Substrate Recognition in HIV-1 Protease: A Computational Study

M. A. S. Perez, P. A. Fernandes, and M. J. Ramos*

*REQUIMTE, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre,
687, 4169-007 Porto, Portugal*

Received: November 18, 2009; Revised Manuscript Received: January 9, 2010

HIV-1 protease is a crucial enzyme for the life cycle of the human immunodeficiency virus, the retrovirus that triggers AIDS. It is well documented that HIV-1 protease mediates the cleavage of Gag, Gag-Pol, and Nef precursor polyproteins and is highly selective concerning the set of 12 different amino acid sequences that cleaves. However, the governing principles and physical parameters, which determine substrate recognition and specificity, remain poorly understood despite the many speculative proposals that abound in the literature. In fact, it has been difficult so far to circumvent the fact that protease's substrates share little sequence identity and lack an obvious consensus binding motif. We have used microsecond time scale MD simulations to quantitatively show that some sequences of the polyprotein Gag-Pol that are not cleaved (nonsubstrates) have in fact a higher affinity to the active site of HIV-1 protease than a substrate; i.e., recognition is not governed by affinity to the active site. On the basis of a detailed analysis of the results and experimental data, we propose that the recognition is based on the geometric specificity of PR:Gag and PR:Gag-Pol multiprotein complex, that selects which residues lie in the specific position that makes them accessible to the active site for cleavage.

Introduction

HIV-1 protease (PR), a retroviral protease member of the aspartic proteinase family of enzymes,[1] is an extremely important enzyme with some of its inhibitors having been established as some of the first drugs developed for the treatment of AIDS. Years later, this enzyme continues to be the fundamental pillar of its therapy.

The enzyme is active as a homodimer, each unit contributing with one catalytic aspartate to the active site.[2–6] PR's flaps that lie above the active site as well as the "fireman's grip", the conserved complex scaffold of hydrogen bonds supporting the active site, are important features[7] that can be observed in Figure 1.

PR is a small enzyme with 99 amino acids per monomer that mediates the cleavage of Gag, Gag-Pol, and Nef precursor polyproteins. These reactions occur late in the viral life cycle, during virion assembly and maturation at the cell surface. The process is highly specific, temporally regulated, and essential for the production of infectious viral particles.[8–11] The main structural proteins are formed by cleavage of the PR55^{Gag} polyprotein, and the viral enzymes are formed by cleavage of PR160$^{Gag\text{-}Pol}$.[8] The PR embedded within the Gag-Pol polyprotein cleaves itself out by specifically cutting peptide bonds at either end of its sequence. Subsequently, it cleaves additional bonds within the remaining fragment of the Gag-Pol polyprotein. In total, 12 proteolytic reactions are required to generate a mature infectious virion.[9] Figure 2 shows the 12 individual PR cleavage sites.

Each reaction occurs sequentially and with high fidelity at seemingly unrelated cleavage sites. The 12 recognition sequences that PR cleaves specifically are shown in Table 1. The substrate used as a reference in the current study has been highlighted, and the cleavage sites are denoted by an asterisk.

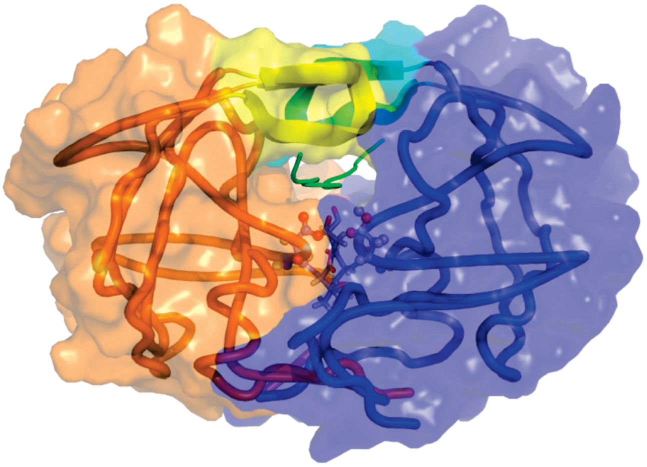

Figure 1. Complex protease:substrate ARVLAEAM. Monomer A is orange and monomer B blue. The flaps of both monomers have been colored yellow and cyan, respectively. Fireman's grip can be seen in sticks, with both catalytic aspartates in ball and stick. Catalytic aspartate from monomer B is shown protonated. In the active center, the eight residues of the substrate that fit in only have been simply displayed in green tube.

Centered in the hydrophobic active site are the two symmetrically disposed aspartate residues (Asp 25 and Asp 25′) involved in the hydrolysis of the peptide bond (Figure 1). Studies have shown that the hydrophobic cavity can hold eight amino acids of substrate bound in an extended β-sheet conformation along this cleft (hydrophobic, charged), through extensive hydrogen bond and van der Waals interactions.[12] It is also understood that PR undergoes substantial conformational changes as the cleft of the active site tightens around an incoming substrate. However, the governing principles and physical parameters, which determine substrate recognition and specificity, remain poorly understood despite the many speculative proposals that progressively have been put forward in the

* To whom correspondence should be addressed. E-mail: mjramos@fc.up.pt.

10.1021/jp910958u
Published on Web 02/01/2010

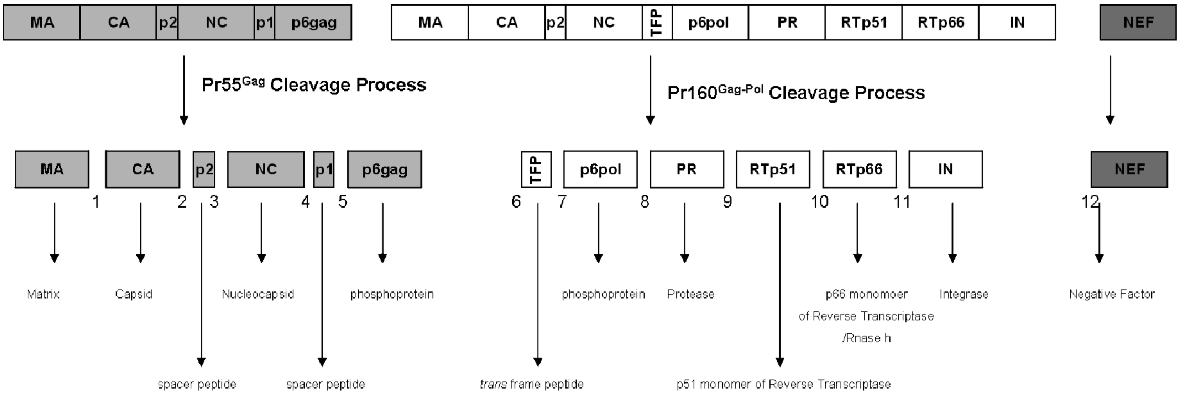

Figure 2. Schematic representation of the Gag (light gray) and Gag-Pol (white) processing sites showing the 12 individual protease cleavage sites. MA, matrix; CA, capsid; NC, nucleocapsid; p2 and p1, two spacer peptides; TFP, *trans* frame peptide; p6gag, an important phosphoprotein required for virion release; p6pol, protein implicated in the regulation of Protease auto-activation; PR, protease; RTp51, p51 monomer of reverse transcriptase; RTp66, p66 monomer of reverse transcriptase/rnase h; IN, integrase; NEF, negative factor, precursor polyprotein also known as F protein.

