe an increased range of conformations compared with conventional MD alone. Zhu et al. apply a constraint on the interflap distance in order to sample open conformations of the flaps.[15] The free energy profile calculated for the apoenzyme shows an initial barrier to opening, and after an opening of approximately 3 Å between the α-carbon residues, the profile is flat, suggesting a significant flexibility of the open conformation.

The transition between the closed and semiopen states in the apoenzyme was modeled using a single reaction path method by Rick et al. in 1998.[16] Although neither state allows access to the active site, the transition between semiopen and closed conformations is enzymatically important, possibly describing a closing motion over bound substrates. A loss of β-sheet structure was noted during the simulations, and significant flexibility was reported in the flap tips. Results suggest that the semiopen state exists at a higher potential energy but is stabilized by entropy.

Hornak et al.,[17] Tozzini et al.,[18] Tóth and Borics,[19] and Hamelberg and McCammon[20] all observed flap opening from semiopen/closed starting structures in their simulations. In the Hornak publication,[17] unrestrained molecular dynamics performed on the apo protein using a continuum Generalized-Born solvation model showed multiple conversions between the closed and semiopen forms, along with several reversible large-scale flap opening events. It was noticeable that the most significant opening events were not associated with flap tip curling, as has been postulated elsewhere.[9] In the publication of Hamelberg and McCammon,[20] the accelerated molecular dynamics approach[21] was used to overcome the time scale limitations, and a closed−semiopen−open conformational transition was observed with the use of explicit solvent. In addition, frequent cis−trans isomerization events of the Gly-Gly ω dihedral angles in the flap tips were observed, suggesting that the flexibility if the Gly-rich flap tips contributes to the opening motion. This conclusion is supported by experimental mutagenesis data, which shows that the Gly residues are nearly intolerant of any substitution. A study by Tóth and Borics[19] used Langevin dynamics, a continuum solvent model, and only considered torsion angles, keeping bond angles and distances rigid. Starting from the semiopen conformation, they sampled the open conformation of the flaps in addition to a range of conformations including curling of the flap tips. In addition, Tozzini et al.[18] also observed flap opening among other conformations starting from a coarse-grained model of the semiopen structure. The simulations revealed four types of flap conformation, closed, open, wide-open, and semiopen. They also show that in the open form, the flap-tips have van der Waals contact with the laterally located residue 80 (residue of P1 loop) of the opposite monomer. This observation agrees with the same observation seen in a crystallographic study of the more open HIV-1 PR structure (pdb code 1TW7).[7]

1.3. Protonation State of the Aspartic Acid Dyad. A range of theoretical and experimental studies have been performed to determine the protonation state of the active site of HIV-1 protease, in which the aspartic acids 25 and 125 are in close proximity. At pH 7, both acids would normally be modeled in the deprotonated state,[22] although their close proximity complicates this issue.

A number of experimental studies show that the protonation state of the catalytic dyad is monoprotonated at the active pH range of 5−6, where the stability and activity of this protease is at a maximum.[23,24] At neutral pH, the dyad would lose its proton and become more unstable.[25] In contrast, there is an NMR study by Smith et al. which concludes that the dianionic form exists in solution.[26]

Theoretical studies also report contrasting protonation states for the aspartic acids. Wang et al.[22] suggest a dianionic state as this form gives inter-residue aspartate distances to be closest to the crystal structure. However, the MD simulations in implicit solvent are only 120 ps and showed little deviation from starting structures for any protonation state. The simulations of Scott et al. have been confirmed as using a dianionic dyad in the active site (personal communications), and the reported six counterions used by Perryman et al. also suggests an unprotonated active site.

In support of the monoprotonated form of the active site aspartates are the MD and ab initio studies of Piana et al.[27,28] In the free enzyme, the dyad was considered to be almost ionic, with the aspartic acid residues sharing one proton. It is suggested that the ionic state of the proton may account for the interpretation of the NMR data by Smith et al.[26] Also, at each step of the proposed catalytic cycle, the aspartic acid dyad exists either in the monoprotonated form, or in a transition state, sharing protons with the substrate or with water molecules.

An MD study of the stability of the monoprotonated and deprotonated forms of the aspartic dyad under physiological conditions[29] concluded similar dynamics for both systems, independent of the protonation state of the dyad. A sodium ion was shown to closely bind to the catalytic dyad in the deprotonated state, which blocks the effects of the repulsion. In the absence of this ion, the dimer became unstable, leading to a fully open structure and perhaps explaining the result of Scott et al.

There is therefore no clear choice of protonation state of the aspartic acid in the apoenzyme, and so, in this study, both the monoprotonated and deprotonated states have been investigated.

1.4. Summary. The HIV-1 PR apoenzyme is believed to exist in a semiopen structure in solution and is known to undergo two conformational motions: a limited motion on a time scale well within 10 ns that involves a few residues on the flap tips, and an opening event on a 100 μs time scale (Ishima et al.[11]). Scott et al.[9] report an opening event during 10 ns of simulation which increases the accessibility of the active site although the motion is not reversibly sampled. Perryman et al.[13] report 22 ns of simulation which did not sample the large opening event seen by Scott et al. but

presented results in agreement with the rapid flap curling motion reported by Ishima et al. Hornak et al.[17] and Hamelberg and McCammon[20] have both observed flap opening, the former using unrestrained molecular dynamics in a continuum solvent, and the latter using an enhanced sampling method and accelerated molecular dynamics.

In light of these results, we have applied our own enhanced sampling approach, Reversible Digitally Filtered Molecular Dynamics (RDFMD), to the simulation of flap motion in HIV-1 PR. This approach is computationally efficient and offers the advantage of being useable with explicit solvent models. In addition, these RDFMD simulations are supplemented by a range of normal and high temperature MD simulations, for the monoprotonated and dianionic aspartic acid dyad.

