Explicit solvent dynamics and energetics of HIV-1 protease flap-opening and closing

S. Kashif Sadiq[1,*] and Gianni De Fabritiis[1,†]

[1] *Computational Biochemistry and Biophysics Laboratory (GRIB-IMIM), Universitat Pompeu Fabra, Barcelona Biomedical Research Park (PRBB), C/ Doctor Aiguader 88, 08003 Barcelona, Spain*

(Dated: June 8, 2010)

An accurate description of the conformational dynamics of the β-hairpin flaps of HIV-1 protease is of central importance in elucidating the functional recognition of the enzyme by ligands. Using all-atom molecular dynamics simulations in explicit solvent, with a total of 461 trajectories of approximately 50 ns each, we report the closed, semi-open, open and wide-open flap conformation of the free wildtype protease. The free energy of flap opening and closing from the semi-open state is 0.9 $\pm$ 0.2 kcal/mol and 2.4 $\pm$ 0.4 kcal/mol respectively. The mean relaxation time of opening is approximately 8 ns, in good agreement with NMR data. The explicit solvent simulations quantitatively confirm the hypothesis that the semi-open state is the dominant population in the free protease whilst fast flap-tip fluctuations lead frequently to an open state. More pronounced flap rearrangements lead to a rare wide-open state with the catalytic site completely exposed to the solvent. The structures of the different flap-conformations provided herein are of general interest for improved drug design of HIV-1 protease, in particular, the wide-open conformation could be favoured by the large Gag and GagPol polyprotein chains. Strategies that take into account multiple flap-gating mechanisms may lead to more effective inhibitors.

I. INTRODUCTION

The HIV-1 protease is a key target for antiretroviral inhibitors[1,2] due to its pivotal role in the maturation of infectious virions via cleavage of polyprotein precursors encoded by the Gag and GagPol genes[3]. The mechanistic basis of protease-ligand recognition and processing and its differences upon mutational variations is important for the design of new inhibitors that circumvent the growing problem of multi-drug resistance[4].

HIV-1 protease is a C_2-symmetric homodimer[5] composed of monomeric chains of 99 amino acids (see Figure 1(a)) that form a dimeric interface consisting of a tightly coupled four stranded interdigitated β-sheet, a loosely coupled pair of β-hairpins termed the 'flaps' and a central region consisting of a strong hydrogen bond network termed the 'fireman's grip' which includes a catalytic aspartic acid dyad that forms the base of the active site cavity recognised by ligands. The active site is laterally obscured by two loop regions (one from each monomer) termed the 'walls'. The result of active site formation through dimer interfacing and the lateral juxtapositioning of the wall loops forms a uniquely directional binding trench that extends diagonally across the protease and forms the binding surface (see Figure 1(b)) for cleavage-site peptidic substrates[6–9].

Understanding the mechanistic basis of flap dynamics is of central importance in elucidating the functional recognition of ligands by the protease. One of the principal insights from crystallographic studies is that the flaps can exist in several conformations: a) a semi-open conformation when unbound[5], defined by minimal inter-flap contact with a cis-handedness whilst maintaining closure of the active site cavity and b) a closed conformation when liganded[7], defined by substantial flap overlap with trans-handedness and resulting in complete cavity closure (see Figure 1(c)). However, to achieve ligand binding, conformational changes that result in the opening of the flaps are required to provide ligands access to the active site. Such conformational changes are difficult to probe experimentally at the atomistic level. Interestingly, a recent crystal structure of the protease in a proposed open flap conformation[10] was determined to be due to crystal packing effects[11]. NMR investigations have revealed the structural flexibility of the flaps as well as insight into different flap conformations. Two classes of conformational opening have been discerned corresponding to a flap-tip flexibility on a $\sim$10 ns timescale and a rare opening event on the 100 μs timescale, resulting in a more disordered flap conformation[12–14].

Computational investigations using molecular dynamics (MD) techniques allow atomistic level insight into such conformational changes. Such studies have supported the structural flexibility of the protease flaps[15–19], especially the curling of the flap-tips, in good qualitative agreement with NMR experiments. However, the flap opening observed in such early studies[18] occurs within the first few nanoseconds and has been suggested to be due to improper equilibration[20]. Previous studies, limited to just few nanosecond timescale trajectories, did not allow to visit more rare conformations of flap opening that might occur on longer timescales. To circumvent this problem, low viscosity implicit solvent has been used as a means of accelerating the dynamics[21] at the cost of a realistic kinetic description. Such studies[21,22] have shown that the flaps spontaneously open in the unbound state into conformations with no inter-flap contact and an accessible active site cavity and subsequently reclose into conformations resembling the native unbound crystal structure. Furthermore, placement of a ligand in an open-state protease results in spontaneous flap closure[23,24] and the reversal of flap-handedness. The

broad mechanistic picture of ligand recognition that thus emerges from both computational and experimental investigations is that an equilibrium of structures exist in the unbound protease, favouring a semi-open conformation; flap opening is a rare event and is subsequently followed by ligand binding after which the flaps close around the ligand reversing handedness.

Despite extensive studies of flap dynamics, there remain crucial ambiguities into what constitutes an open state as well as the exact conformational changes and energetics associated with flap opening. Scott and Schiffer[18] and Toth and Borics[22,24] described the curling-in of the flap tips to expose an active site cavity large enough for substrate binding. The open structure of Hornak *et al.*[21] described upward motion of the flaps away from the dyad and correlated downward motion of the flap elbows and fulcrum. In addition to this, coarse-grained simulations by Tozzini *et al.*[25] have shown opening (termed semi-open therein) in accordance with the open crystal structure mentioned above[10] where flap motion occurs laterally outwards in an opposite direction to that observed by Scott and Schiffer[18]. Furthermore, although early investigations of the energetics of flap dynamics using picosecond MD showed favourable free energy changes from the closed to the semi-open conformations[26], a thorough description of flap opening and closing requires sampling over much larger timescales.

Coarse-grained models have had moderate success in describing the kinetics of flap opening[25,27], although they lack the atomistic detail required to distinguish conformations as clearly as fully-atomistic models can. This coupled with the fact that there are inconsistencies in the definitions of the less well ordered open state requires a more complete elucidation of protease flap dynamics and energetics using fully-atomistic models in explicit solvent that reach the timescales required for accurate flap kinetics to be discerned.

The volunteer-based GPUGRID project (http://gpugrid.net) delivers the capability of performing large ensembles of explicit solvent molecular dynamics simulations via a custom built molecular dynamics software, ACEMD[28], for use on GPUs. Here, using the GPUGRID infrastructure, we perform an ensemble of 461 fully atomistic molecular dynamics simulations in explicit solvent of unbound wildtype dimeric HIV-1 protease each with an average simulation time of ∼51 ns. Utilising such extensive sampling, we are able to elucidate a more complete description of flap conformations and transitions including the energetics of flap opening and closing. In concurrence with previous studies, the semi-open state is shown to be substantially more favoured whilst a significant population of open-state conformations, in agreement with the curled-open structure of Toth and Borics[22], and a minor population of closed-state conformations are additionally exhibited. The use of explicit solvent makes possible the absolute kinetic comparison of conformational transition timescales with previous NMR investigations and we determine the mean relaxation time between the semi-open and the curled-open conformation in good agreement with the experimental data. Furthermore, we are able to distinguish another class of open conformation, termed the wide-open, in agreement with the structure of Hornak *et al.*[21], thus providing an explicit solvent template structure for the docking of potential inhibitors.

II. MATERIALS AND METHODS

A. Molecular simulation protocol

Atomic coordinates for wildtype dimeric HIV-1 protease were extracted from the 1HHP crystal structure in the Protein Data Bank[29]. The standard AMBER forcefield (ff03)[30] was used to describe all protease parameters. A dianionic state was assigned to the catalytic dyad as, at physiological pH, the unbound protease is unprotonated, whilst mono-protonation is thought to occur upon or after ligand binding[31,32]. The LEAP module of the AMBER 10 software package[33] was then used to build the initial systems. Each system was solvated using atomistic TIP3P water[34] in a cubic box with at least 10 Å distance around the complex and then electrically neutralised with a total ionic concentration of 0.15 M, resulting in a fully atomistic system of 50,437 atoms.