TABLE 1: The 12 Recognition Sequences Cleaved by HIV-1 Protease, HXB2, Group M, Subtype B (HIV-1 is Divided into Three Groups, (M (the Most Common), N, and O), and Each Group Is Divided into Subtypes and CRFs (Circulating Recombinant Forms)[48])

substrate sequences	cleavage domain
SQNY*PIVQ	MA-CA
ARVL*AEAM	**CA-p2**
ATIM*MQRG	p2-NC
RQAN*FLGK	NC-p1
RQAN*FLRE	NC-TFP
PGNF*LQSR	p1-p6gag
DLAF*LQGK	TFP-p6pol
SFNF*PQVT	p6pol-PR
TLNF*PISP	PR-RTp51
AETF*YVDG	RTp51-RTp66
RKVL*FLDG	RTp66-INT
DCAW*LEAQ	NEF

literature. In fact, it has been difficult so far to circumvent the fact that protease's substrates share little sequence identity and lack an obvious morphologic and electrostatic consensus binding motif. The current approach traditionally invoked to explain multiple substrate binding in HIV-1 PR defends that the specificity of PR is determined by the ability of substrates' amino acid side chains to bind into eight individual subsites within the enzyme.[13]

In this work, we quantitatively show that several sequences of the polyprotein Gag-Pol that are not cleaved (nonsubstrates) have a higher affinity to the active site of HIV-1 PR than its substrates. This fact is enough to disclaim the above-mentioned traditional active site substrate recognition approach. A detailed analysis of the existing experimental data as well as the results presented here shows that the models presently used to explain substrate recognition are inappropriate, and we put forward a new consistent hypothesis in which the geometry of protein–protein PR:Pr55$^{\text{Gag}}$ and PR:Pr160$^{\text{Gag-Pol}}$ complexation is the basis for substrate recognition by HIV-1 PR.

Methods

To prove the existence of such high PR affinity from nonsubstrates, we have calculated the relative free energy difference in the binding of a substrate as compared to some nonsubstrates, $\Delta\Delta G_{\text{bind}}^{\text{S}\rightarrow\text{nS}}$. Values for $\Delta\Delta G_{\text{bind}}^{\text{S}\rightarrow\text{nS}}$ have been calculated with the very accurate thermodynamic integration (TI) method[14–16] as well as the faster, albeit still accurate, molecular mechanics Poisson–Boltzmann surface area (MMPBSA)[17] approach. Such calculated values represent the affinity of the nonsubstrates, relative to the substrate, for the PR active site. MMPBSA has several appealing features as compared to TI, in particular the much inferior computational time that is required for the calculations, which allows for the exploration of the binding properties of more nonsubstrates than otherwise possible. Therefore, we have determined the value of $\Delta\Delta G_{\text{bind}}^{\text{S}\rightarrow\text{nS}}$ for one nonsubstrate with TI and for the other nonsubstrates with MMPBSA, calibrated as to reproduce the TI values, in the fashion that we hasten to explain.

Structures. The crystal structure of HIV-1 PR in complex with the substrate was obtained in the database Protein Data Bank[18] with reference 1F7A.[12] The structure is presented at 2.0 Å resolution, a complex between an inactive variant of HIV-1 protease (D25N), with the catalytic aspartic acids mutated by asparagines, and a long substrate peptide, Lys-Ala-Arg-Val-Leu-Ala-Glu-Ala-Met-Ser. To begin with, we have mutated the asparagines back to the necessary catalytic aspartic acids (residues 25 and 25′, each one of chains A and B, respectively). It is known that one of the two aspartate residues in protease is protonated and therefore we have added one hydrogen atom to aspartate 25 of chain B, Asp 25′, in agreement with the most widely accepted catalytic scheme considered for the cleavage of the polyprotein by the protease of HIV-1[19] and with our own findings in a previous work.[20] It was verified that in protease the side chains of His residues, frequently found in a nonstandard protonated state (His pK_a = 6.1), were exposed to the solvent and therefore considered to be in the neutral state. Standard protonation states were assumed for all other residues. We have worked with eight amino acids of the substrate, those that fit inside the binding pocket, and excluded the last and the first amino acids.

Our nonsubstrates are sequences of the polyprotein PR160$^{\text{Gag-Pol}}$, not cleaved by PR. The 3D structure of the polyprotein Gag-Pol is unknown, and we have extracted the amino acid sequence from SWISS-PROT,[21] P04585. In Figure 3, substrates are underlined and nonsubstrates used in the current work are shown in bold.

Nonsubstrate TKALTEVI was selected because it has more common properties with substrate ARVLAEAM, and nonsubstrates ARASVLSG and SQVTNSAT were chosen randomly. The three-dimensional structures of complexes PR:TKALTEVI, PR:SQVTNSAT, and PR:ARASVLSG were built using the complex PR:ARVLAEAM as a reference. We wanted to fit the nonsubstrates into the protease with a maximum number of interactions. To achieve this, we mutated each amino acid side

G**ARASVLSG**GELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQIL
GQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAAAD
TGHSNQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATP
QDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTS
TLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRFY
KTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPAATLEEMMTACQGVGGPGHKAR
VLAEAM**SQVTNSAT**IMMQRGNFRNQRKIVKCFNCGKEGHTARNCRAPRKKGCWKCGKEGH
QMKDCTERQANFLREDLAFLQGKAREFSSEQTRANSPTRRELQVWGRDNNSPSEAGADRQ
GTVSFNFPQVTLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIGGF
IKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPISPIETVPVKLKP
GMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPVFAIKKKDSTKWRKL
VDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEDFRKYTAFTIPS
INNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDIVIYQYMDDLYVGSDLE
IGQHRTKIEELRQHLLRWGLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWTV
NDIQKLVGKLNWASQIYPGIKVRQLCKLLRG**TKALTEVI**PLTEEAELELAENREILKEPV
HGVYYDPSKDLIAEIQKQGQGQWTYQIYQEPFKNLKTGKYARMRGAHTNDVKQLTEAVQK
ITTESIVIWGKTPKFKLPIQKETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPI
VGAETFYVDGAANRETKLGKAGYVTNRGRQKVVTLTDTTNQKTELQAIYLALQDSGLEVN
IVTDSQYALGIIQAQPDQSESELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSA
GIRKVLFLDGIDKAQDEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKGEAMHGQV
DCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKTIH
TDNGSNFTGATVRAACWWAGIKQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKT
AVQMAVFIHNFKRKGGIGGYSAGERIVDIIATDIQTKELQKQITKIQNFRVYYRDSRNPL
WKGPAKLLWKGEGAVVIQDNSDIKVVPRRKAKIIRDYGKQMAGDDCVASRQDED

Figure 3. Complex UniProtKB/Swiss-Prot P04585, Gag-Pol polyprotein (HIV-1 isolate HXB2 group M subtype B). Substrates are underlined, and nonsubstrates are shown in bold.

chain of the substrate into the correspondent amino acid side chain of the nonsubstrate using a rotamer library[22] for the initial conformational guess and choosing the rotamer that leads to the greater number of interactions.