2. Computational Methodology

2.1. Reversible Digitally Filtered Molecular Dynamics. RDFMD is a method of amplifying motions of a specific frequency during a simulation. It has been used to promote conformational motions in the pentapeptide YPGDV,[30] the Syrian hamster prion protein[31] and *E. coli* dihydrofolate reductase.[32] There are a number of interdependent parameters that are associated with the method and a protocol optimized to maximize induced motion in dihedral angles will be applied here.[32,33] This parameter set was previously shown to be suitable for application to flexible regions of protein systems.[32]

RDFMD is performed by the combination of filter sequences separated by periods of traditional molecular dynamics. During the filter sequences, a short simulation is performed in the NVE ensemble with the internal velocities, $v_{\mathrm{int},i}$, saved at each step, i, to create a velocity buffer. A digital filter is a set of coefficients designed to create a certain frequency response. A suitable digital filter is applied to the velocity buffer by multiplying the coefficients of the filter, c_i, by the Cartesian components of the internal velocities, $v_{\mathrm{int},i}$, and summing across the velocity buffer (of length $2m + 1$ steps), as shown in eq 1. Depending on the digital filter applied, the set of velocities produced, v_{filt}, will have certain frequencies of motion amplified or suppressed. The filtered velocities are then multiplied by an amplification factor, A, which determines the level of amplification performed by each filter application and are summed with the original velocities from the center of the velocity buffer, v_0. This creates a new set of velocities, v', as shown in eq 2.

$$v_{\mathrm{filt}} = \sum_{i=-m}^{m} c_i v_{\mathrm{int},i} \tag{1}$$

$$v' = v_0 + A v_{\mathrm{filt}} \tag{2}$$

Simulation can be continued using the coordinate set corresponding to the center of the velocity buffer, with the new velocity set, v'. Another velocity buffer can be filled, to which the digital filter can again be applied, thus progressively amplifying motions of a specific frequency. It is desirable to control the number of steps between filters, d, and if this is to be less than half the buffer length (i.e., if $d < m$), the simulation after the first filter must be run both forward and backward in time, as illustrated in Figure 2. If the time between filter applications is too long, the effects of each filter dissipate before the next is applied, but if too short, the filter sequences are computationally expensive and the system has no time to respond to the amplification of velocities.[32]

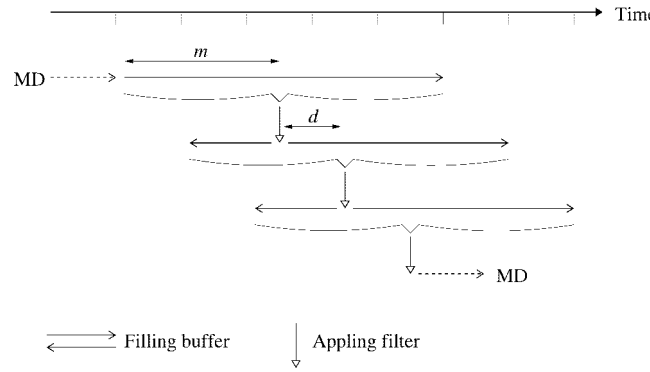

Figure 2. RDFMD sequence showing the delay parameter, d, and the filter length, $2m + 1$.

The repeated application of filters is performed until the internal energies of the targeted residues breach a specified internal temperature cap. This parameter will be discussed in a later section. Once this cap is reached, the filter sequence is ended, and molecular dynamics is performed for a period of time before another set of filters are applied. During this period the effects of the filters are allowed to dissipate, and the simulation temperature quickly returns to that set by a thermostat.

A suitable set of parameters for use with flexible regions of a protein has been optimized on a pentapeptide system[32] and shall be applied here. This includes the use of a digital filter designed to amplify frequencies between $0-100\ \mathrm{cm}^{-1}$ (for which 201 coefficients has been found to be sufficient), an amplification factor of 2, and a delay between filters of either 50 or 100 steps (any value in this region is suggested to be suitable). Filter sequences are separated by 4 ps of molecular dynamics simulation in either the NVT or NPT ensembles. This is sufficient time for the system temperature to return to the desired 300 K, and during this time conformers are sampled for analysis.

Full details of RDFMD, digital filter design, and the methods of optimizing RDFMD parameters can be obtained from previous publications.[30−32]

2.2. Computational Details. The 1HHP pdb structure[6] of the apo enzyme in a semiopen conformation was used as a starting point for both the monoprotonated and unprotonated active site systems. This structure consists of a single monomer, and a transformation of (x, y, z) to $(y, x, -z)$ is used to generate the symmetrical dimer. The crystal structure was checked, and polar hydrogen atoms were placed, using WHAT IF.[34] All other hydrogen atoms were placed and the structures solvated within the XLEAP utility of the AMBER package.[35] In the monoprotonated form of the enzyme, the proton was placed on the OD2 oxygen of aspartate 125, which lies close to the catalytic aspartate of the second monomer. Solvent was added to give a minimum distance of 12 Å from the protein surface to the cell boundary. The

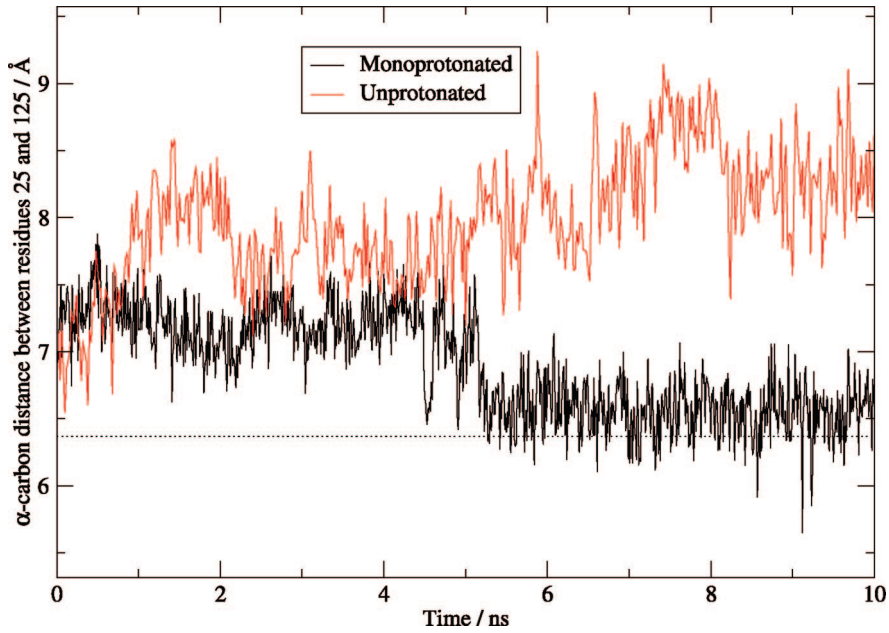

Figure 3. Inter-residue distance between catalytic aspartic acids 25 and 125. Results from 10 ns of MD simulation using the unprotonated and monoprotonated conformations are shown. The dotted line indicates the inter-residue distance in the 1HHP crystal structure.

overall charge was neutralized by the addition of chloride ions within XLEAP. The CHARMM27 force field[36] has been used with TIPS3P[37] explicit solvent.