Equilibration simutions were carried out using the NAMD2.6 program[35] on a local cluster. During equilibration the position of the heavy protein atoms were restrained by a 10 kcal/mol/Å^2 spring constant and conjugate-gradient minimisation performed for 1000 steps. The hydrogen atoms and water molecules were then allowed to evolve for a total of 64 ps at 300 K to ensure thorough solvation of the complex and to prevent premature flap collapse[20]. The magnitude of the restraining spring constant was kept at 10 kcal/mol/Å^2 for the first 4 ps, then set to 1 kcal/mol/Å^2 for 20 ps, to 0.05 kcal/mol/Å^2 for 40 ps and then finally to zero. The temperature was maintained at 300 K using a Langevin thermostat with a low damping constant of 0.1/ps and the pressure maintained at 1 atm using a Nose-Hoover barostat. The sytem was finally equilibrated for 10 ns of unrestrained simulation in the isothermal-isobaric ensemble (NPT). The long range Coulomb interaction was handled using the particle mesh Ewald summation method (PME)[36] with grid size 84x84x84. A non-bonded cut-off distance of 9 Å was used with a switching distance of 7.5 Å. For the equilibration runs the SHAKE algorithm[37] was employed on all atoms covalently bonded to a hydrogen atom with an integration timestep of 2 fs.

All production simulations were carried out using ACEMD[28] using a GPU implemention of the PME method[38] on the GPUGRID compute infrastructure in the NVT ensemble with the box size (79.9115Å , 80.0859Å, 80.0437Å) which resulted from the equilibra-

tion at 1 atm and maintained at 300 K using a Langevin thermostat with damping of $0.1/ps$. A longer timestep of 3.5 fs was used thanks to the use of the hydrogen mass repartitioning scheme[39] implemented in ACEMD. This elegant scheme takes advantange of the fact that individual atom masses do not appear explicitly in the equilibrium distribution, therefore changing them only affects the dynamic properties of a system marginally[39], but not the equilibrium distribution. The change in the diffusion coefficient is minimal (10%) and small relative to the approximation that the TIP3 water model has anyway against real water[39]. All other parameters were kept the same as the equilibration phase. Initially 500 runs were submitted on GPUGRID; 39 were not returned by the server, resulting in an ensemble of 461 generated production trajectories. All data collection started after an initial transient time of 3.5 ns in order to achieve some conformational randomization between runs. Coordinate snapshots were generated every 20000 timesteps. On average each trajectory achieved a production length of ∼51 ns yielding a total of ∼340000 conformational states for subsequent post-production analysis.

B. Structural analysis

The molecular dynamics simulations were validated by several methods. A sample trajectory of length 50 ns which remained in the crystal structure flap-conformation was chosen as a representative trajectory. Amide backbone S^2 order parameters from this trajectory were then compared with NMR experimental values obtained from Freedberg *et al.*[13]. Furthermore, root mean squared fluctuations (RMSFs) and positional cross correlations were compared with those obtained from the mean pairwise comparisons of a large set of crystal structures. Structural analysis was performed using the PTRAJ module of the AMBER 10 software package[33] and the Visual Molecular Dynamics (VMD) package[40]. Snapshots were aligned across all amino acid backbone atoms.

The S^2 order parameters were calculated using the method by Hornak *et al.*[41]; the mean of the plateau-region in the last 7 ns in the autocorrelation function for each N-H vector was used. The autocorrelation function, $A(\tau)$, was calculated as:

$$A(\tau) = \frac{\langle \Delta r(t) \Delta r(t+\tau) \rangle_t}{\sqrt{\langle \Delta r(t)^2 \rangle \langle \Delta r(t+\tau)^2 \rangle}}, \tag{1}$$

where $\Delta r(t)$ and $\Delta r(t+\tau)$ are the displacement of the C_α atom of residue i, at times t and $t+\tau$, from the C_α atom of the same residue in the time averaged structure and $\langle \rangle_t$ represents the ensemble average over the running initial time t. The root mean squared fluctuations ($RMSF_i$) of the C_α atoms of each residue were calculated as:

$$RMSF_i = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (\mathbf{r}_i(t) - \langle \mathbf{r}_i \rangle)^2}, \tag{2}$$

where T is the number of snapshots considered in the time trajectory, $\mathbf{r}_i(t)$, the position of the C_α atom of residue i at time t and $\langle \mathbf{r}_i \rangle$, the time-averaged position of the C_α atom of residue i. The positional cross-correlation coefficient between residues i and j was calculated as:

$$C_{ij} = \frac{\langle \Delta r_i(t) \Delta r_j(t) \rangle}{\sqrt{\langle \Delta r_i(t)^2 \rangle \langle \Delta r_j(t)^2 \rangle}}, \tag{3}$$

where $\Delta r_i(t)$ is the displacement of the C_α atom of residue i, at time t, from the C_α atom of the same residue in the time averaged structure. Cross correlation maps (CCMs) were then constructed using arrays of protein residue pairs.

At the time this study was undertaken, there were over 200 structures of HIV-1 protease in the Protein Data Bank, the majority complexed to a variety of inhibitors and peptide sequences and only a minority in the unbounded form. Structures with missing C_α atoms were discarded, resulting in a subset of 171 crystal structures. Analysis of these structures was performed using the Visual Molecular Dynamics (VMD) package[40]; the backbone atoms of each pair of proteases within a set being analysed were first aligned. The average of all pairwise RMSDs of the C_α atoms was then calculated across all residues. Positional cross-correlations between each of the amino acid residues were also calculated using pairwise superimposition and both sets of results compared with the MD simulation trajectory.

C. Free energy analysis

Once suitable reaction coordinates are used to define a number of different conformational states, the relative free energy of the states can be calculated from the population of conformations that exist at a particular point on the reaction coordinate space. Two reaction parameters, α and β, corresponding to the wall-tip to flap-tip distances for each monomer, were used to differentiate the open, closed and semi-open states. Specifically, these were the C_α distances between residues P81 and I50, and P180 with I149 respectively (see Figure 4). The free energy (G) at a point (α, β) in the reaction coordinate space was then given by:

$$G(\alpha, \beta) = -k_B T ln(p(\alpha, \beta)) + C, \tag{4}$$

where T is the temperature, k_B, the Boltzmann constant, $p(\alpha, \beta)$, the conformational population density at each value of the reaction coordinate space and C, a constant

that adjusted the relative zero-point of the data set. Reaction parameters were calculated for all snapshots in the production ensemble and a bin size of 0.5 Å was used to calculate the population density distribution. The error of the free energy at a point was computed by calculating the difference between two half-sets of data over the same reaction coordinate space and is provided in the Supporting Information.

D. Conformational kinetics analysis

For the estimation of the forward and reverse rates we have assumed the simplest two-state approximation[42,43]. For the two conformations of the protein, the semi-open (S) and open (O), the reaction scheme is then given by

$$S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} O \tag{5}$$

where $k_1 = 1/\tau_1$ and $k_{-1} = 1/\tau_{-1}$ are the forward and reverse rate constants and τ_1 and τ_{-1} the forward and reverse relaxation times respectively. The forward and reverse transitions can be measured from the probability that a transition has occurred within time t for the forward ($F_1(t)$) and reverse ($F_{-1}(t)$) cumulative ensemble fractions, where $F_1(t) = 1 - exp(-t/\tau_1)$ and $F_{-1}(t) = 1 - exp(-t/\tau_{-1})$. The mean relaxation time τ_m of the conformational transitions can then be calculated from:

$$\frac{1}{\tau_m} = \frac{1}{\tau_1} + \frac{1}{\tau_{-1}}, \tag{6}$$

The kinetics of flap opening were investigated by estimating the forward, τ_1, and reverse, τ_{-1}, relaxation times between S and O, based on the above two-state approximation. The mean relaxation time, τ_m, could then be calculated through Equation 6 and directly compared to the mean timescales observed in NMR experiments, thus allowing a direct comparison with the experimental data. Linear fitting of $ln(1 - F(t)) = \frac{-t}{\tau}$ allowed τ_1 and τ_{-1} to be determined from the inverse of the gradients. The accuracy of these results were further checked by additionally computing S_{eq} and O_{eq}, the equilibrium populations of conformations S and O respectively and comparing with $\tau_{-1}/\tau_1 = O_{eq}/S_{eq}$.