Thermodynamic Integration. An important goal of computational chemistry is the accurate prediction of free energies in molecular systems. It provides a direct link between the microscopic structure and fluctuations of a system and its most fundamental thermodynamic property, the Gibbs energy. Estimation of the difference in binding free energies for two ligands, substrate (S) ARVLAEAM and nonsubstrate (nS) TKALTEVI, to HIV-1 PR is our case study.

The most accurate methods to compute free energy are thermodynamic integration, (TI) and free energy perturbation, (FEP).[14,23–28] Much has been said in the literature about FEP versus TI, but the difference mainly pertains to the formula used for evaluating the free energy. In this work, we have used TI, but FEP and TI are comparably efficient.[29]

The free energy difference between two states A and B can formally be obtained from Zwanzig's formula:[30]

$$\Delta G_{\mathrm{A}\rightarrow\mathrm{B}} = G_{\mathrm{B}} - G_{\mathrm{A}} = -\beta^{-1} \ln\langle \exp - \beta(V_{\mathrm{B}} - V_{\mathrm{A}})\rangle_{\mathrm{A}} \tag{1}$$

where $1/kT = \beta$ and $\langle\,\rangle_{\mathrm{A}}$ denote an MD or MC generated ensemble average that is sampled using the V_{A} potential. Equation 1 assumes that the configurational sampling is carried out under constant temperature and pressure conditions (isothermal−isobaric ensemble). A kinetic contribution to the free energy difference, e.g., due to a possible change in atomic masses, is not considered, since it will always cancel out by virtue of the equipartioning theorem and the relevant thermodynamic cycle.

Equation 2, usually referred to as the TI formula for free energy, is in fact exact, just as is eq 1, and can be derived directly from the configuration integral

$$\Delta G_{\mathrm{A}\rightarrow\mathrm{B}} = \int_0^1 \left(\frac{\partial V(\lambda)}{\partial \lambda}\right)_{\lambda} \mathrm{d}\lambda \tag{2}$$

The difference in Gibbs free energy between the substrate model, ARVLAEAM, and the nonsubstrate model, TKALTEVI, was evaluated using the thermodynamic cycle shown in Figure 4.

$$\begin{tikzcd}
\text{ARVLAEAM}_{(aq)} + \text{Protease}_{(aq)} \arrow[r, "\Delta G^{S}_{bind}"] \arrow[d, "\Delta G^{S\rightarrow nS}_{solv}"'] & \text{Protease:ARVLAEAM}_{(aq)} \arrow[d, "\Delta G^{S\rightarrow nS}_{P}"] \\
\text{TKALTEVI}_{(aq)} + \text{Protease}_{(aq)} \arrow[r, "\Delta G^{nS}_{bind}"'] & \text{Protease:TKALTEVI}_{(aq)}
\end{tikzcd}$$

Figure 4. Thermodynamic cycle for calculating the free energy difference between the substrate ARVLAEAM and nonsubstrate TKALTEVI, bound to protease. $\Delta G^{\mathrm{S}}_{\mathrm{bind}}$ and $\Delta G^{\mathrm{nS}}_{\mathrm{bind}}$ are binding free energies for substrate and nonsubstrate, respectively, and $\Delta G^{\mathrm{S}\rightarrow\mathrm{nS}}_{\mathrm{solv}}$ and $\Delta G^{\mathrm{S}\rightarrow\mathrm{nS}}_{\mathrm{P}}$ are the nonphysical transmutation free energies from substrate to nonsubstrate in solution and bound to protease, respectively.

From the thermodynamic cycle, we get eq 3:

$$\Delta\Delta G^{\mathrm{S}\rightarrow\mathrm{nS}}_{\mathrm{bind}} = \Delta G^{\mathrm{nS}}_{\mathrm{bind}} - \Delta G^{\mathrm{S}}_{\mathrm{bind}} = \Delta G^{\mathrm{S}\rightarrow\mathrm{nS}}_{\mathrm{P}} - \Delta G^{\mathrm{S}\rightarrow\mathrm{nS}}_{\mathrm{solv}} \tag{3}$$

We need to calculate $\Delta G^{\mathrm{S}\rightarrow\mathrm{nS}}_{\mathrm{P}}$ and $\Delta G^{\mathrm{S}\rightarrow\mathrm{nS}}_{\mathrm{solv}}$. Both quantities have no meaning in the real world, as they refer to free energy differences between different molecules.

The necessary transformation to compute $\Delta G^{\mathrm{S}\rightarrow\mathrm{nS}}_{\mathrm{P}}$ and $\Delta G^{\mathrm{S}\rightarrow\mathrm{nS}}_{\mathrm{solv}}$ represents a severe challenge for explicit free energy calculations, as the mutations from the substrate to a nonsubstrate involve creating and deleting a very large number of atoms, the highest number ever tried as far as we know.

There is no direct correspondence between the number of atoms in the substrate and in the nonsubstrate, so "dummy atoms"[24] were used. A dummy atom is an atom for which the nonbonded interactions with all other atoms are zero. Dummy atoms have masses, but masses only affect kinetic properties; they do not enter in thermodynamics and therefore have no influence in the free energy.

A well-known problem in free energy calculations is the so-called "end-point catastrophe" associated with vanishing or created atoms.[15] This is particularly pertinent for our case. The problem is basically twofold. First, there is a possible numerical instability of the free energy formulas associated primarily with the infinite repulsive ($1/r^{12}$) Lennard-Jones term (LJ) for $r = 0$. That is, in the state in which a certain atom has no interactions, other atoms can lie on top of it and therefore the energy of the state in which this atom becomes suddenly present with its nonzero interaction would become infinite, as well as its derivative. The second problem is perhaps more serious and has to do with a deficient configurational sampling introduced by appearing/disappearing atoms. It can be understood from the fact that the repulsive Lennard-Jones potential of a very small atom, i.e., one whose interactions are scaled by a very small value of λ, is still infinite for $r = 0$. This means that sampling of the positions occupied by vanishing atoms cannot be accomplished until these atoms have completely disappeared, i.e., at the end-point of the λ-range. This is particularly problematic in confined geometries, such as a protein binding site, where it is possible that the space occupied by vanishing atoms cannot be properly filled until the very last end-point simulation.