All simulations presented here have been performed using the NAMD package[38] (NAMD was developed by the Theoretical Biophysics Group in the Beckman Institute at Urbana−Champaign) with cuboid periodic boundary conditions, a particle mesh Ewald treatment of electrostatics (with an interpolation order of 6), a 2 fs time step, and a switching function applied to Lennard−Jones interactions between 9 Å, and the 10.5 Å cutoff. All bonds containing hydrogen atoms have been constrained to equilibrium lengths using SHAKE with a tolerance of 10^{-8} Å. Simulations performed in the NVT ensemble use a Langevin thermostat, with an associated damping parameter. Simulations in the NPT ensemble also use a Nosé−Hoover Langevin barostat (with piston period and decay parameters).

During system setup a total of 96 000 steps of minimization were performed using the conjugate gradient line-search algorithm.[38] The system was then heated in the NVT ensemble for 40 ps at temperatures from 50 to 300 K, at 50 K intervals. A 10 ps^{-1} thermostat damping parameter was used to control the system temperature. An NPT simulation was then used to equilibrate the system pressure to a target of 1 atm; 100 ps of molecular dynamics was performed using a decay parameter of 100 fs and a piston period of 200 fs. A further 200 ps was then performed using a decay parameter of 300 fs and a piston period of 500 fs. The equilibrated system contained 9370 water molecules and 7 chloride ions and had cell dimensions of 61.08, 60.98, and 82.37 Å.

Molecular dynamics simulations are performed in the NVT ensemble using a 5 ps^{-1} thermostat damping parameter. The RDFMD simulations use either the NVT or NPT ensembles between filter sequences (which are performed under NVE conditions), using a 5 ps^{-1} thermostat damping parameter, a barostat decay parameter of 300 fs and a piston period of 500 fs as applicable.

3. Results

It is likely that the protonation state of the catalytic dyad is important for determining the forces in the cavity of the protein, and these may influence the conformational motions of the protein flaps. Initial simulations have therefore been performed with the monoprotonated and unprotonated dyad states and the distance between the α-carbons of the aspartic acid residues are shown in Figure 3. The distance observed in the crystal structure of 1HHP is shown with a dotted line.

The monoprotonated dyad sees two distinct states. The first has an α-carbon separation between residues 25 and 125 of over 7 Å, and the second has a distance less than 7 Å. Visualization of the simulation trajectories (not presented here) indicates that this change, which occurs rapidly at 5 ns, involves the removal of a water molecule from between the aspartic acid sidechains. This water molecule entered the region during the equilibration stage. This is not unexpected, as a water molecule is involved in the reaction catalyzed by HIV-1 PR. The second state has a similar inter-residue distance to that seen in the crystal structure.

The unprotonated dyad begins with an aspartic acid separation similar to that seen in the crystal structure, but this distance quickly increases due to the insertion of several water molecules between the aspartic acid side chains. Similar insertions of water molecules have been previously observed during MD simulations.[29] The resulting structure indicates a distance approximately 2 Å larger than that seen in the crystal structure.

It would therefore appear that the dimer interface is more stable when using a monoprotonated dyad, and subsequent simulation results presented here are all performed on the monoprotonated dyad system.

3.1. Molecular Dynamics. It is difficult to describe the flap motions of HIV-1 PR in graphical form, and analysis is shown by characteristic interaction distances and by snapshots of the protein displayed in a manner consistent to Figure 1. Trajectories of simulations presented in this study may be available by request from the corresponding author. Three conformations are identified during MD simulations, which can be defined by the location of the flap tip isoleucine residues 50 and 150, with respect to important hydrophobic side chains. How open the structure is can be described using the distance between the two flap tips, and the distances between the flap tips and the catalytic residues.

The starting semiopen structure, 1HHP, requires the side chain of the isoleucine flap tips to be closely associated with the side chain of the phenylalanine residue on the opposite flap (for example, an interaction between residues 50 and 153). The closed conformation, seen in crystal structures with inhibitors or substrate analogues, is characterized by the two flaps reaching 'past' each other, interacting with the P1 loop (residues 79 to 81) of the opposite monomer. This can be most clearly seen when looking at the distance between the side chains of the isoleucine flap tips and a proline residue in the P1 loop (residue 81 or 181). The semiopen and closed conformers can be identified using rmsd against the 1HHP and 1G6L pdb structures presented in Figure 1. Superposition is performed over residues 1 to 45, 55 to 99, 101 to 145, and 155 to 199 (those residues deemed not to be part of protein flaps) and the rmsd of the remaining flap residues (46 to 54 and 146 to 154) measured. A third conformation, with the flap tips curled back toward the P1 loop of the same monomer is also frequently observed, and can be defined by the close interaction between the isoleucine 50 to proline 81 (or residue 150 to 181) side-chain distance.

No significant opening events are sampled by the flaps during the 10 ns MD simulation. The first flap (assigned the lower residue numbers) fluctuates between the semiopen and curled conformations (demonstrated by the close distances between residues Ile50 and Phe153 (semiopen) and Ile50 and Pro81 (curled); see Supporting Information) and the distance between the flap tips and the catalytic residues is maintained at approximately 15 Å, indicating the stability of the cavity structure.

The second flap also begins in a state similar to the semiopen form; however, after 0.7 ns, a clearly defined closed conformation is obtained. The rmsd against the semiopen structure increases, and the rmsd against the known closed structure falls from 6 to 3 Å. This change is accompanied the movement of flap tip residue 150 into close proximity with proline 81. After 1.7 ns, the second flap returns to a semiopen conformation. After this point, the flap interchanges between the semiopen and curled forms, in a similar manner to that seen in the first flap.

Conformations adopted by the HIV-1 PR flaps during this simulation are representative of those referred to as semiopen, curled, and closed. Figure 4 shows examples of each of these states. For clarity, only the flap and P1 loop regions of the protein as shown in the top views in Figure 1 are displayed. The nature and timescales of the motions observed suggest that it is the interconversion between these conformers that account for the fast motions (<10 ns) observed by Ishima et al. There is no evidence of flap-opening or of increased accessibility to the active site. It is also worth noting that although the closed conformation is sampled, this only occurs in one flap rather than in both as seen in X-ray structures with bound inhibitors.