Based on an analysis of the two corresponding minima for the semi-open (S) and open (O) conformations exhibited in the free energy analysis (see Figure 4) and centred at $\alpha_S = \beta_S = 12.75$ Å and $\alpha_O = \beta_O = 8.75$ Å respectively, the criteria used to define the S and O conformations was that, for each conformation, both α and β were within a circle of radius 1 Å from the corresponding minima. The free energy of transition from state S to O was determined as $\Delta G_{S\rightarrow O} = -k_B T ln(\tau_{-1}/\tau_1)$ and compared to the results obtained from the free energy analysis.

The initial ensemble of conformations containing subpopulations of S and O comprised every snapshot from the first half of each trajectory from a subset of 453 trajectories that had each achieved 35 ns of simulation. The forward $F_1(t)$ and reverse $F_{-1}(t)$ cumulative ensemble fractions were then determined based on first time transitions within 17.5 ns from the corresponding initial snapshot. Finally, in order to test the sensitivity of the results to the choice of radius, the radial value was varied from 0.8 Å to 1.2 Å and the relative change in the kinetic parameters assessed.

III. RESULTS AND DISCUSSION

A. Structural flexibility

Molecular dynamics trajectories were validated by comparing properties of structural flexibility obtained from a sample trajectory with those obtained from NMR experiments[13] as well as an ensemble of crystal structures. Figure 3(a) shows the comparison of amide backbone S^2 order parameters from the experimental results of Freedberg *et al.*[13] with a sample MD trajectory of 50 ns that remained in the semi-open conformation. The computed results exhibit good correlation with the experimental results; the trends characterising the regions of high and low flexibility are reproduced well. Furthermore, the results obtained here are in relatively good quantitative agreement with the experimental results being only marginally greater for the characteristically flexible regions such as the flaps and the fulcrum tip. As expected, these results are more similar to experiment than those used previously in an implicit solvent model by Hornak *et al.*[44] showing that low-viscosity effects may indeed be a factor in altering quantitative agreement.

Figure 3(b) shows the comparison of the C_α root mean squared fluctuation (RMSF) across all protease residues from the same trajectory with the mean pairwise C_α RMSD of 171 crystal structures. In general the MD simulation exhibits significantly increased flexibility compared to the crystal structures. This is due to the fact that the simulation is of the free form of the protease whilst the majority of the crystal structures are taken from the bound form; reduced flexibility is thus expected. The flexibility profile from the MD simulation is very similar to that of the crystal structures, exhibiting peaks as expected in the flap elbows, the flap tips, the fulcrum tip and the cantilever and the wall tip (see Figure 1 for definitions). There are extra peaks at the location of the fireman's grip and the tip of the dimer interface. The existence of ligands in the crystal structures thus reduce the flexibility of the dimer interface and the catalytic region in good agreement with previous investigations[15].

Cross correlation analysis (see Figure 3) yields similar secondary structural relationships between (a) the ensemble of crystal structures and (b) the MD simulation. The β-hairpin structures such as the fulcrum (purple), cantilever (cyan), flap elbows (yellow) and flap tips

(red) are represented by highly correlated regions perpendicular to the diagonal. The α-helices (gray: residues 85-95 and 184-194) are represented by the correlational broadening along the diagonal. Correlational features between different secondary structure regions include the high correlation between the fulcrum and the cantilever, the wall (orange) and the fireman's grip (blue) and the interdigitated β-sheeted dimer interface (green) as well as anti-correlated motion between the flaps with the cantilever, the flap-elbows with the fulcrum tip and the flap elbows with the dimer interface. High anti-correlation is also shown between the flaps of each monomer with the fireman's grip region of the other as well as the flap of each monomer with the other flap. The crystal structure flap tips exhibit no correlation with each other whilst the MD simulation exhibit strong correlation. This is again explained by the different flap conformations between the predominantly bound crystal structures (which are in the trans-handed form) and the unbound semi-open conformations of the MD simulation (which are in the cis-handed form) and concurs well with previous investigations[25].

B. Energetics of flap opening and closing

Ensemble MD simulations from 461 different trajectories with an average simulation time of ∼51 ns allowed a large conformational ensemble to be generated (∼340000 snapshots). In order to describe the energetics of flap opening, assuming the flap curling mechanism described previously[18,22,24], appropriate reaction coordinates were defined, followed by the determination of the frequency distribution of conformations across such coordinates. These were defined as α and β corresponding to the C_α distances between P81-I50 and P180-I149 residues respectively (see Figure 4(a)). The free energy, derived from the frequency distribution, across the two dimensional reaction space was then plotted as a contour map (see Figure 4). The error associated to this map was determined by computing the difference between two half-data sets separated by over 20 ns and was within 0.2 kcal/mol for most of the reaction coordinate space accessed by the flap conformations (see Supplementary Online Material).

A small value of α and β corresponds to the curling back of the flaps of monomer A and B respectively towards their wall residues and describes the curled open state reported previously. A large value of α and β requires the flaps to cross over each other and thus describes the closed conformation where flap-handedness is reversed. Five distinct regions are identifiable on the map. The two principal regions correspond to a) the semi-open conformation (S), characterised by a free energy minima of -2.7 ± 0.1 kcal/mol in a region centred at α=β= 12.75 Å and b) the open conformation (O), characterised by a free energy minima of -1.8 ± 0.1 kcal/mol centred at α=β= 8.75 Å. The open conformation results in the exposure of the active site cavity (especially the aspartyl dyad) directly along the substrate binding trench, thus making it geometrically possible for a substrate to bind in the downward direction from the top of the flaps. Furthermore, there is a $\Delta G_{S\rightarrow O}$ = 0.9 ± 0.2 kcal/mol penalty to flap opening probably related ttoo the molecular rearrangement of the I50 and I149 residues on the flap tip that exhibit hydrophobic interactions with each other in the semi-open conformation but with the wall residues, specifically V32,P79, T80 and P81, in the open-conformation.

Two more regions, c) curl1 and d) curl2 correspond to either the flaps of monomer A or B respectively being curled back in an open conformation whilst the other flap is extended towards the curled flap and over the active site. Structurally, they are accounted for by the fact that the hydrophobic contacts between the flaps in such conformations are still maintained whilst the curled flap also manifests hydrophobic interactions as expected with its respective wall residues. Visual inspection of such conformations confirms that they proceed rapidly either to the open or semi-open states within any given trajectory. Although obscured from the top, the curling back of one flap does laterally expose the binding pocket to a significant degree. Previous investigations have reported the likelihood of lateral inhibitor dissociation from mutant proteases[45], whilst a recent computational study showed small peptide binding laterally to HIV-1 protease without the conventional opening of the flaps[46].

A final identified conformation is e) the closed conformation (C), characterised by a shallow free energy minima centred at α=β= 19 Å and structurally characterised by the flaps both extending away from their respective walls and across the centre of the active site bringing each flap tip closer to the opposite wall residues. This requires the reversal of flap-handedness and therefore corresponds to the closed conformation. The free energy penalty of flap closing from the semi-open state is $\Delta G_{S\rightarrow C}$ = 2.4 ± 0.4 kcal/mol. This qualitatively agrees with and refines early calculations on flap closing from picosecond umbrella sampling studies[26] which gave a flap closing free energy of 7 ± 3 kcal/mol; individual estimates varied from 2 ± 2 kcal/mol to 13 ± 5 kcal/mol. The quantitative discrepancy of these earlier investigations is likely to arise from the limited sampling available at a picosecond timescale, although our results are in good agreement with the lower bound obtained by Rick *et al*[26]. The global minima associated with the semi-open state, together with the local minima of the open and closed states, thermodynamically support previous experimental findings that the semi-open conformation is dominant in the unbound form of the protease, with a smaller population of open conformations and minimal population of closed conformations[12–14].

The free energy of flap closing is of particular importance to end-state thermodynamic methods that calculate free energies of ligands binding to HIV-1 protease. Such methods compute thermodynamic differ-

ences between unbound and bound states for both the solvated and vacuum phase, instead of sampling direct binding events in solution[47–50], but do not implicitly factor the free energy differences associated with conformational changes upon binding. Including the 2.4 kcal/mol calculated here due to conformational change, together with other possible state-change phenomena upon binding that are usually excluded from such models, such as the free energy of explicit water-mediated protein-ligand interactions[51,52] and the changes in key protonation states upon binding[32] could allow accurate HIV-1 protease-ligand binding free energies to be recaptured[53].