It is clear that the use of the soft-core nonbonded potentials, the use of a denser distribution of λ points near problematic end-points, and simultaneously decoupling the changes of the LJ and Coulombic interactions can improve sampling considerably.[16]

We have considered four separate energy transformations, a first one in which the charges on atoms that will disappear are removed and on atoms that will be converted are changed (ΔG^1), a second one in which the LJ interactions between the atoms that will disappear are morphed to zero and between the atoms that will be converted are changed (ΔG^2), a third one in which LJ interactions on the nascent atoms are created (ΔG^3), and a fourth and last one in which the charges of the nascent atoms are turned on (ΔG^4). We note that it is impossible to directly assign a Coulombic or vdW partitioning of the total free energy, as these are both path dependent.[31]

$$\Delta\Delta G_{\mathrm{bind}}^{\mathrm{S}\rightarrow\mathrm{nS}} = \Delta G_{\mathrm{P}}^{\mathrm{S}\rightarrow\mathrm{nS}} - \Delta G_{\mathrm{solv}}^{\mathrm{S}\rightarrow\mathrm{nS}} = (\Delta G_{\mathrm{P}}^{1} + \Delta G_{\mathrm{P}}^{2} + \Delta G_{\mathrm{P}}^{3} + \Delta G_{\mathrm{P}}^{4}) - (\Delta G_{\mathrm{solv}}^{1} + \Delta G_{\mathrm{solv}}^{2} + \Delta G_{\mathrm{solv}}^{3} + \Delta G_{\mathrm{solv}}^{4}) \quad (4)$$

We have computed the free energy of each transformation with the GROMACS package version 3.3.[24]

For each lambda value running in each transformation, we have carried out a minimization with the l-bfgs algorithm, a minimization with the leapfrog stochastic dynamics integrator, a constant volume equilibration run, a constant pressure equilibration run, and then a constant pressure production run. Free energy derivatives were collected independently for each lambda from the production run.

These simulations were performed in explicit solvent and under periodic boundary conditions[24] with 16 λ values (0.00, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.00). We have used λ increments of 0.05 when the free energy derivative was changing quickly and increments of 0.10 when the free energy derivative was changing slowly.

During the λ path, *soft-core* potentials were used for nonbonded interactions. Differences in constraints or in long-range interactions are properly handled. Long-range dispersion corrections for energy and pressure were applied. With the long-range vdW corrections turned on, the energies and pressures will not be influenced by the vdW cutoff. We have used a 9 Å cutoff for the Coulombic interactions, within a particle-mesh Ewald methodology, and an 8−9 Å switched cutoff for the LJ interactions.[24]

The substrate in free and bound states was centered in a cubic box with 10 and 12 Å, respectively, from the molecule periphery. The Cornell force field[32] was used to describe both the protein and the ligand. The TIP3P water model[33] was used, which is rigid and includes no Lennard-Jones term on the hydrogens. Integration of the equations of motion was perfomed using the velocity Verlet algorithm.[34] The time step used for integrating the equations of motion was 0.001 ps. A total of 100 ps was simulated for constant volume equilibration and constant pressure equilibration, and a total of 3 ns for production near the end points (λ = 0.00, 0.05, 0.10, 0.85, 0.90, 0.95, 1.00) and 2 ns for the other λ values. The ensemble average of $\partial V/\partial\lambda$ for each lambda point was calculated over the production. The total simulation time to mutate the substrate into the nonsubstrate was 168.8 ns, and we have performed this in both free and bound states as well as in the direct and reverse directions, making up a total time that approaches the microsecond time scale (675.2 ns). Berendsen pressure control was used to produce controlled pressure simulations. To maintain the temperature at 300 K, we have used the Langevin[35] dynamics. The hydrogen bonds were constrained using the LINCS algorithm.[24]

We have calculated the hysteresis in the free energy performing the mutations forward and backward. This is not actually a valid uncertainty estimate (it is related to systematic error[16]), but it is frequently smaller than the actual statistical uncertainty of the measurement.

Free energy is highly dependent on the protocols. Such simulations require careful attention to the computational setup and to the interpretation of results and should not be performed routinely. We have rigorously tested the methodology used, calculating $\Delta\Delta G$ for some mutations for which experimental data are available. Test simulations were conducted with our system to examine that as the sampling time is increased the difference in free energy forward and backward decreases. Moreover, the free energy profile forward and backward of such simulations has indicated that most of the difference is due to the discrepancy at λ values near the end points. Thus, near the end points, the derivative $\partial V/\partial\lambda$ was averaged over 3 ns instead of 2 ns.

Molecular Mechanics/Poisson−Boltzmann/Surface Area. The MMPBSA[17] script implemented in Amber8[36] was used to calculate the binding free energies for all complexes.

This approach is based on an analysis of molecular dynamics trajectories using a continuum solvent approach and resulting in the "average" free energy of a state given as

$$\langle G\rangle = \langle E_{\mathrm{MM}}\rangle + \langle G_{\mathrm{PBSA}}\rangle - T\langle S_{\mathrm{MM}}\rangle \quad (5)$$

in which $\langle E_{\mathrm{MM}}\rangle$ is an average molecular mechanical energy that typically includes bond, angle, torsion, vdW, and electrostatic terms from a regular force field, evaluated with no cutoff. Solvation free energies are calculated using a numerical solution of the Poisson−Boltzmann equation and, together with a surface area based estimate of the nonpolar free energy, constitute the $\langle G_{\mathrm{PBSA}}\rangle$ term. Both $\langle E_{\mathrm{MM}}\rangle$ and $\langle G_{\mathrm{PBSA}}\rangle$ were obtained by averaging over a sample of representative geometries extracted from an MD trajectory of a system. The last term, $-T\langle S_{\mathrm{MM}}\rangle$, is the solute entropy, which is usually considered as being equivalent in gas phase and solution, getting canceled (and therefore not determined) in the calculation of the solvation free energy.

For the MMPBSA calculations, we have made minimizations and molecular dynamics simulations with the Amber software package.[36]

During the calculations, the solvent was modeled with a modified generalized Born solvation model.[37] To release the bad contacts in the crystallographic structures, the complexes were minimized in three stages: in the first stage, only the hydrogen atoms added with xLeap were minimized, in the second stage, the backbone was also minimized, and in the third and last stage, the entire system was minimized. About 1500 steps were used for each stage, with the first 500 steps performed using the steepest descent algorithm and the remaining steps carried out using a conjugate gradient.