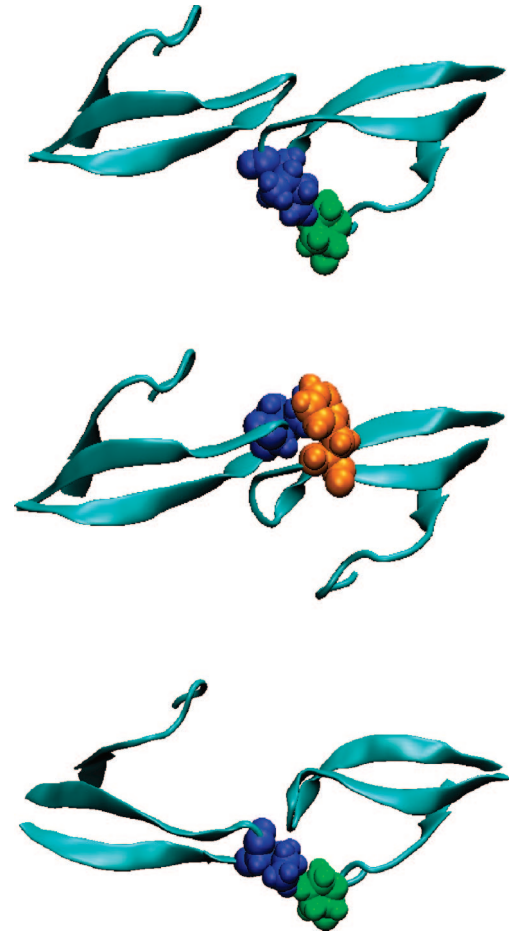

Figure 4. Top: Right-hand flap exhibits curled conformation with close contact between residues 150 (blue) and 181 (green). Middle: Left-hand flap in the semiopen conformation as seen by the close contact between residues 50 (blue) and 153 (orange). Bottom: Left-hand flap in closed conformation, with close contacts between residues 50 (blue) and 181 (green). Conformers obtained after 10 ns (top and middle) or 1.3 ns (bottom) of monoprotonated dyad MD simulation at 300 K. Residues 44 to 59 and 76 to 83 of each flap are shown, as presented in Figure 1.

3.2. Raised Temperature Molecular Dynamics. To assess the flexibility of protein residues, and to determine the stability of the HIV-1 protease system, a range of 10 ns simulations were performed at 50 K intervals from 350 to 500 K. Simulations began from the same starting point as the 300 K simulation, and the temperature was adjusted using the Langevin thermostat. The results are briefly discussed here using the locations of the flap tip residues isoleucine 50 and 150 to describe the conformers observed.

The flap conformations sampled in the 350 K MD simulations are similar to those of the 300 K MD simulation. As in the 300 K simulation, the closed conformation is only attained in one flap, indicating that although mobile between the semiopen, closed and curled conformers, the change to

a stable state with both closed flaps may be outside the 10 ns time scale. It is worth noting the consistency with the results of Perryman et al., performed on a mutant similar to the wild-type protein used here, which sampled transitions from the closed to semiopen form in only one flap within a 22 ns simulation.

The 400 K simulation samples conformations to those at 300 and 350 K. However, the simulation at 450 K observes a significant flap opening event where the second flap moves away from the active site, and another, smaller flap opening event through the curling back of the flap tips (see Supporting Information). This opening could be similar to that reported by Scott et al., although Scott et al. sampled an immediate opening.

At 500 K, the two flaps both drop toward the active site near the start of the simulation, completely excluding water from the protein cavity. There is no experimental data to suggest the existence of this conformation, although a similar event was sampled during a simulation performed by David et al.[39] with an implicit solvent distance-dependent dielectric model. It is possible that the accessibility of this conformation may be associated with a change in the solvation properties of the water model with temperature, but it is more likely to indicate a deficit of the simulation procedures and model for this system using a rapid jump to such a high temperature.

3.3. RDFMD. RDFMD has the ability to selectively enhance motions in certain regions of the protein, without increasing the energy in all degrees of freedom. This is ideal for the HIV-1 PR system, which has been shown to be highly flexible and unstable at elevated temperatures. The flap tip residues are known to be mobile and have been shown to adopt distinct conformers. Indeed, in the recent work of Hamelberg and McCammon,[20] the flexibility of the Gly-rich flap tips has been shown explicitly to contribute to the opening and motions of the flaps. RDFMD is therefore applied to the flap tip residues 49 to 51 and 149 to 151 using the previously described protocol. One hundred filter sequences are performed in an RDFMD simulation, with each sequence being separated by 4 ps of NPT or NVT MD simulation which is pieced into a trajectory for analysis. The protocol used is designed to increase dihedral angle motion in targeted regions as previously described. An important limitation of the RDFMD method when applied to increase dihedral motions is that the frequencies of these motions are similar to those in the ω peptide dihedral. However, by limiting the energy put into the system, ω angle isomerization can be excluded because of the higher energy barriers of isomerization of this dihedral, compared with those of ψ and ϕ. Within each filter sequence, amplification of low frequency motions is therefore repeated only until the targeted residues reach a desired temperature cap. No simulations presented here include isomerization of peptide bonds.

Twelve RDFMD simulations have been performed, visiting each permutation of protocol found by varying either NVT or NPT MD between filter sequences, a temperature cap of 900, 1100, or 1300 K, and 50 or 100 steps between filter applications. A range of protocols are used so to investigate the response of the system to different parameter choices, and to test the reproducibility of results when using this nonequilibrium method. Two of the simulations with a 1300 K temperature cap saw isomerization of the peptide bonds and have therefore been discarded. A total of ten RDFMD trajectories are therefore considered here.

Similar motions to those observed during molecular dynamics simulations are seen using RDFMD, with transitions occurring at an increased rate between closed, curled and semiopen states. However, the RDFMD results discussed here can be grouped into three categories, each sampling a significant motion to a greater extent than seen (if seen at all) during the MD simulations: reversible flap opening, transitions between the semiopen and closed conformations, and flap separation via curled conformations. For each category, the clearest simulation results obtained are presented, along with an indication of how often this type of motion is observed. Simulations that are not presented contain lesser movements or do not sample transitions as completely. Where flap opening motions are noted, the shortest distance between all atoms of residues 45 to 55 and those of 145 to 155 is reported (giving the separation of the flaps looking at the flap tips and four residues either side). This is intended to indicate the extent of accessibility to the active site. Simulation snapshots are considered a clear way of describing conformers sampled and are included as a replacement to full graphical analysis, which is included only where of particular interest.

When considering the motions induced by RDFMD it is important to note that only three residues at the tip of each flap are targeted by the applied protocol.

3.3.1. Reversible Flap Opening. Reversible flap opening is observed in two of the ten 400 ps RDFMD simulations. The simulation presenting the largest opening event used a filter delay of 100 steps, a temperature cap of 1100 K, and NPT MD simulation. Prior to the opening event, both flaps undergo dynamics similar to those reported for MD, with the first flap remaining in the semiopen conformation and the second flap converting between curled and semiopen conformations (indicated by proximity of the isoleucine 150 side chain with the side chains of phenylalanine 53 and proline 181, see Supporting Information).