The various conformational transitions of the flaps, exhibited here, also gives rise to additional structural signatures. Specifically, these are differing RMSDs of the flaps relative to closed and semi-open crystal structures, flap-tip distances and protease backbone RMSDs. The dynamics of two representative trajectories with these metrics as well as the corresponding α and β metrics are provided in the Supplementary Online Material. Representative structures of all conformations described herein are also provided in PDB format in the Supplementary Online Material.

C. Kinetics of rapid flap opening

Freedberg *et al.*[13] revealed a rapid opening fluctuation of the flap tips within the 10 ns timescale of overall tumbling. The timescales sampled in this study permit a kinetic analysis of this conformational transition, defined herein as the semi-open to open transition. Our use of explicit solvent models allows us to investigate whether the open conformation exhibited here corresponds to the NMR experiments as well as directly compare our calculations to the experimental mean relaxation times of such a conformational fluctuation.

By calculating the forward ($F_1(t)$) and reverse ($F_{-1}(t)$) cumulative ensemble fractions of first transitions, we determined the forward (τ_1) and reverse (τ_{-1}) rates from the semi-open (S) to the open (S) state using the two state approximation for conformational transitions. Figure 5 shows the linear fitting of $ln[1 - F_1(t)]$ and $ln[1 - F_{-1}(t)]$ for trajectories that make transitions into the open conformation and in the reverse direction. Both the $ln[1-F_1(t)]$ and $ln[1-F_{-1}(t)]$ functions exhibit good linearity and result in $\tau_1 = 36.5$ ns and $\tau_{-1} = 10.4$ ns. This gives a mean relaxation time, $\tau_m = 8.1$ ns, in good agreement with the 10 ns timescale of the NMR data. Furthermore, the reverse rate is far steeper than the forward, in good accordance with the relative free energies of the two states (see Supplementary Online Material for plots of the forward and reverse cumulative ensemble fractions).

If the ensemble of runs is in equilibrium, the validity of the kinetics calculated above can be assessed from the normalised equilibrium populations of the semi-open (S_{eq}) and open (O_{eq}) conformations. Both conformations exhibit stable populations within the ensemble confirming that the ensemble is in equilibrium; the corresponding equilibrium population is dominated by the semi-open state, with $S_{eq} = 0.10$ and $O_{eq} = 0.02$. The free energy of transition from state S to O, is calculated both from $\Delta G_{S\rightarrow O} = -k_B T ln(\tau_{-1}/\tau_1)$ and from $\Delta G_{S\rightarrow O} = -k_B T ln(O_{eq}/S_{eq})$, yielding $\Delta G_{S\rightarrow O} = 0.8$ kcal/mol and $\Delta G_{S\rightarrow O} = 1$ kcal/mol respectively. Both results are in excellent agreement with $\Delta G_{S\rightarrow O} = 0.9 \pm 0.1$ kcal/mol obtained from Figure 4 as well as the timescale exhibited in previous NMR data.

When interpreting these results it is important to consider the sensitivity of the kinetics to changing the radius used to define each state. Altering the radius between 0.8 - 1.2 Å gives τ_1 in the range 46.5 - 29.7 ns and τ_{-1} in the range 13.5 - 9.0 ns. This corresponds to a change in the forward and reverse rate constants within a factor of two. The overall mean relaxation rate τ_m varies between 10.4 - 6.9 ns, indicating that the results are fairly insensitive to a radial variation.

D. Wide-open conformational dynamics

In addition to the rapid semi-open to open transition revealed by Freedberg *et al.*[13], Ishima *et al.*[12] additionally revealed a slower mode of opening on the 100 μs timescale with a more disordered flap conformation. In our investigation, as well as the five flap conformations already described, a further rare conformation is exhibited, which is not clearly differentiable in terms of the α and β metrics for flap curling, but is in terms of other metrics. This conformation is denoted the wide-open conformation and is characterised by a substantially increased opening of the protease active site.

Figures 6(i) and 6(ii) show the conformation from the top and lateral views respectively. Computation of relative C_α distances show that a large difference is exhibited between the wide-open and the open conformation (see Supplementary Online Information). The wide-open conformation is structurally characterised by a large vertical motion of the flaps away from the active site centre with averaged C_α distances increasing up to 4 Å away from other regions of the protease as compared to the open structure. This is coupled to a downward motion of the flap elbows and the cantilever towards the dimer interface with C_α distances decreasing by up to 4 Å. This is consistent with a large rotational rearrangement of each monomer around its fulcrum and results in the opening up of the active site binding trench both laterally and from the top substantially more than the open conformation. Not only is the catalytic dyad clearly exposed, but further, the flap tips are pulled back beyond the wall tips; unlike the open conformation, the lateral active site cavity of the wide-open conformation is thus bounded not by the closest distance between the flaps but by the wall-tip distance. The binding trench is thus able to accommodate substantially larger substrates than in the

open conformation.

Analysis of the structural metrics using a representative trajectory leading to the wide-open state (see Figure 6(iii)) reveals the distinct signature of the wide-open conformation. The initial semi-open state, characterised by C_α flap-tip distances smaller than 10 Å, is followed by curled states between 14 ns and 20 ns and then an open state, with flap-tip distances smaller than 20 Å, α and β metrics approximately 9 Å and where the flap-elbows (yellow) approach the fulcrum-tip (purple). After a further 5 ns, the flap C_α RMSD relative to both closed and open structures increases to over 10 Å and 7.5 Å respectively. The flap tip distance approaches a maximum of 30 Å and the backbone RMSD reaches values in excess of 4.5 Å. Furthermore, although the α and β metrics exhibit a large range of values, they do not represent any single conformation described in Figure 4. The substantial change in all these metrics at 25 ns correspond to a transition into the wide-open conformation.

Our results suggest that initial curling of the flap-tips, which results in the open state, may be a pre-requisite for the subsequent transition into the wide-open state. Once the wide-open state is accessed, all curling conformations are accessible and uncurling of the flap-tips does not return the system to the semi-open state. Instead, upon uncurling, the flaps move further away from the other regions of the protease, explaining the large differences in C_α-distances exhibited between the wide-open and the open conformations.

Previous studies have shown that a significant rearrangement of the hydrophobic core of the protease may facilitate conformational changes[54]. A similar rearrangement is observed here. The hydrophobic core is principally composed of residues L33, M36 and L38 in the flap elbows, residues V11, I13, I15 and A22 in the fulcrum, residues I62, I64, I66, V75 and V77 in the cantilever and residue I85 just before the α-helix. The transition from the open to the wide-open conformation is triggered by structural rearrangement of the M36 residue, which is initially exposed to solvent, into the hydrophobic core consisting of L33, L38, L62, V75 and V77. This causes conformational instability within the core leading to further rearrangements, specifically I66 towards V11 and I64 away from I13, A22 and I85, resulting in the sliding of the cantilever down towards the dimer interface and the wide-opening of the active site.

The wide-open conformation identified here is exhibited in only four simulations in the entire ensemble thus precluding the accurate determination of either energetics or kinetics for this mode of opening. For this reason it is not possible to calculate the free energy profile of the wide-open structure due to the overwhelming relative frequency of the semi-open conformation. However, the conformation is in agreement with the structure obtained by Hornak *et al*[21] using an implicit solvent model. Furthermore, the much smaller frequency of exhibition of the wide-open state with respect to the open state indicates that this mode of opening is qualitatively likely to be the mode corresponding to the 100μs opening event captured in the NMR experiments.

IV. CONCLUSIONS

Using an ensemble of 461 fully-atomistic molecular dynamics simulations in explicit solvent, with an average simulation time of 50 ns/run, we have characterised the conformational states available to the β-hairpin flaps of HIV-1 protease and investigated the energetics of flap opening and closing. Our investigation confirms that the semi-open conformation of the free protease is the dominant population, in agreement with existing crystal structures, as well as reproducing the rapid timescale flap-tip fluctuations which result in the curled open structure (termed 'open' herein) observed in solution NMR data[13] and in previous implicit solvent molecular dynamics studies[18,22,24]. Furthermore, we also reproduce a further full flap opening (termed 'wide-open' herein) also observed in solution NMR[12] and recent MD studies[21] as well as the closed flap conformation corresponding to the reversal of flap-handedness exhibited in bound crystal structures.