Subsequently, we performed 40 ns molecular dynamics (MD) runs starting from each of the minimized structures. All simulations presented in this work were carried out using the sander module, implemented in the Amber8[36] simulations package, with the *Cornell* force field.[32,36] Bond lengths involving hydrogens were constrained using the SHAKE algorithm,[38] and the equations of motion were integrated with a 2 fs time step using the Verlet leapfrog algorithm and the nonbonded interactions truncated with a 16 Å cutoff. The temperature of the system was regulated by the Langevin thermostat to maintain the temperature of our system at 300 K.[35,39]

The MMPBSA script[40] was used to perform a postprocessing treatment of all complexes by using its structure, and calculating

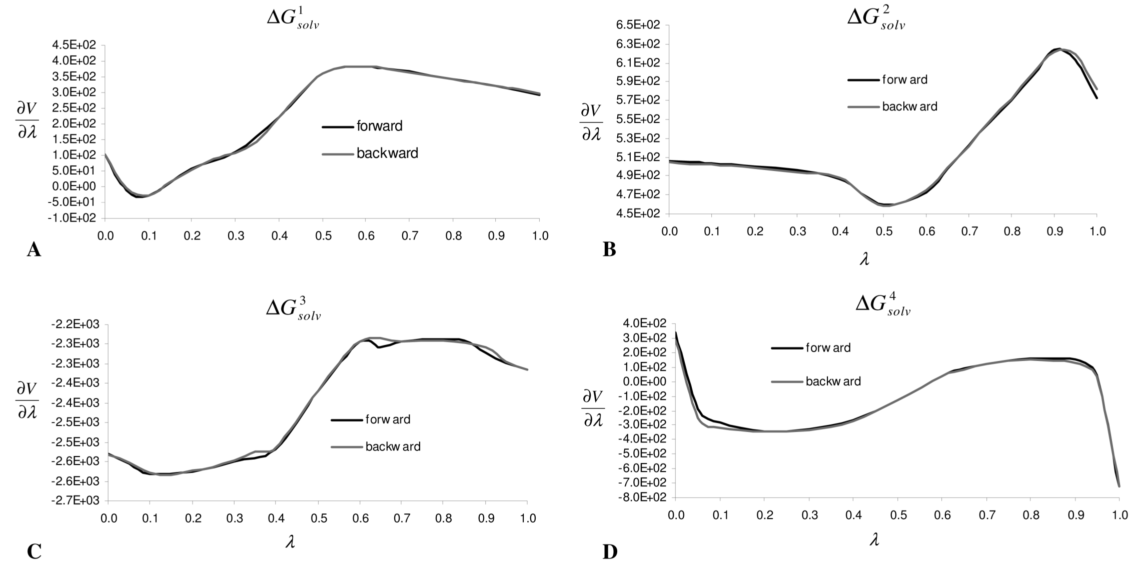

Figure 5. Derivative $\partial V(\lambda)/\partial\lambda$ (kcal·mol^{-1}) as a function of λ for each transformation free in solution in direct and reverse directions. $\Delta G^{1}_{\text{solv}}$, $\Delta G^{2}_{\text{solv}}$, $\Delta G^{3}_{\text{solv}}$, and $\Delta G^{4}_{\text{solv}}$ are presented in parts A, B, C, and D, respectively.

the respective energies for the complex and all interacting monomers. For the binding free energy calculations, 1000 snapshots of the complexes were extracted every 100 steps for the last 100 000 steps (i.e., 20 ns) of the molecular dynamics simulations.

$$\Delta G^{\text{nS}}_{\text{bind}} = \langle G_{\text{complex}}\rangle - \langle G_{\text{protease}}\rangle - \langle G_{\text{nonsubstrate}}\rangle \quad (6)$$

$$\Delta G^{\text{S}}_{\text{bind}} = \langle G_{\text{complex}}\rangle - \langle G_{\text{protease}}\rangle - \langle G_{\text{substrate}}\rangle \quad (7)$$

The binding free energies differences between the nonsubstrates and substrate complexes are defined as:

$$\Delta\Delta G^{S\rightarrow nS}_{bind} = \Delta G^{nS}_{bind} - \Delta G^{S}_{bind} \quad (8)$$

The internal energy (bond, angle, and dihedral), electrostatic, and van der Waals interactions were calculated using the *Cornell* force field[24] with no cutoff. The electrostatic solvation free energy was calculated by solving the Poisson−Boltzmann equation with the software Delphi v.4.[41,42] The accuracy of this method depends on the self-consistency of the model parameters used to solve the finite difference method implemented in DelPhi. The key parameters used were based on a detailed study of the effect of their variation of key parameters in several systems as well as the correspondent computational time.[43] The nonpolar contribution to solvation free energy due to van der Waals interactions between the solute and the solvent and cavity formation was modeled as a term that is dependent on the solvent accessible surface area of the molecule. It was estimated using an empirical relation

$$\Delta G_{\text{nonpolar}} = \alpha A + \beta \quad (9)$$

in which A is the solvent-accessible surface area that was estimated using the molsurf program, that is based on the idea primarily developed by Mike Connolly.[44] α and β are empirical constants, and the values used were 0.00542 kcal Å^{-2} mol^{-1} and 0.92 kcal mol^{-1}, respectively. The entropy term was not calculated because it was assumed that its contribution to $\Delta\Delta G^{\text{S}\rightarrow\text{nS}}_{\text{bind}}$ is mainly canceled and becomes negligible.[40]

While the choice of the external dielectric constant depends on the solvent media, the choice of the internal dielectric constant has been the subject of discussion and controversy because the dielectric constant is not a universal constant but simply a parameter that depends on the model and the methodology used.[40] In this work, we could reproduce computationally the binding free energy obtained with thermodynamic integration,[14−16] emphasizing the accuracy and reliability of the MMPBSA protocol under these circumstances. The value of the external dielectric constant used was 80.0. The value of the internal dielectric constant was set to 9.00.

Results

The TI method was applied to calculate the relative free energy of binding between complexes protease:ARVLAEAM and protease:TKALTEVI. The TI calculation required the virtual transformation of the substrate into the nonsubstrate, which involved mutating a very large number of amino acids. Consequently, problems with sampling and convergence were of great concern and large simulation times were needed. To the best of our knowledge, the TI calculation presented here represents the largest mutation that has ever been attempted so far. It involved a total simulation time close to a microsecond (675 ns of molecular dynamics (MD) simulations). The free energy profiles of the separate transformations considered, $\Delta G^{1}_{\text{solv}}$, $\Delta G^{2}_{\text{solv}}$, $\Delta G^{3}_{\text{solv}}$, $\Delta G^{4}_{\text{solv}}$, ΔG^{1}_{P}, ΔG^{2}_{P}, ΔG^{3}_{P}, and ΔG^{4}_{P}, are reported in Figures 5 and 6. To check the dependency of the results on the direction of the mutation (hysteresis), the free energy profiles of the reverse mutations are also presented.