At 280 and 305 ps, significantly open conformations which are believed to be similar to those reported by Scott et al.[9] are sampled. The flap separation reaches a maximum of 10.1 Å at 305 ps, indicating an opening motion of more than 6.5 Å (see Supporting Information). Snapshots illustrating the significant simulation conformations are shown in Figure 5. As Figure 5a and 5b show, in the most open conformation, flap 2 adopts a curled conformation and the first flap slowly increases its distance from the active site with no interaction with the P1 loop, previously seen to stabilize the curled state. After 320 ps, the opening is reversed, indicated by the decrease in the distance between the nearest flap residues and the flaps return to a conformation where the active site is inaccessible with the flaps approaching a closed conformation at the end of the simulation (Figure 5c and 5d).

This simulation appears to sample more completely the motion reported by Scott et al., with both an opening motion away from the active site, and a closing that Scott et al. did not observe. The secondary structure observed during this

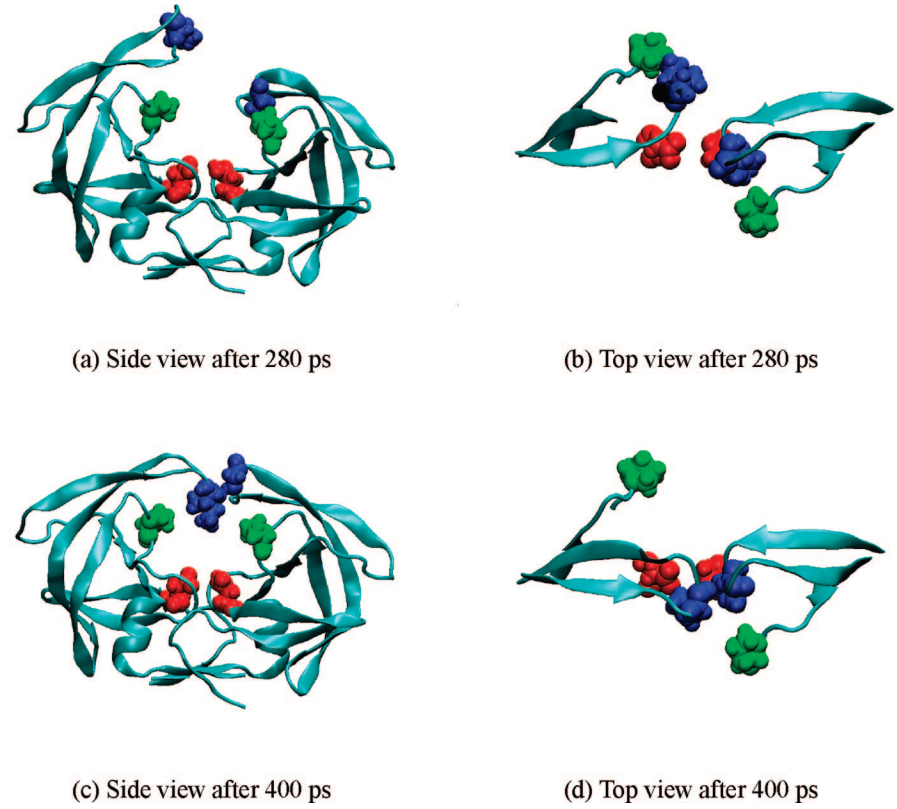

Figure 5. Snapshots of RDFMD simulation using a filter delay of 100 steps, an internal temperature cap of 1100 K and NPT simulation between filter sequences. Presentation is as described in Figure 1.

RDFMD simulation shows no significant differences to that obtained from the 10 ns MD trajectory (analysis not shown), with limited variation in the β-sheet content of the flaps. This is the case for all RDFMD simulations performed in this study.

A complete opening and corresponding closure was also obtained by the RDFMD simulation with a filter delay of 100 steps, a temperature cap of 900 K, and NVT MD simulation between filter sequences. An opening of 8.3 Å is observed in a similar fashion with one flap lifting from the protein cavity, and the other flap adopting a curled conformation. Analysis shows similar results for this lesser opening event (not presented). The opening motion of one flap has therefore been observed in two out of the ten RDFMD simulations performed.

3.3.2. Transitions between the Semiopen and Closed Conformations. One of the ten RDFMD simulations sampled a closed conformation from the initial semiopen starting structure, a conformational change not sampled during the MD simulations where, similar to several of the RDFMD simulations, only one flap moves toward a closed state. In Figure 6, six snapshots from this trajectory are presented. The conformation sampled by the RDFMD simulation where both flaps are in the closed state is similar to the closed 1G6L X-ray structure obtained from a tethered apo dimer and is commonly seen in the presence of an inhibitor. In this conformation, the flap tips swap sides (as shown in Figure 1d and 1e). This conformational change starts after 50 ps during the RDFMD simulation and does not require any significant separation in the flaps (see Supporting Information).

We consider the pathway between the semiopen and closed conformers to be similar to that seen in a reaction path study, published in 1998, by Rick et al.[40] There is little movement in the flap residues that are involved in the β-sheet structure of the protein flaps, and the flap tips (residues 48 to 52 and 148 to 152) do not cross above or below each other but pass directly in front. This is facilitated in the work of Rick et al. by the tips bending toward the direction they are to move in. However, in this simulation, the bend is toward the P1 loop in the opposite direction. The movement therefore is initially similar to the flap curling motion. Once the flaps have crossed, the flap tips straighten to give the closed conformation as seen in Figure 6 at 82 ps. Interestingly, as noted by Rick et al., it is the flap tip residues that are primarily involved in this motion, and these are assigned by Ishima et al. to the faster motion observed in a time scale of less than 10 ns by NMR. However, the single occurrence of this event in this study suggests that the transition of both flaps between semiopen and closed conformations does not occur on this time scale, and it is the semiopen to curled motion that is observed by Ishima et al.