Extensive sampling allows the unbiased calculation of the free energy of flap-conformational changes. The free energy of flap opening from the semi-open state to the open state is 0.9 ± 0.2 kcal/mol; the free energy change of flap closing from the semi-open state is 2.4 ± 0.4 kcal/mol. The latter term is an important correction factor for thermodynamic end-state methods that do not factor in conformational changes upon binding[47,48,53].

The use of explicit solvent allows direct comparison of mean relaxation times with solution NMR data[12,13]. Kinetic analysis confirms the mean relaxation time of flap-opening to be approximately 8 ns, in excellent agreement with the NMR result of within 10 ns. However, some care has to be taken when interpreting the results. Even though the use of explicit solvent should improve the accuracy of the kinetics, the energetics and kinetics of flap opening due to curling may be also be partially sensitive to the choice of forcefield. Recently, the Amber03 forcefield was shown to significantly stabilize helical structure[55]. However, as the protease is a largely β-sheeted protein, such effects are likely to be minimal. Wide-opening of HIV-1 protease flaps is a rare event, occurring only four times in our investigations and thus not permitting any energetic or kinetic analysis. However, the rare occurrence of wide-opening compared to that of opening suggests that the wide-open structure has a relaxation several orders of magnitude slower than the open structure and may therefore correlate with the slower ($\sim$100 μs) mode of opening reported in the NMR measurements.

The fully atomistic resolution of the two distinct open structures (open and wide-open) of HIV-1 protease raise important questions regarding the various modes of ligand binding. The open structure is geometrically capa-

ble of binding small peptides and/or several inhibitors as noted previously by Scott and Schiffer[18]. This is supported by the study of Toth and Borics[24] showing the full closure of the flaps from the open structure, once a ligand is placed in the active site and recently by Pietrucci *et al.*[46] showing that binding of small peptides does not require the opening of the flaps if the peptide is small enough to bind laterally to the active site.

However, the open structure is unlikely to be able to bind GagPol precursors effectively, owing to the larger size of the latter inducing considerable steric obstructions. It is more likely that GagPol binding by HIV-1 protease predominantly occurs through a collision complex in the wide-open state. The speed of flap-gating is an important factor moderating the association rates of ligands. The four orders of magnitude difference in the mean relaxation times for open and wide-open states correspond to the 'fast' and 'slow' gating mechansims formulated by Chang *et al.*[56] suggesting that the binding site always appears open to small ligands but not necessarily to GagPol chains. Only binding of small ligands to HIV-1 protease is then likely to be diffusion limited whilst GagPol processing will be limited by the wide-opening transition timescale of the protease. Furthermore, this may explain the observed differences in the rate of processing of various cleavage site peptides as compared to Gag and GagPol chains[57,58].

Finally, the explicit solvent structures of the flap-conformations reported herein are provided in PDB format in the Supplementary Online Material. These structures constitute additional conformational targets that may be of general interest to future drug-design and docking investigations that make use of the multiple flap-gating mechanisms reported herein in the search for more effective inhibitors of HIV-1 protease.

Acknowledgments

We thank Mathew J. Harvey and Dr. Toni Giorgino for assistance. We are grateful to the volunteers of GPU-GRID for providing the computational time for this investigation. GDF acknowledges support from the Ramon y Cajal scheme and from the Spanish national research plan.

* Electronic address: kashif.sadiq@upf.edu
† Electronic address: gianni.defabritiis@upf.edu

[1] Wlodawer, A. and Erickson, J. W. Structure-based inhibitors of HIV-1 Protease. Annual Review of Biochemistry 62:543–585, 1993.
[2] Wlodawer, A. and Vondrasek, J. Inhibitors of HIV-1 protease: A major success of structure-assissted drug design. Ann. Rev. Biophys. Biomol. Struct. 27:249–284, 1998.
[3] Alfano, M. and Poli, G. The HIV life cycle: Multiple targets for antiretroviral agents. Drug Design Reviews - Online 1:83–92, 2004.
[4] Ohtaka, H., Schon, A., and Freire, E. Multidrug resistance to HIV-1 protease inhibition requires cooperative coupling between distal mutations. Biochemistry 42:13659–13666, 2003.
[5] Wlodawer, A., Miller, M., Jaskólski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., and Kent, S. B. H. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. Science 245:616–621, 1989.
[6] Prabu-Jeyabalan, M., Nalivaika, E., and Schiffer, C. A. How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. Journal of Molecular Biology 301:1207–1220, 2000.
[7] Prabu-Jeyabalan, M., Nalivaika, E., and Schiffer, C. A. Substrate shape determines specificity of recognition for HIV-1 protease: Analysis of crystal structures of six substrate complexes. Structure 10:369–381, 2002.
[8] Prabu-Jeyabalan, M., Nalivaika, E., King, N. M., and Schiffer, C. A. Viability of a drug-resistant human immunodeficiency virus type 1 protease variant: Structural insights for better antiviral therapy. Journal of Virology 10(2):1306–1315, 2003.
[9] Prabu-Jeyabalan, M., Nalivaika, E., King, N. M., and Schiffer, C. A. Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. Journal of Virology 78(22):12446–12454, 2004.
[10] Martin, P., Vickrey, J. F., Proteasa, G., Jimenez, Y. L., Wawrzak, Z., Winters, M. A., Merigan, T. C., and Kovari, L. C. "Wide-open" 1.3 A structure of a multidrug-resistant HIV-1 protease as a drug target. Structure 13(12):1887–1895, 2005.
[11] Layten, M., Hornak, V., and Simmerling, C. The open structure of a multi-drug-resistant HIV-1 protease is stabilized by crystal packing contacts. J. Am. Chem. Soc. 128:13360–13361, 2006.
[12] Ishima, R., Freedberg, D. I., Wang, Y. X., Louis, J. M., and Torchia, D. A. Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function. Structure 7:1047–1055, 1999.
[13] Freedberg, D. I., Ishima, R., Jacob, J., Y-X.Wang, Kustanovich, I., Louis, J. M., and Torchia, D. A. Rapid structural fluctuations of the free HIV protease flaps in solution: Relationship to crystal structures and comparison with predictions of dynamics calculations. Protein Science 11:221–232, 2002.
[14] Katoh, E., Louis, J. M., Yamazaki, T., Gronenborn, A. M., Torchia, D. A., and Ishima, R. A solution NMR study of the binding kinetics and the internal dynamics of an HIV-1 protease-substrate complex. Protein Sci 12(7):1376–1385, Jul, 2003.
[15] Zoete, V., Michielin, O., and Karplus, M. Relation between sequence and structure of HIV-1 protease inhibitor complexes: A model system for the analysis of protein flexibility. J. Mol. Biol. 315:21–52, 2002.
[16] Piana, S., Carloni, P., and Rothlisberger, U. Drug resistance in HIV-1 protease: Flexibility-assisted mechanism

of compensatory mutations. Protein Science 11:2393–2402, 2002.

[17] Seibold, S. A. and Cukier, R. I. A molecular dynamics study comparing a wild-type with a multiple drug resistant HIV protease: differences in flap and aspartate 25 cavity dimensions. Proteins 69(3):551–565, Nov, 2007.

[18] Scott, W. R. P. and Schiffer, C. A. Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. Structure 8:1259–1265, 2000.

[19] Perryman, A. L., Lin, J., and McCammon, J. A. HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: Possible contributions to drug resistance and a potential new target site for drugs. Protein Science 13:1108–1123, 2004.

[20] Meagher, K. L. and Carlson, H. A. Solvation Influences Flap Collapse in HIV-1 Protease. Proteins: Struct. Funct. Bioinf. 58:119–125, 2005.

[21] Hornak, V., Okur, A., Rizzo, R. C., and Simmerling, C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. Proc. Natl. Acad. Sci. U. S. A. 103:915–920, 2006.

[22] Toth, G. and Borics, A. Flap opening mechanism of HIV-1 protease. J. Mol. Graphics Modell. 24:465–474, 2006.

[23] Hornak, V., Okur, A., Rizzo, R. C., and Simmerling, C. HIV-1 protease flaps spontaneously close to the correct structure in simulations following manual placement of an inhibitor into the open state. J. Am. Chem. Soc. 128(9):2812–2813, Mar, 2006.

[24] Toth, G. and Borics, A. Closing of the flaps of HIV-1 protease induced by substrate binding: A model of a flap closing mechanism in retroviral aspartic proteases. Biochemistry 45:6606–6614, 2006.