Even though the MD simulations ran for hundreds of nanoseconds, which is more than enough to explore the slow rotameric space, we must emphasize that, if conformational sampling was still incomplete, it would benefit substrate binding rather than nonsubstrate binding, as the starting structure was an X-ray structure of the enzyme−substrate complex. This means that eventual conformational sampling limitations and hypothetical long time scale induced-fit active site backbone rearrangement, in the presence of nonsubstrates, would make the affinity for the nonsubstrates increase even further. The free

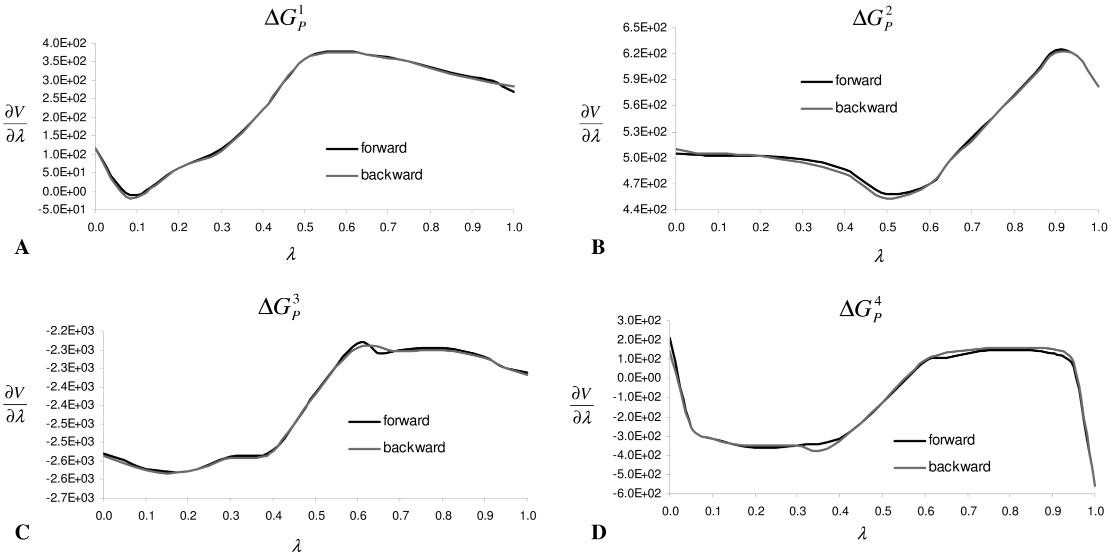

Figure 6. Derivative $\partial V(\lambda)/\partial\lambda$ (kcal·mol^{-1}) as a function of λ for each transformation while bound to the protein in direct and reverse directions. ΔG_P^1, ΔG_P^2, ΔG_P^3, and ΔG_P^4 are presented in parts A, B, C, and D, respectively.

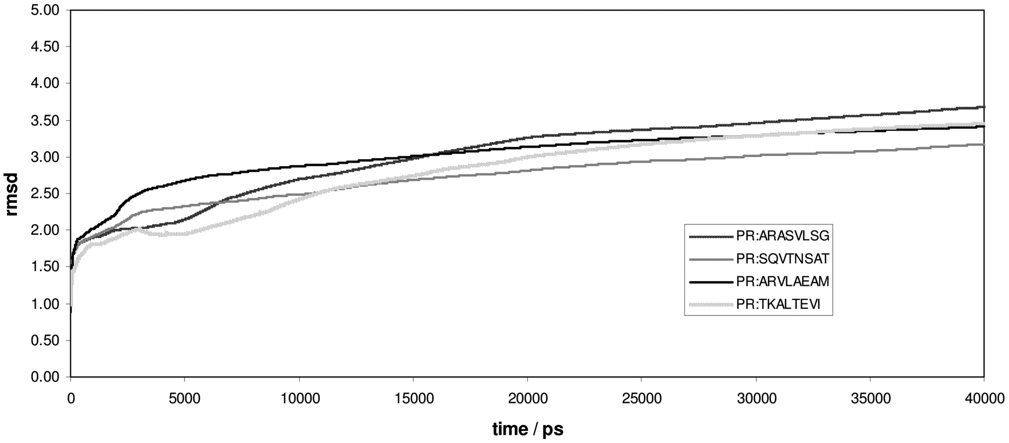

Figure 7. rmsd of the trajectory's structures as compared to the starting structures for complexes PR:ARVLAEAM, PR:TKALTEVI, PR:SQVTNSAT, and PR:ARASVLSG.

energy for mutating ARVLAEAM into TKALTEVI free in solution, $\Delta G_{\mathrm{solv}}^{\mathrm{S}\rightarrow\mathrm{nS}}$, is −1754.0 kcal/mol, and while bound to the protein $\Delta G_{\mathrm{P}}^{\mathrm{S}\rightarrow\mathrm{nS}}$ is −1758.9 kcal/mol. These values have no meaning by themselves, but the difference between them, −4.9 kcal/mol, is the difference in the free energy of binding between the nonsubstrate and the substrate. The free energy for the same mutation but calculated in the reverse direction free in solution, $\Delta G_{\mathrm{solv}}^{\mathrm{S}\rightarrow\mathrm{nS}}$, is −1763.9 kcal/mol; while bound to the protein, $\Delta G_{\mathrm{P}}^{\mathrm{S}\rightarrow\mathrm{nS}}$ is −1767.7 kcal/mol and thus $\Delta\Delta G_{\mathrm{bind}}^{\mathrm{S}\rightarrow\mathrm{nS}}$ is −3.8 kcal/mol.

We have achieved an excellent level of convergence considering the magnitude of the mutations involved. The energy profiles forward and backward indicate that, with an overall microsecond time scale simulation, we can obtain accurate estimates of the binding free energy. No experimental data on the binding affinities are available, and there are some problems such as the propagation error through the phases (random errors) that were not treated in this paper. However, our results confidently predict that nonsubstrate TKALTEVI has a higher affinity to the protease than substrate ARVLAEAM even though PR never cleaves the former but always cleaves the latter. Therefore, the specific selectivity for the substrate cannot be based on the affinity for the active site.

To evaluate the binding affinities of the other nonsubstrates, we have used MMPBSA after calibration by reproducing exactly the average obtained with TI for the nonsubstrate TKALTEVI. With TI, the average value obtained for $\Delta\Delta G_{\mathrm{bind}}^{\mathrm{S}\rightarrow\mathrm{nS}}$ was −4.4 kcal/mol; with MMPBSA, we set the internal dielectric constant to 9.00 and obtained −4.4 kcal/mol.

To assess the quality of the simulations, we monitored the root-mean-square deviations of the trajectory structures as compared to the starting structures for the complexes, for the substrate and the nonsubstrates, for each monomer of protease, and for the backbone. The rmsd's for complexes PR:ARVLAEAM, PR:TKALTEVI, PR:ARASVLSG, and PR:SQVTNSAT are shown in Figure 7.

It can be seen that, at the last 20 ns of the 40 ns long MMPBSA simulations, every system has achieved equilibrium and remains stable. Table 2 summarizes the values of the free energy differences for the nonsubstrates relative to the substrate ARVLAEAM.

The obtained results reveal that two (out of three) nonsubstrates (TKALTEVI, SQVTNSAT) of the polyprotein Gag-Pol that are not cleaved have greater affinity to the protease active site than the used substrate. Nonsubstrate ARASVLSG has to be considered a poor ligand, as shown by the positive value obtained for $\Delta\Delta G_{\mathrm{bind}}^{\mathrm{S}\rightarrow\mathrm{nS}}$.