Also noted in this simulation is the occurrence of a brief opening event just before 100 ps, during which a maximum flap separation of 6.5 Å is observed at 127 ps. Figure 6 shows snapshots of the simulation trajectory, in which, at 127 ps, a conformation is seen with strong interactions between the flap tips and the P1 loop of the opposite protein monomers but with no contact between the two flap tips, as in pdb structure 1TW7. From this state the second flap undergoes an opening motion, reaching a maximum flap separation of 9.8 Å and making no contact with the opposite P1 loop. This

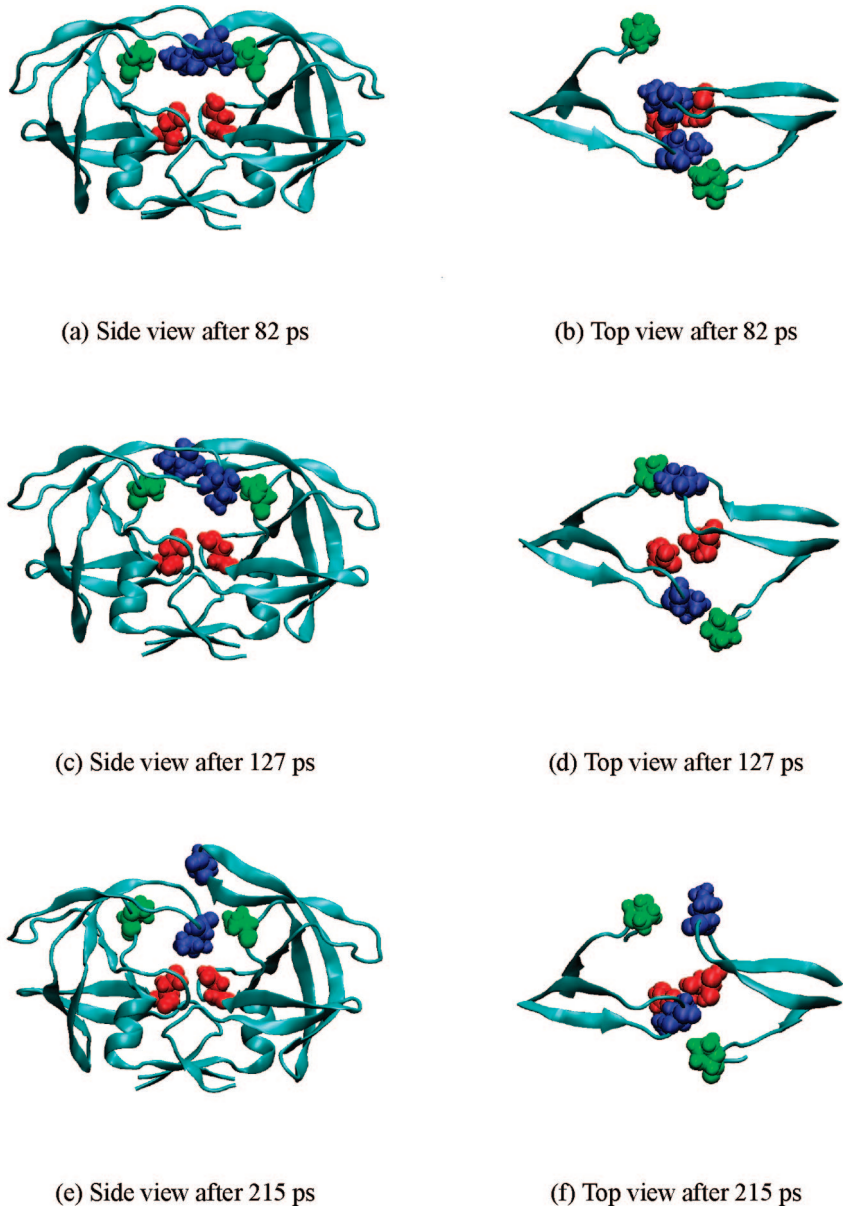

Figure 6. Conformers observed during the 400 ps RDFMD simulation of HIV-1 PR using a filter delay of 100 steps, an internal temperature cap of 1300 K, and NPT simulation between filter sequences. Presentation is as described in Figure 1.

is most clearly seen in Figure 6 at 215 ps. Following this opening motion, the flaps temporarily close back to the closed conformation, with both flaps in contact. The flaps do separate again to a lesser extent, with contact between the flap tips and the P1 loop of the opposite monomers maintained.

This RDFMD simulation has therefore shown a transition between the semiopen and closed conformations in both flaps that was not sampled in the 300 K or raised temperature MD simulations. This motion was observed in only one of the ten RDFMD simulations and could be an important step of the protein's catalytic cycle. It was noted that no flap separation was required to move between states. Also suggested is that flap opening from the closed conformation is possible, perhaps making use of an intermediate state with no contact between the flap tips. Such a state is observed in the open conformation pdb structure 1TW7.

3.3.3. Flap Separation via Curled Conformations. A separation of flaps achieved by both flaps adopting tightly curled conformations was seen in several RDFMD simulations. The simulation with a filter delay of 50 steps, a temperature cap of 900 K, and NPT between simulations showed the largest separation which reached a maximum of 8.4 Å by 400 ps. Of greater interest, despite the smaller opening of 6.7 Å, is the RDFMD simulation with a filter delay of 100 steps, temperature cap of 1100 K, and NVT between simulations, which sampled the opening in a reversible manner (see Supporting Information). Three other RDFMD simulations also sample flap separation via both flaps adopting curled conformations, and with five out of the ten simulations observing this event, it is the most frequently seen of the large motions presented here using RDFMD. It was also noted to a lesser extent very briefly during the 400 K MD simulation, where flap contact was broken for approximately 15 ps.