[25] Tozzini, V., Trylska, J., en Chang, C., and McCammon, J. A. Flap opening dynamics in HIV-1 protease explored with a coarse-grained model. J Struct Biol 157(3):606–615, Mar, 2007.

[26] Rick, S. W., Erickson, J. W., and Burt, S. K. Reaction path and free energy calculations of the transition between alternate conformations of HIV-1 protease. Proteins: Struct. Funct. Genet. 32:7–16, 1998.

[27] Chang, C.-E., Shen, T., Trylska, J., Tozzini, V., and McCammon, J. A. Gated binding of ligands to HIV-1 protease: Brownian dynamics simulations in a coarse-grained model. Biophysical Journal 90:3880–3885, 2006.

[28] Harvey, M. J., Giupponi, G., and Fabritiis, G. D. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. Journal of Chemical Theory and Computation 5(6):1632–1639, June, 2009.

[29] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., , and Bourne, P. E. The protein data bank. Nucleic Acids Research 28:235–242, 2000.

[30] Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., and Lee, T. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J. Comput. Chem. 24:1999–2012, 2003.

[31] Kovalskyy, D., Dubyna, V., Mark, A. E., and Korenelyuk, A. A molecular dynamics study of the structural stability of HIV-1 protease under physiological conditions: The role of Na+ ions in stabilizing the active site. Proteins: Struct. Funct. Bioinf. 58:450–458, 2005.

[32] Wittayanarakul, K., Hannongbua, S., and Feig, M. Accurate prediction of protonation state as a prerequisite for reliable MM-PB(GB)SA binding free energy calculations of HIV-1 protease inhibitors. J. Comput. Chem. 29(5):673–685, 2008.

[33] Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, III, T. E., DeBolt, S., Ferguson, D., Seibel, G., and Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Computer Physics Communications 91:1–41, 1995.

[34] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79:926–935, 1983.

[35] Kale, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K., and Schulten, K. NAMD2: Greater scalability for parallel molecular dynamics. J. Comput. Phys. 151:283–312, 1999.

[36] Essmann, U., Perera, L., Berkowitz, M. L., and Darden, T. A smooth particle mesh Ewald method. J. Chem. Phys. 103:8577–9593, 1995.

[37] Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. J. Comput. Phys. 23:327–341, 1977.

[38] Harvey, M. and De Fabritiis, G. An Implementation of the Smooth Particle Mesh Ewald Method on GPU Hardware. Journal of Chemical Theory and Computation 5(9):2371–2377, 2009.

[39] Feenstra, K., Hess, B., and Berendsen, H. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. Journal of Computational Chemistry 20(8):786–798, 1999.

[40] Humphrey, W., Dalke, A., and Schulten, K. VMD - Visual molecular dynamics. J. Mol. Graphics 14:33–38, 1996.

[41] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65:712–725, 2006.

[42] Voter, A. F. Parallel replica method for dynamics of infrequent events. Phys. Rev. B 57(22):R13985–R13988, Jun, 1998.

[43] Marianayagam, N. J., Fawzi, N. L., and Head-Gordon, T. Protein folding by distributed computing and the denatured state ensemble. Proc Natl Acad Sci U S A 102(46):16684–16689, 2005.

[44] Hornak, V., Okur, A., Rizzo, R. C., and Simmerling, C. HIV-1 protease flaps spontaneously close to the correct structure in simulations following manual placement of an inhibitor into the open state. J. Am. Chem. Soc. 128(9):2812–2813, 2006.

[45] Sadiq, S. K., Wan, S., and Coveney, P. V. Insights into a mutation-assisted lateral drug escape mechanism from the HIV-1 protease active site. Biochemistry 46:14865–14877, 2007.

[46] Pietrucci, F., Marinelli, F., Carloni, P., and Laio, A. Substrate binding mechanism of HIV-1 protease from explicit-solvent atomistic simulations. J. Am. Chem. Soc. 131(33):11811–11818, 2009.

[47] Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham,

T. E. r. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc. Chem. Res. 33(12):889–897, 2000.

[48] Wang, W. and Kollman, P. A. Computational study of protein specificity: The molecular basis of HIV-1 protease drug resistance. Proc. Natl. Acad. Sci. U. S. A. 98:14937–14942, 2001.

[49] Lepsik, M., Kriz, Z., and Havlas, Z. Efficiency of a second-generation HIV-1 protease inhibitor studied by molecular dynamics and absolute binding free energy calculations. Proteins: Struct. Funct. Bioinf. 57:279–293, 2004.

[50] Stoica, I., Sadiq, S. K., and Coveney, P. V. Rapid and accurate prediction of binding free energies of saquinavir-bound HIV-1 proteases. J. Am. Chem. Soc. 130:2639–2648, 2008.

[51] Hamelberg, D. and McCammmon, J. A. Standard free energy of releasing a localized water molecule from the binding pockets of proteins: Double-decoupling method. J. Am. Chem. Soc. 126:7683–7689, 2004.

[52] Lu, Y., Yang, C.-Y., and Wang, S. Binding free energy contributions of interfacial waters in HIV-1 protease/inhibitor complexes. J. Am. Chem. Soc. 128(36):11830–11839, 2006.

[53] Sadiq, S. K., Wright, D. W., Kenway, O. A., and Coveney, P. V. Accurate Ensemble Molecular Dynamics Binding Free Energy Ranking of Multidrug-Resistant HIV-1 Proteases. Journal of Chemical Information and Modeling, 2010.

[54] Foulkes-Murzycki, J., Scott, W., and Schiffer, C. Hydrophobic sliding: a possible mechanism for drug resistance in human immunodeficiency virus type 1 protease. Structure 15(2):225–233, 2007.

[55] Best, R. and Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. Journal of Physical Chemistry B 113(26):9004–9015, 2009.

[56] Chang, C.-E., Shen, T., Trylska, J., Tozzini, V., and McCammon, J. A. Gated binding of ligands to HIV-1 protease: Brownian dynamics simulations in a coarse-grained model. Biophysical Journal 90:3880–3885, 2006.

[57] Maschera, B., Darby, G., Palu, G., Wright, L. L., Tisdale, M., Myers, R., Blair, E. D., and Furfine, E. S. Human immunodeficiency virus: Mutations in the viral protease that confer resistance to saquinavir increase the dissociation rate constant of the protease-saquinavir complex. J. Biol. Chem. 271:33231–33235, 1996.

[58] Pettit, S. C., Henderson, G. J., Schiffer, C. A., and Swanstrom, R. Replacement of the P1 amino acid of human immunodeficiency virus type 1 gag processing sites can inhibit or enhance the rate of cleavage by the viral protease. Journal of Virology 76(20):10226–10233, 2002.

Figure Legends

Figure 1: (a) The three-dimensional structure of dimeric HIV-1 protease composed of two identical monomers (Monomer A: residues 1-99 and Monomer B: residues 100-198). The dimer interface is formed principally from an interdigitated 4-stranded β-sheet (residues 1-5,96-99,100-104 and 195-198: green) and the fireman's grip (residues 24-29 and 123-128: blue) that provide the base for the active site and the catalytic Asp dyad (D25/D124). The active site is enclosed on top by flexible β-hairpin structures called the 'flaps' (residues 43-58 and 142-157: red) and laterally obscured to some extent by the 'walls' (residues 80-83 and 179-182: orange). Additional features of each monomer are the flap elbows (residues 37-42 and 136-141: yellow), the cantilever (residues 59-75 and 158-174: cyan), the fulcrum (residues 10-23 and 109-122: purple) and the α-helix (residues 87-95 and 186-194: gray). (b) Space filling representation of a peptidic substrate in the active site (from the 1KJ4 crystal structure) showing the wall residues (orange) and the fireman's grip (blue). Lateral obscuring of the active site creates a uni-directional diagonal binding trench along which the substrate can bind. (c) Unbound crystal structures are found in the semi-open conformation (1HHP), bound structures in the closed flap conformation (1KJ4). The two conformations have opposite flap handedness (top); the binding of the substrate causes closure of the active site cavity (bottom). The substrate has been removed from the closed structure for clarity.

Figure 2: Comparison of (a) experimental S^2 order parameters (black) with those calculated from molecular dynamics simulation (red). (b) Comparison of the residue decomposed C_α RMSFs of crystal structures (black) with those calculated from molecular dynamics simulation (red), averaged across both monomers. Calculated errors represent the difference between monomers. Corresponding regions of the protease are additionally colour coded using the scheme from Figure 1(a).