Protease is highly specific, cleaving only a very specific set of amino acid sequences; on the other hand, substrates share little sequence identity and morphologic/electrostatic similarity,

TABLE 2: Values of the Free Energy Differences between the Nonsubstrates and the Substrate ARVLAEAM and Their Uncertainties[a]

nonsubstrate	$\Delta\Delta E^{\mathrm{ele}}$	$\Delta\Delta E^{\mathrm{vdw}}$	$\Delta\Delta G^{\mathrm{nonpol}}$	$\Delta\Delta G^{\mathrm{solv}}$	$\Delta\Delta G_{\mathrm{bind}}^{\mathrm{S}\to\mathrm{nS}}$	$\delta\Delta\Delta G_{\mathrm{bind}}^{\mathrm{S}\to\mathrm{nS}}$
TKALTEVI	−14.8	5.8	−0.2	4.5	−4.4	1.1
ARASVLSG	3.0	8.3	0.0	−7.6	3.6	1.1
SQVTNSAT	2.2	−7.2	−0.2	0.3	−4.7	1.1

[a] Also shown are the individual contributions for$\Delta\Delta G_{\mathrm{bind}}^{\mathrm{S}\to\mathrm{nS}}$, i.e., the differences in the electrostatic $\Delta\Delta E^{\mathrm{ele}}$ and van der Waals $\Delta\Delta E^{\mathrm{vdw}}$ energies of interaction of the solute, as well as the differences in the nonpolar $\Delta\Delta G^{\mathrm{nonpol}}$ and the sum of polar and nonpolar contributions to solvation $\Delta\Delta G^{\mathrm{solv}}$ free energies of solvation. All values are in kcal/mol.

and we have shown that some sequences of the polyprotein PR160$^{\text{Gag-Pol}}$ that are not cleaved (nonsubstrates) have higher affinity to the HIV-1 PR than a substrate. This means that the specificity of HIV-1 PR is not exclusively determined by the ability of the substrate to bind to the enzyme; i.e., recognition is not active-site-based.

Discussion

The proposal that the specificity of PR is determined by the ability of substrates' amino acid side chains to bind into eight individual subsites within the enzyme[13] is at the very least incomplete. Protease can bind a large variety of peptides, and some studies have shown that (I) a variety of amino acid residues can be accommodated in each of the enzyme subsites;[45] (II) individual enzyme subsites are capable of acting independently in the recognition of amino acids in the corresponding substrate position;[13] (III) in the natural substrates, amino acids with different properties and side chain sizes can occupy the same position in the PR binding pocket.[45] There are no proposals that can explain PR specificity based only on the interactions of the substrates inside the binding pocket, because several nonsubstrates can establish stronger interactions.[13]

We have no doubt that the polyprotein conformation is extremely important in substrate recognition. Experimental data superficially explores this fact and shows that cleavage by HIV-1 PR is dependent on the polyprotein folding.[46] In fact, a hidden site whose cleavage by HIV-1 PR is dependent on the polyprotein folding has been identified. This site was detected only in mutants destabilized by peptide insertion. Insertion of amino acids has produced a new stable conformation combining activity and exposure that allow sufficient interaction with the PR.[46]

The above-mentioned arguments, together with our results, show clearly that a lock-and-key model is not appropriate to describe substrate recognition. The theoretical results also indicate that an induced-fit binding cannot be invoked either, as nonsubstrates can bind the *substrate-fitted* active site with higher affinity.

Based on all of these observations we propose a new, consistent hypothesis to explain substrate recognition by PR. *HIV-1 protease only cleaves substrates that are sufficiently exposed for a given time; i.e., many nonsubstrates are not cleaved, despite having greater affinity, because they are buried inside the polyprotein and physically hidden from PR. Among the exposed residues, the recognition is based on the geometric specificity of PR:Gag and PR:Gag-Pol multiprotein complex that determines which residues lie in a position accessible to the active site.*

Therefore, more important than the affinity of the peptide is the geometry of the complex between HIV-1 PR and the successive polyprotein fragments. It is well-known that most protein−protein associations occur with a highly specific pose; therefore, the accessibility of the active site to the polyprotein will be extremely constrained by the protein: protein complexation geometry. In fact, this would be the only viable selection mechanism for a selective enzyme exhibiting promiscuous binding at the active-site level. This explains the fact that cleavage of Gag-Pol occurs sequentially and with high fidelity at unrelated cleavage sites.[47] In contrast, active-site recognition cannot explain the sequential nature of the cleavage. The 3D structure of the polyprotein is unknown, but it seems logical that when the first substrate is cleaved a conformational/folding rearrangement occurs (or at least a new region becomes PR accessible) and the new polyprotein conformation, as well as the new PR:polyprotein complex, promote the cleavage of the second substrate and so on.

Moreover, it is known that alterations in the active site of PR might allow the virus to escape inhibition by antiviral compounds while maintaining the necessary points of cleavage to produce structural proteins.[47] Again, our proposal is in agreement with this, and therefore, independently of the substrate having more or less affinity, the same substrates in the same order are cleaved, *i.e., in the order in which the multiprotein PR:Gag and PR:Gag-Pol complex places them exposed to the active site for HIV-1 protease to bind and to cleave them.*

Additionally, the increase in viral fitness due to the frequent case of a mutation in the polyprotein (outside the cleavage motif) together with a mutation in PR (outside the active site) can be explained by the recognition model proposed here. As both mutations occur outside the reactive centers, the increase in fitness must be due to interactions outside the active site. It can well be due either to a more favorable interaction and binding between the two binding partners or to a better, more accurate recognition, with less cleavage errors.

Experimental observation of the several PR:Gag and PR:Gag-Pol successive complexes is a daunting task. Up until now, not even the precursor polyproteins have been crystallized. However, the amount of evidence that has been gathered together here seems to be more than enough to show that recognition is not active-site-based and to predict that recognition is ultimately based on the multiprotein interface.

The present findings are very exciting, as they point to the exploration of new targets in anti-HIV1 therapy, i.e., the PR: Gag and PR:Gag-Pol interfaces. They further direct a way to the discovery and development of a whole class of anti-HIV1 drugs, i.e., molecules that bind at the recognition sites of the multiprotein complex.

References and Notes

(1) Toh, H.; et al. Retroviral Protease-Like Sequence in the Yeast Transposon Ty1. *Nature* **1985**, *315* (6021), 691.

(2) Meek, T. D.; et al. Human Immunodeficiency Virus-1 Protease Expressed in Escherichia-Coli Behaves as a Dimeric Aspartic Protease. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86* (6), 1841–1845.

(3) Wlodawer, A.; et al. Conserved Folding in Retroviral Proteases - Crystal-Structure of a Synthetic Hiv-1 Protease. *Science* **1989**, *245* (4918), 616–621.