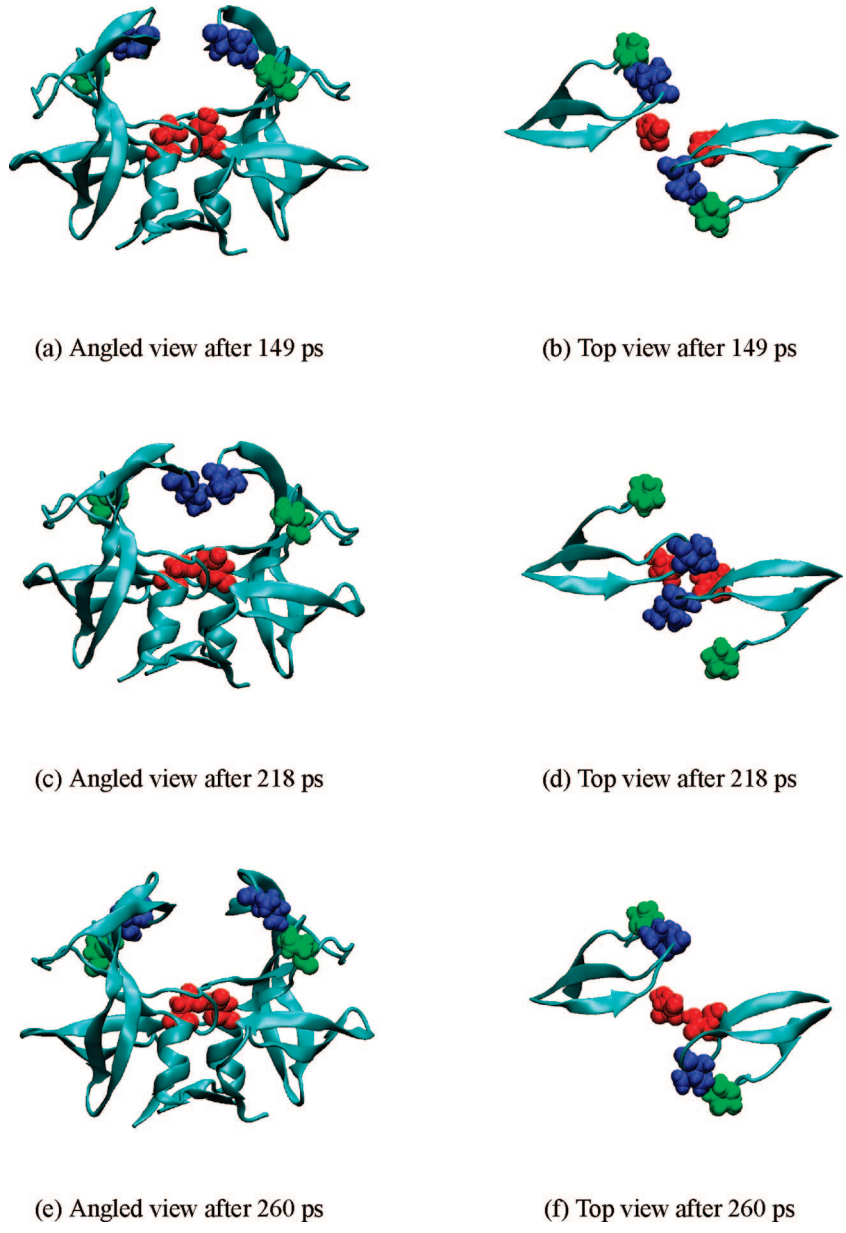

Figure 7. Conformers observed during the 400 ps RDFMD simulation of HIV-1 PR using a filter delay of 100 steps, an internal temperature cap of 1100 K and NVT simulation between filter sequences. Presentation is as described in Figure 1, with an angled view replacing the side view.

Initially, in the simulation with a filter delay of 100 steps, temperature cap of 1100 K, and NVT between simulations, the first flap is seen in a semiopen state and the second flap is curled. After 120 ps, the first flap swaps to the curled conformation. With both flaps curled, the flaps are separated (see Supporting Information). A channel is then revealed with a clear path to the active site for which no movement of the flaps away from the protein cavity and active site is required (as seen in RDFMD simulations already presented here). The flap curling is also an extension of the motion seen frequently within the 10 ns MD simulation and appears to be inherent to the HIV PR protein.

After 212 ps, the extent of the flap curling is decreased, and the contact between the flaps is regained. However, shortly afterward at 250 ps, the second flap moves again into a tightly curled conformation, and another separation of the protein flaps is seen. At 268 ps the second flap returns to the semiopen state and the flaps again close, maintaining contact until the end of the simulation. Snapshots from the simulation showing the reversible nature of the curled opening motion are shown in Figure 7. The side view shown for other simulations is replaced by a view taken from an angle to the protein, revealing the channel that opens above the active site.

4. Discussion

Analysis of the molecular dynamics simulations indicates a stable dimer interface and cavity region using the monoprotonated dyad. A conformational change to and from the closed state in one flap occurs, and the semiopen and curled conformers interconvert at a time scale well within that of the simulation. These two motions are likely to correspond to the fast event described in the NMR study by Ishima et

al. The flaps do not separate in either simulation as seen by Scott et al. although we believe our results to be consistent with those of Perryman et al., who also sampled a closed protein cavity and curling of the flap tips at room temperature.

The opening event described by Ishima et al. as occurring on a 100 μs time scale has not been sampled in the molecular dynamics simulations, and investigation was therefore continued at elevated temperatures. At 350 and 400 K, the protein flaps interconvert between clearly defined semiopen, closed, and curled conformations. Only a brief separation of flap tips is observed at 400 K, and the 350 and 400 K simulations otherwise yield results considered to be similar to those at 300 K. At 450 K, an opening event was sampled in which one flap moves away from the active site. The requirement of such high temperatures to break the otherwise stable contact of the flap tips adds strength to the conclusion that flap opening is outside the time scale of the faster motion observed by Ishima et al. At 500 K, both flaps moved into the protein cavity, adopting a collapsed conformation that indicates an instability of the protein or a breakdown of the simulation model at such a high temperature.

RDFMD simulations show three types of large scale conformational change: the transition of both flaps between the known semiopen and closed conformations, and the increased accessibility of the active site via both reversible flap opening and by flap curling. The flap opening is similar to that reported by Scott et al. although the corresponding closure is also sampled here in two of the ten RDFMD simulations.

The opening via both flaps adopting curled conformations is seen in five of the ten RDFMD simulations presented in this study, of which four sampled the corresponding closure. It is possible that this motion involves overcoming a comparatively smaller energy barrier than an opening by a flap moving away from the cavity. However, since the opening is an extension of a curling motion seen frequently in the MD simulations, it is also possible that the RDFMD method, which amplifies only motions that exist in a simulation buffer, is more likely to sample the double curling event.

It is important to note that RDFMD is seen in this study to induce large motions in the protein by targeting very few residues. This suggests that RDFMD is not necessarily driving the motion, but is assisting the protein in overcoming the initial barriers to flap tip separation. This is consistent with the observation of Hamelberg and McCammon[20] that flexibility of the flap tips contributes to their opening, although it should be noted that the simulations reported here explicitly exclude cis−trans isomerization of ω dihedral angles. The corresponding RDFMD-induced openings and conformational changes involve more residues than those targeted. It is also an important result that the reverse of opening motions are sampled within periods in which filter amplifications are applied. This confirms that RDFMD is not simply disrupting the structure as frequency independent heating may be expected to do, but is successfully assisting in conformational motions inherent to the system. The repeated sampling of the opening events induced by RDFMD, and the similarity of the closed conformer obtained to a known X-ray structure, also adds confidence in RDFMD as a conformational sampling tool.

5. Conclusions

In this manuscript, the application of an enhanced sampling technique, Reversible Digitally Filtered Molecular Dynamics, to the motion of the flaps in HIV-1 PR has been described. A range of motions are sampled in the simulations including the reversible opening of one of the flaps in a fashion similar to that reported by others, transitions between the known semiopen and closed conformations, and finally the reversible opening of the flaps by a mechanism in which both flaps curl back toward the P1 loop of their own monomer. It is important to note that all these motions are seen to be reversible, lending confidence to the results. Furthermore, all these transitions are brought about by amplifying the low frequency vibrational motions in the flap tips.