Figure 3: Comparison of the cross correlation map (CCM) of C_α atoms of all residues from (a) crystal structures with (b) CCM from molecular dynamics simulation. Cross correlations are mapped from 1 for perfect correlation to -1 for perfect anti-correlation. Corresponding regions of the protease are additionally colour coded using the scheme from Figure 1(a).

Figure 4: Energetic contour map in kcal/mol of flap conformations of HIV-1 protease calculated using the frequency distribution over metrics α and β corresponding to the C_α distances between P81-I50 and P180-I149 residues respectively. The wall and flaps are in orange and red ribbons respectively and the aspartyl dyad (D25/D124) in blue space filling representation. (a) The main minima (G = -2.7 $\pm$ 0.1 kcal/mol) corresponds to the semi-open (S) conformation with (b) a secondary minima (G = -1.8 $\pm$ 0.1 kcal/mol) representing the open (O) conformation in which both the flaps curl towards the wall exposing the aspartyl dyad. (c) The curl 1 and (d)

curl 2 conformations correspond to the flap of only one monomer curling towards the wall, the other remaining semi-open and vice versa respectively. (e) A final minor minima represents the closed (C) conformation (G= -0.3 $\pm$ 0.3 kcal/mol) in which flap handedness is reversed and both flap tips are maximally distant from their respective wall tip residues. From the semi-open conformation the free energy of flap closing is thus $\Delta G_{S\rightarrow C}$ = 2.4 $\pm$ 0.4 kcal/mol whilst the free energy of flap opening is $\Delta G_{S\rightarrow O}$ = 0.9 $\pm$ 0.2 kcal/mol. Representative structures of the semi-open open, closed, curl1 and curl2 conformations are available in PDB format in the Supporting Information as well as the free energy error contour map across the reaction coordinate space.

Figure 5: Plot of $ln[1 - F_1(t)]$ (black line) and $ln[1 - F_{-1}(t)]$ (red line) where $F_1(t)$ and $F_{-1}(t)$ are the cumulative ensemble fractions of trajectories that make forward and reverse transitions between the semi-open and open conformations respectively. The semi-open and open conformations are circles of radius 1 Å centred at the free energy minima $\alpha_S = \beta_S$ = 12.75 Å and $\alpha_O = \beta_O$ = 8.75 Å respectively. Linear interpolation yields gradients -1/τ_1 and -1/τ_{-1}, where τ_1 = 36.5 ns and τ_{-1} = 10.4 ns are the forward and reverse relaxation times for the conformational change.

Figure 6: (i) Top and (ii) lateral view of the wide open conformation of HIV-1 protease using the same representation scheme as Figure 4. (iii) The characteristic structural signature of the wide open conformation calculated using various distance and RMSD metrics from a representative trajectory. Structures at corresponding points in the trajectory are also shown. The wide open-conformation is preceded by the open conformation and marked by flap C_α RMSDs relative to closed and semi-open structures, greater than 10 Å and 7 Å respectively. The C_α flap-tip distance approaches 30 Å and a large backbone RMSD compared to 1HHP (more than 4.5 Å). The conformation is characterised by upwards vertical movement of the flaps and downward movement of the flap-elbows and cantilever associated with a relative rotation of each monomer about the dimer interface. This is caused by internal rearrangement of the hydrophobic core of each monomer and results in large exposure to the active site aspartyl dyad. A representative structure of the wide-open conformation is available in PDB format in the Supplementary Online Material.

Figures

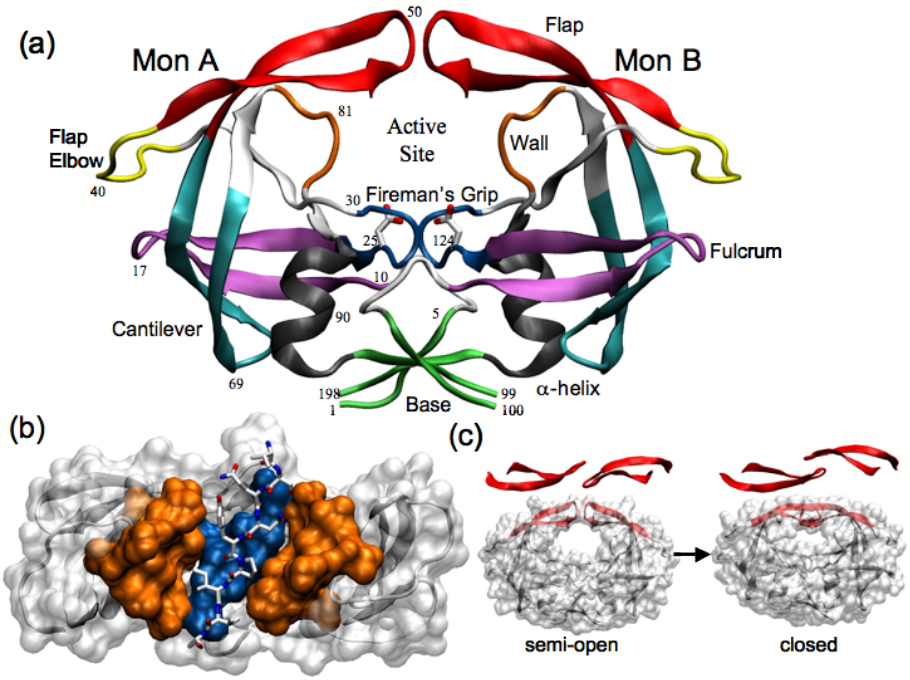

FIG. 1:

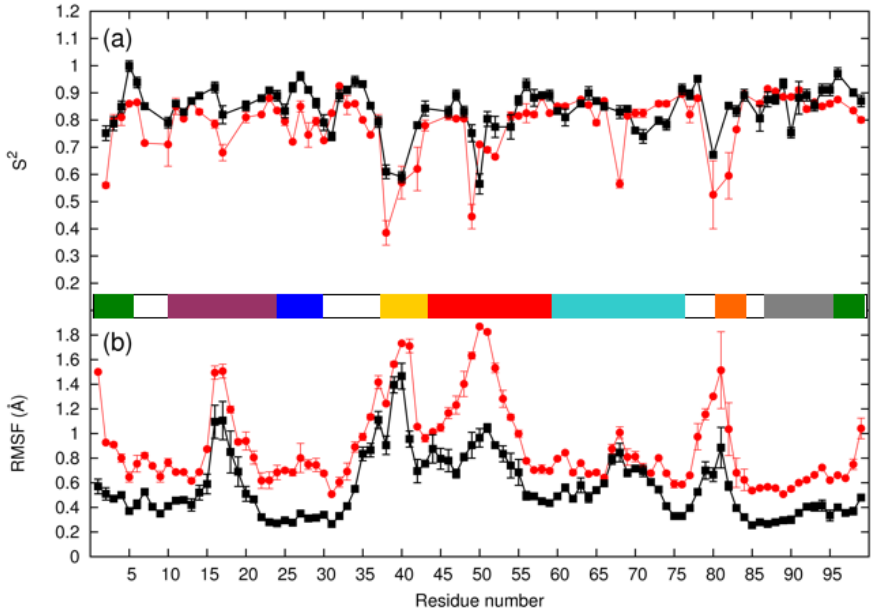

FIG. 2:

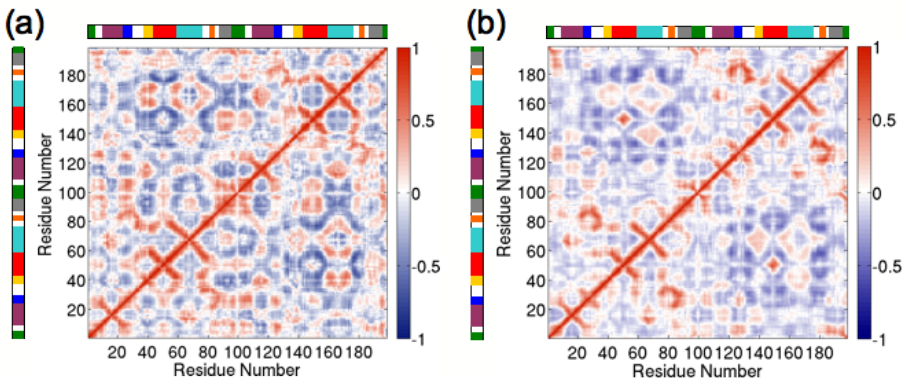
(a)
(b)
Residue Number
Residue Number
20 40 60 80 100 120 140 160 180
1
0.5
0
-0.5
-1