(4) Lapatto, R.; et al. X-Ray-Analysis of Hiv-1 Proteinase at 2.7 a Resolution Confirms Structural Homology among Retroviral Enzymes. *Nature* **1989**, *342* (6247), 299–302.

(5) Navia, M. A.; et al. 3-Dimensional Structure of Aspartyl Protease from Human Immunodeficiency Virus Hiv-1. *Nature* **1989**, *337* (6208), 615–620.

(6) Miller, M.; et al. Crystal-Structure of a Retroviral Protease Proves Relationship to Aspartic Protease Family. *Nature* **1989**, *337* (6207), 576–579.

(7) Ingr, M.; et al. Kinetics of the dimerization of retroviral proteases: The "fireman's grip" and dimerization. *Protein Sci.* **2003**, *12* (10), 2173–2182.

(8) Jacks, T.; et al. Characterization of Ribosomal Frameshifting in Hiv-1 Gag-Pol Expression. *Nature* **1988**, *331* (6153), 280–283.

(9) de Oliveira, T.; et al. Variability at human immunodeficiency virus type 1 subtype C protease cleavage sites: an indication of viral fitness. *J. Virol.* **2003**, *77* (17), 9422–9430.

(10) Krausslich, H. G.; et al. Activity of Purified Biosynthetic Proteinase of Human Immunodeficiency Virus on Natural Substrates and Synthetic Peptides. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86* (3), 807–811.

(11) Swanstrom, R.; Wills, J. W. Retroviral gene expression. II. Synthesis, processing, and assembly of viral proteins. In *Retroviruses*; Coffin, J. M., Hughes, S. H., Varmus, H. E., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1997; pp 263−334.

(12) Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A. How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. *J. Mol. Biol.* **2000**, *301* (5), 1207–1220.

(13) Ridky, T. W.; et al. Human immunodeficiency virus, type 1 protease substrate specificity is limited by interactions between substrate amino acids bound in adjacent enzyme subsites. *J. Biol. Chem.* **1996**, *271* (9), 4709–4717.

(14) Blondel, A. Ensemble variance in free energy calculations by thermodynamic integration: Theory, optimal "Alchemical" path, and practical solutions. *J. Comput. Chem.* **2004**, *25* (7), 985–993.

(15) Pitera, J. W.; Van Gunsteren, W. F. A comparison of non-bonded scaling approaches for free energy calculations. *Mol. Simul.* **2002**, *28* (1−2), 45–65.

(16) Shirts, M. R.; et al. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.* **2003**, *119* (11), 5740–5761.

(17) Kollman, P. A.; et al. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33* (12), 889–897.

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

(19) Piana, S.; et al. Reaction mechanism of HIV-1 protease by hybrid carparrinello/classical MD simulations. *J. Phys. Chem. B* **2004**, *108* (30), 11139–11149.

(20) Perez, M. A. A.; Fernandes, P. A.; Ramos, M. J. Drug design: New inhibitors for HIV-1 protease based on Nelfinavir as lead. *J. Mol. Graphics Modell.* **2007**, *26*, 634–642.

(21) Boeckmann, B.; et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **2003**, *31* (1), 365–370.

(22) Biosym technologies, S.D., *InsightII v. 2.3.0.* 1993.

(23) Wang, W.; et al. Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211–243.

(24) van der Spoel, D.; Lindahl, E.; Hess, B.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sjibers, A. L. T. M.; Feenstra, K. A.; van Drunen, R.; Berendsen, H. J. C. *Gromacs User Manual*, version 3.3; 2005.

(25) Van der Spoel, D.; et al. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26* (16), 1701–1718.

(26) Straatsma, T. P.; McCammon, J. A. Computational Alchemy. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407–435.

(27) Kollman, P. Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93* (7), 2395–2417.

(28) Wang, W.; et al. Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211–243.

(29) Pearlman, D. A. A Comparison of Alternative Approaches to Free-Energy Calculations. *J. Phys. Chem.* **1994**, *98* (5), 1487–1493.

(30) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method 0.1. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22* (8), 1420–1426.

(31) Mark, A. E.; van Gunsteren, W. F. Decomposition of the free energy of a system in terms of specific interactions. Implications for theoretical and experimental studies. *J. Mol. Biol.* **1994**, *240* (2), 167–176.

(32) Cornell, W. D.; et al. A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179–5197.

(33) Jorgensen, W. L.; et al. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.

(34) Swope, W. C.; et al. A Computer-Simulation Method for the Calculation of Equilibrium-Constants for the Formation of Physical Clusters of Molecules - Application to Small Water Clusters. *J. Chem. Phys.* **1982**, *76* (1), 637–649.

(35) Izaguirre, J. A.; et al. Langevin stabilization of molecular dynamics. *J. Chem. Phys.* **2001**, *114* (5), 2090–2098.

(36) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, H. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California: San Francisco, CA, 2004.

(37) Tsui, V.; Case, D. A. *Biopolymers* **2001**, *56*, 275−291.

(38) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341.

(39) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin Dynamics of Peptides - the Frictional Dependence of Isomerization Ratesof N-Acetylalanyl-N'-Methylamide. *Biopolymers* **1992**, *32* (5), 523–535.

(40) Huo, S.; Massova, I.; Kollman, P. A. Computational alanine scanning of the 1: 1 human growth hormone-receptor complex. *J. Comput. Chem.* **2002**, *23* (1), 15–27.

(41) Rocchia, W.; Alexov, A.; Honig, B. Extending the applicability of the nonlinear Poisson-Boltzmann equation: Multiple dielectric constants and multivalent ions. *J. Phys. Chem. B* **2001**, *105* (28), 6507–6514.

(42) Rocchia, W.; et al. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects. *J. Comput. Chem.* **2002**, *23* (1), 128–137.

(43) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Accuracy of the numerical solution of the Poisson-Boltzmann equation. *THEOCHEM* **2005**, *729* (1−2), 11–18.

(44) Connolly, M. L. Analytical Molecular-Surface Calculation. *J. Appl. Crystallogr.* **1983**, *16* (OCT), 548–558.

(45) Poorman, R. A.; et al. A Cumulative Specificity Model for Proteases from Human-Immunodeficiency-Virus Type-1 and Type-2, Inferred from Statistical-Analysis of an Extended Substrate Data-Base. *J. Biol. Chem.* **1991**, *266* (22), 14554–14561.

(46) Hazebrouck, S.; et al. Local and spatial factors determining HIV-1 protease substrate recognition. *Biochem. J.* **2001**, *358*, 505–510.

(47) Dunn, B.; Goodenow, M.; Gustchina, A.; Wlodawer, A. Retroviral proteases. *Genome Biology* **2002**, *3* (4).

(48) Billich, S.; et al. Synthetic Peptides as Substrates and Inhibitors of Human Immune-Deficiency Virus-1 Protease. *J. Biol. Chem.* **1988**, *263* (34), 17905–17908.

JP910958U