It must of course be noted that although RDFMD is efficient and capable of being used in explicit solvent, it is a nonequilibrium simulation method. For this reason, a number of RDFMD trajectories have been produced for analysis, and consistency in the conformational transitions observed has been sought. That reversible transitions from the initial semiopen to the known closed structure are observed supports the assertion that RDMFD is successfully assisting in sampling conformational motions inherent to the system. The preponderance of RDFMD trajectories following the double-curling opening mechanism may therefore reflect the presence of this type of motion in normal molecular dynamics.

Acknowledgment. This work has been funded by EPSRC. We acknowledge the contributions of S. Phillips, M. Swain, C. Edge, and R. Gledhill in the development of the RDFMD method. We thank the authors of the NAMD package for the provision of the program and for its continuous improvement. We also thank the EPSRC for the provision of computer resources (grant no. GR/R06137/01). Software to perform RDFMD is available from the authors on request.

Supporting Information Available: Analysis of MD and RDFMD simulations using characteristic interaction distances. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Oroszlan, S.; Luftig, R. B. *Curr. Top. Microbiol. Immunol.* **1990**, *157*, 153–185.

(2) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(3) Prasanna, M. D.; Vondrasek, J.; Wlodawer, A.; Bhat, T. N. *Proteins* **2005**, *60*, 1–4.

(4) Prasanna, M. D.; Vondrasek, J.; Wlodawer, A.; Rodriguez, H.; Bhat, T. N. *Proteins* **2006**, *63*, 907–917.

(5) Zoete, V.; Michielin, O.; Karplus, M. *J. Mol. Biol.* **2002**, *315*, 21–52.

(6) Spinelli, S.; Liu, Q. Z.; Alzari, P. M.; Hirel, P. H.; Poljak, R. J. *Biochimie* **1991**, *73*, 1391–1396.

(7) Martin, P.; Vickrey, J. F.; Proteasa, G.; Jimenez, Y. L.; Wawrzak, Z.; Winters, M. A.; Merigan, T. C.; Kovari, L. C. *Structure* **2005**, *13*, 1887–1895.

(8) Pillai, B.; Kannan, K. K.; Hosur, M. V. *Proteins: Struct. Funct. Genet.* **2001**, *43*, 57–64.

(9) Scott, W. R. P.; Schiffer, C. A. *Structure* **2000**, *8*, 1259–1265.

(10) Cigler, P.; Koziesk, M.; Rezacova, P.; Brynda, J.; Otwinowski, Z.; Pokorna, J.; Plesek, J.; Gruner, B.; Doleckova-Maresova, L.; Masa, M.; Sedlacek, J.; Bodem, J.; Kraeusslich, H. G.; Kral, V.; Konvalinka, J. *Proc. Natl. Acad. Sci. U.S.A* **2005**, *102*, 15394–15399.

(11) Ishima, R.; Freedberg, D. I.; Wang, Y. X.; Louis, J. M.; Torchia, D. A. *Structure* **1999**, *7*, 1047–1055.

(12) Katoh, E.; Louis, J. M.; Yamazaki, T.; Gronenborn, A. M.; Torchia, D. A.; Ishima, R. *Protein Sci.* **2003**, *12*, 1376–1385.

(13) Perryman, A. L.; Lin, J. H.; McCammon, J. A. *Protein Sci.* **2004**, *13*, 1108–1123.

(14) Meagher, K. L.; Carlson, H. A. *Proteins* **2005**, *58*, 119–125.

(15) Zhu, Z. W.; Schuster, D. I.; Tuckerman, M. E. *Biochemistry* **2003**, *42*, 1326–1333.

(16) Rick, S. W.; Erickson, J. W.; Burt, S. K. *Proteins* **1998**, *32*, 7–16.

(17) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 915–920.

(18) Tozzini, V.; Trylska, J.; Chang, C. E.; McCammon, J. A. *J. Struct. Biol.* **2007**, *157*, 606–615.

(19) Toth, G.; Borics, A. *J. Mol. Graph. Model.* **2006**, *24*, 465–474.

(20) Hamelberg, D.; McCammon, J. A. *J. Am. Chem. Soc.* **2005**, *127*, 13778–13779.

(21) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919–11929.

(22) Wang, W.; Kollman, P. A. *J. Mol. Biol.* **2000**, *303*, 567–582.

(23) Hyland, L. J.; Tomaszek, T. A. J.; Meek, T. D. *Biochemistry* **1991**, *30*, 8454–8463.

(24) Ido, E.; Han, H.; Kezdy, F. J.; Tang, J. *J. Biol. Chem.* **1991**, *266*, 24359−24366.

(25) Cheng, Y. E.; Yin, F. H.; Foundling, S.; Blomstrom, D.; Kettner, C. A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9660–9664.

(26) Smith, R.; Brereton, I. M.; Chai, R. Y.; Kent, S. B. H. *Nat. Struct. Biol.* **1996**, *3*, 946–950.

(27) Piana, S.; Carloni, P. *Proteins* **2000**, *39*, 26–36.

(28) Piana, S.; Carloni, P.; Parrinello, M. *J. Mol. Biol.* **2002**, *319*, 567–583.

(29) Kovalskyy, D.; Dubyna, V.; Mark, A. E.; Kornelyuk, A. *Proteins* **2005**, *58*, 450–458.

(30) Phillips, S. C.; Swain, M. T.; Wiley, A. P.; Essex, J. W.; Edge, C. M. *J. Phys. Chem. B* **2003**, *107*, 2098–2110.

(31) Phillips, S. C.; Essex, J. W.; Edge, C. M. *J. Chem. Phys.* **2000**, *112*, 2586–2597.

(32) Wiley, A. P.; Swain, M. T.; Phillips, S. C.; Edge, C. M.; Essex, J. W. *J. Chem. Theory Comput.* **2005**, *1*, 24–35.

(33) Wiley, A. P.; Gledhill, R. J.; Phillips, S. C.; Swain, M. T.; Edge, C. M.; Essex, J. W. In *The Hilbert-Huang Transform in Engineering: The analysis of molecular dynamics simulations by the Hilbert-Huang transform*; Huang, N., Attoh-Okine, N. O., Eds.; CRC Press: Boca Raton, FL, 2005; pp 246−263.

(34) Vriend, G. *J. Mol. Graph.* **1990**, *8*, 52–56.

(35) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1–41.

(36) Mackerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-Mccarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(37) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(38) Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 283–312.

(39) David, L.; Luo, R.; Gilson, M. K. *J. Comput. Chem.* **2000**, *21*, 295–309.

(40) Rick, S. W.; Erickson, J. W.; Burt, S. K. *Proteins* **1998**, *32*, 7–16.

CT800152D