FIG. 3:

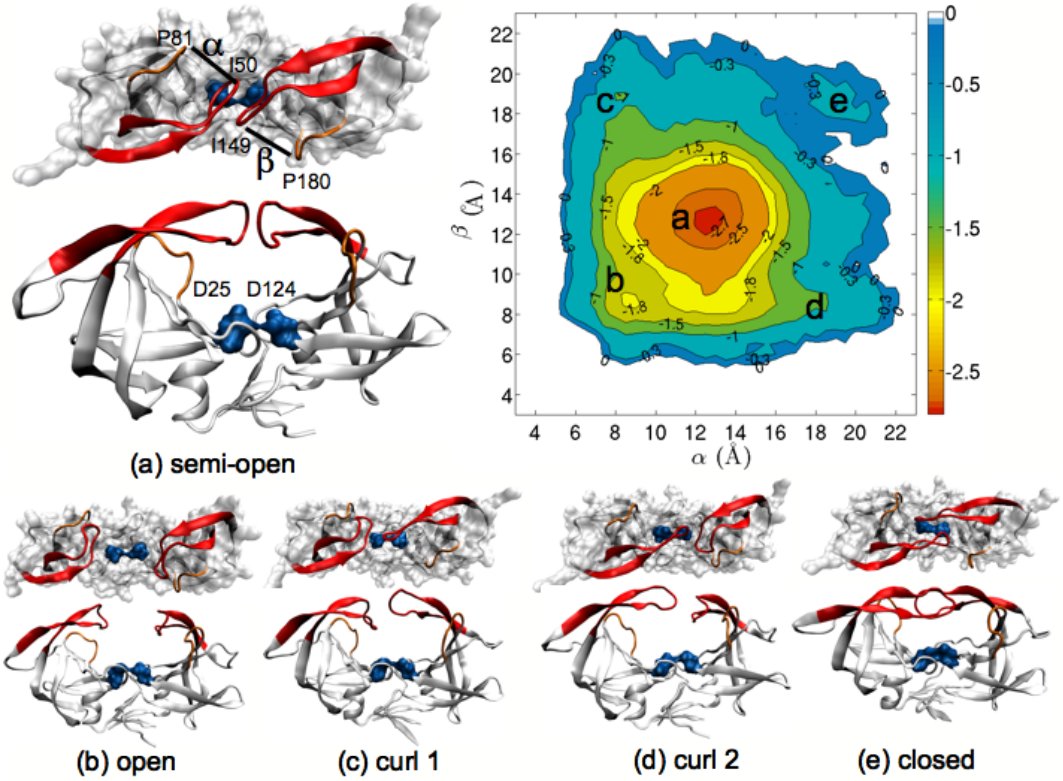
P81
α
I50
I149
β
P180
D25
D124
(a) semi-open
a
b
c
d
e
β (Å)
α (Å)
(b) open
(c) curl 1
(d) curl 2
(e) closed

FIG. 4:

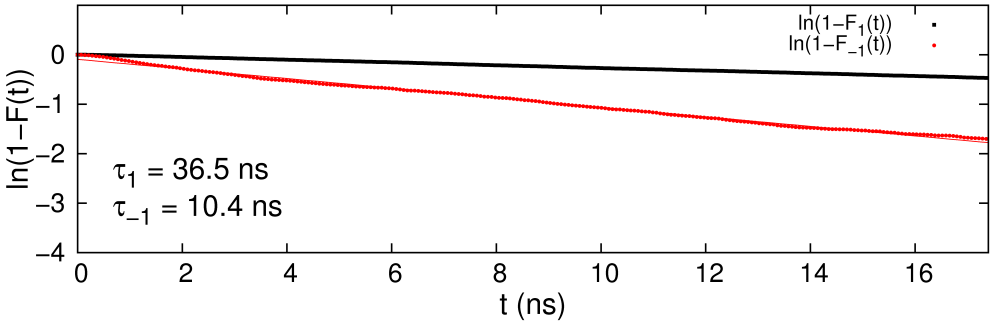
ln(1−F(t))
t (ns)
ln(1−F1(t))
ln(1−F−1(t))
τ1 = 36.5 ns
τ−1 = 10.4 ns
0
−1
−2
−3
−4
0
2
4
6
8
10
12
14
16

FIG. 5:

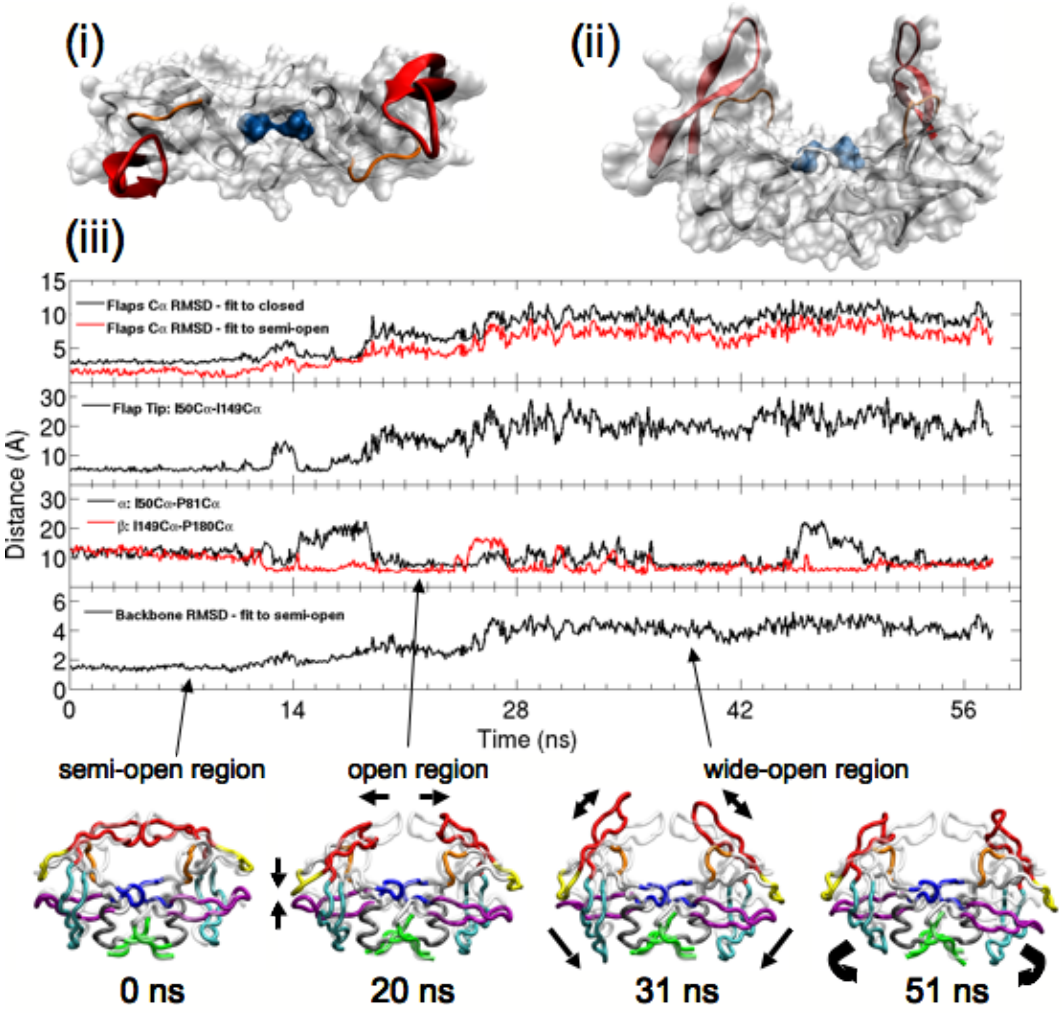
(i)
(ii)
(iii)
Flaps Cα RMSD - fit to closed
Flaps Cα RMSD - fit to semi-open
Flap Tip: I50Cα-I149Cα
α: I50Cα-P81Cα
β: I149Cα-P180Cα
Backbone RMSD - fit to semi-open
Distance (A)
Time (ns)
semi-open region
open region
wide-open region
0 ns
20 ns
31 ns
51 ns

FIG. 6: