PROTEINS: Structure, Function, and Genetics 32:7–16 (1998)

RESEARCH ARTICLES

Reaction Path and Free Energy Calculations of the Transition Between Alternate Conformations of HIV-1 Protease

Steven W. Rick,[1]* John W. Erickson,[2] and Stanley K. Burt[1]
[1]*Frederick Biomedical Supercomputing Center, SAIC-Frederick, NCI-Frederick Cancer Research and Development Center, Frederick, Maryland*
[2]*Structural Biochemistry Program, SAIC-Frederick, NCI-Frederick Cancer Research and Development Center, Frederick, Maryland*

ABSTRACT **Two different structures of ligand-free HIV protease have been determined by X-ray crystallography. These structures differ in the position of two 12 residue, β-hairpin regions (or "flaps") which cap the active site. The movements of the flaps must be involved in the binding of substrates since, in either conformation, the flaps block the binding site. One of these structures is similar to structures of the ligand-bound enzyme; however, the importance of both structures to enzyme function is unclear. This transformation takes place on a time scale too long for conventional molecular dynamics simulations, so the process was studied by first identifying a reaction path between the two structures and then calculating the free energy along this path using umbrella sampling. For the ligand-free enzyme, it is found that the two structures are nearly equally stable, with the ligand-bound-type structure being less stable, consistent with X-ray crystallography data. The more stable open structure does not have a lower potential energy, but is stabilized by entropy. The transition occurs through a collapse and reformation of the β-sheet structure of the conformationally flexible, glycine-rich flap ends. Additionally, some problems in studying conformational changes in proteins through the use of a single reaction path are addressed. Proteins 32:7–16, 1998.** © 1998 Wiley-Liss, Inc.

Key words: conformational change; free energy calculations; HIV protease; molecular dynamics simulations; protein structure

INTRODUCTION

In the pursuit of antiviral drugs for acquired immune deficiency syndrome (AIDS), more crystal structures of the human immunodeficiency virus-1 (HIV-1) protease have been determined than perhaps of any other protein.[1] These structures indicate a significant difference between the ligand-free and inhibitor-bound enzyme. The enzyme is a homodimer with each subunit containing 99 residues. The active site region is capped by two identical β-hairpin loops, the flaps, from each monomer. In inhibitor-bound structures such as 4HVP, the flaps adopt a closed conformation, forming direct contacts with the substrate.[2] In ligand-free structures such as 3HVP, the flap is more open, with the position of the flap tip residues about 7 Å from the inhibitor-bound conformation (Fig. 1).[3,4,5] The flaps are, therefore, flexible and must move in order for substrates to bind since, in all of the known structures, they block access to the active site. There are X-ray structures for the closely related simian immunodeficiency virus (SIV) and HIV-2 proteases as well, which are about 50% homologous to HIV-1 and about 90% to each other. Of the 12 residues comprising the flaps, four residues, which are near the start and end of the loops, are different between HIV-1 and HIV-2. These four residues are identical in SIV and HIV-2. For SIV protease, the observed crystal structure of the ligand-free enzyme resembles that of the inhibitor-bound closed conformation, while the crystal structure of ligand-free HIV-2 protease resembles the open conformation.[6,7] In addition, free HIV-1 protease may also adopt the closed conformation as well[8]. The structure of another retroviral aspartyl protease, from the

The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the U.S. Government.

*Correspondence to: Steven W. Rick, Frederick Biomedical Supercomputing Center, SAIC-Frederick, NCI-Frederick Cancer Research and Development Center, Frederick, MD 27102.

Received 29 September 1997; Accepted 6 February 1998

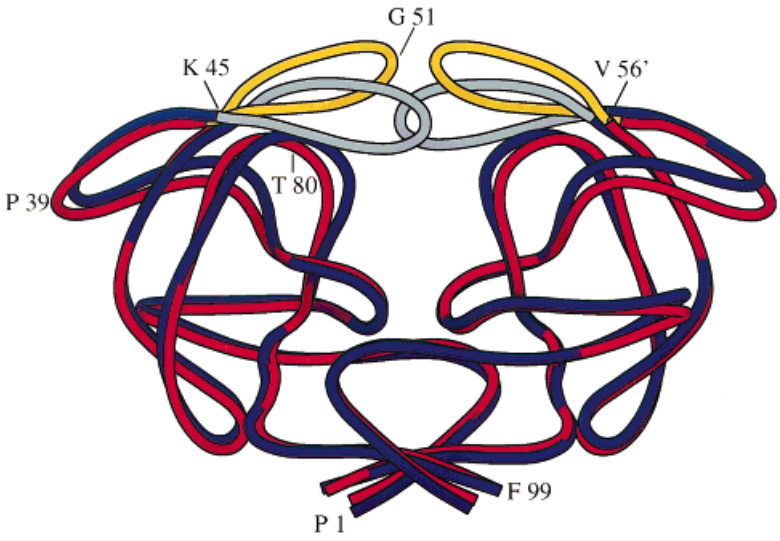

Fig. 1. C_α traces of the open[4] (red) and closed[2] (blue) structures of HIV-1 protease, with flap residues 45–56 of each monomer shown in gray (closed) and orange (open). (This figure was generated using MOLSCRIPT.[44])

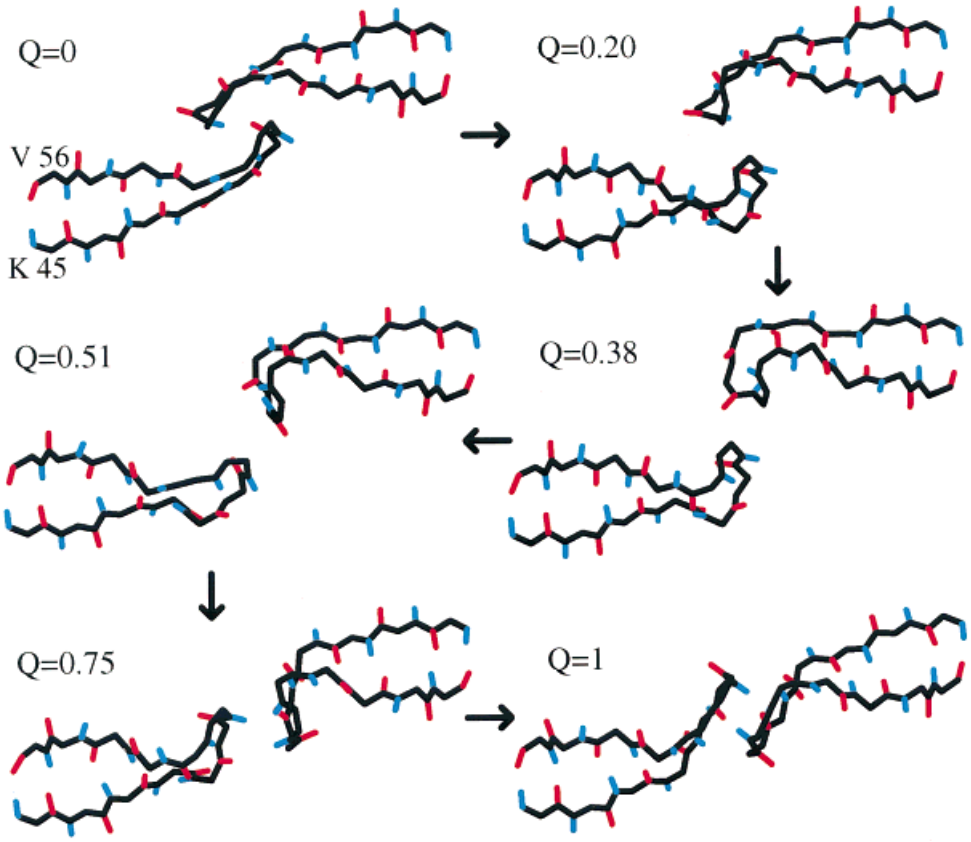

Fig. 2. Backbone atoms (with the NH bond in blue and the C=O bond in red) of residues 45–56 of each monomer viewed from the top of Figure 1, down the C_{2V} axis toward the active site for selected structures along the reaction path, Q. (This figure was generated using MOLSCRIPT.[44])

Rous Sarcoma Virus, has flap residues which could not be resolved, suggesting the absence of a single stable conformation of the flaps.[9] Therefore, it seems that for HIV protease, both the open and closed structures represent local free energy minima, with movement between the two structures limited.

Conformational changes in proteins have been observed in a variety of ligand binding and other biological processes.[10] On the other hand, observed conformational differences in X-ray structures may be the result of crystallization conditions[11] or crystal lattice forces[12] and not necessarily related to the protein's function. Computer simulations could help to characterize the energetic and dynamical relationships of the different conformations; however, the time scales of these changes for proteins are typically too long for conventional molecular dynamics or Monte Carlo simulations. Umbrella sampling can be used to study such rare event problems, but for this method to be used an appropriate reaction coordinate must be identified.[13] Torsional angles, atom-to-atom distances, and radius of gyrations have all been used as umbrella sampling coordinates for proteins.[14,15,16] For transitions from one structure to another which cannot be described in terms of a single torsional angle or atom-to-atom distance, a more general method is needed. These processes can be studied by determining a low energy path between the known structures.[17,18,19,20,21,22] Once such a pathway is identified, then umbrella sampling can be used to calculate the free energy difference between the two states. Whether the actual conformation change of the protein can be represented by a single pathway remains to be determined and this question will be addressed in later sections. Nevertheless, the free energy difference between the structures can be determined, since it does not depend on the path.

Experimental data provide some estimates about the time scales of flap movements. Kinetic studies of inhibitor binding support both a one-step mechanism, with flap motion accompanying binding, and a two-step process, with the formation of a collision complex and then the slower conversion to a tighter complex, depending on the inhibitor.[23,24,25] More tightly bound inhibitors seem to bind as a one-step process; however, both types of binding may in fact involve two steps, with the more potent inhibitors having a faster second flap-closing step and appearing as a one-step process in which the lifetime for the collision complex is very short.[23,24] In this interpretation of the data, the time scale for flap closing is strongly influenced by the ligand. The rate constants of the second step is 7 s^{-1} for a 0.14 μM inhibitor[23] and greater than 370 s^{-1} for a 9 nM inhibitor.[24] The k_{cat} of the enzyme is about 10 s^{-1},[24] so the binding of natural (μM) substrates may compete with the cleavage of the scissile peptide bond as the rate-limiting step of catalysis. Another estimate of the time scale of flap conformation changes comes from nuclear magnetic resonance (NMR) experiments, which find that the peptide bond at the tips in HIV-1 protease of the flaps undergo a 180° flip on a μs time scale.[26] The role of this localized motion in the conformation change will be examined. In addition, flap motion may play an important role in drug resistance. Maschera et al.[25] assert that the reduction in affinity of mutant proteases for saquinavir is due to an increase in the inhibitor-enzyme dissociation rate constant, which might involve conformational changes in the flaps as the inhibitor is released. A molecular dynamics study of the ligand-free enzyme also showed that the movement of the flap may be affected by mutations.[27]

The transition between the two known crystal structures of the substrate-free enzyme represents a first step toward a computational study of the more complex thermodynamics and kinetics of substrate and inhibitor binding. In addition, this study will determine whether the open or closed form of the enzyme is more stable and why two crystal structures are observed.

METHODS

Determining a Reaction Path

In an initial calculation, a path between the two structures is found using the Self-Penalty Walk method, which replicates the system M times, with replicas 1 and M being the starting and ending structures, respectively, and the others representing intermediates.[19,28] An energy, S, involving all the replicas is defined as

$$S = \sum_{i=1}^{M} V_i + \gamma \sum_{i=1}^{M-1} (d_{i,i+1} - \langle d \rangle)^2 + \rho \sum_{i=1}^{M-2} \exp\left(-\lambda d_{i,i+2}^2 / \langle d \rangle^2\right) \quad (1)$$

where V_i is the potential energy of replica i, $d_{i,j} = |\mathbf{R}_i - \mathbf{R}_j|$, the distance between replicas i and j with $\mathbf{R}_i$ the positions of the atoms of replica i, <d> the average distance between adjacent replicas,

$$\langle d \rangle = \frac{1}{M-1} \sum_{1=1}^{M-1} d_{i,i+1}$$

and γ, ρ, and λ are parameters to be chosen. Given an initial guess for the replicas, the positions of all the intermediates are moved to minimize the energy S. The structures which minimize S are then an approximation to structures along the minimum energy path between the two endpoints. The second term in Equation 1, the distance between nearest neighbors, keeps all the replicas evenly spaced along the path and the third term is a repulsion between next nearest neighbors, which keeps the path from turning back on itself.

Umbrella Sampling

Once the path is known, then the potential of mean force (PMF) can be calculated using umbrella sampling.[13] The system coordinates, **R**, are made to sample regions along the path using a biasing poten-

tial, $E_{rest} = \alpha(\mathbf{R} - \mathbf{R}_j)^2$, centered at each replica j. In the present application, we chose to restrain the mainchain heavy atoms (C, C_α, N, and O) of the residues in the flap region which move the most from structure to structure (residues 45–56 and 45′–56′) where the prime denotes the other monomer.

The Reaction Coordinate

The reaction path is a set of straight line segments connecting $\mathbf{R}_1$ with $\mathbf{R}_2$, $\mathbf{R}_2$ with $\mathbf{R}_3$,..., and $R_{M-1,}$ with $\mathbf{R}_M$. The reaction coordinate of a point on this path is chosen to be the distance along the path as measured by following successive segments, starting at $\mathbf{R}_1$. The reaction coordinate of each replica is then defined as

$$Q_i = \sum_{j=1}^{i-1} |R_{j+1} - R_j| / Q_{tot} \tag{2}$$

where

$$Q_{tot} = \sum_{j=1}^{M-1} |R_{j+1} - R_j|.$$

We normalized the path by Q_{tot} so that $Q_1 = 0$ and $Q_M = 1$. In general, the dynamics simulation produces conformations not lying on the reaction path; the reaction coordinate of an arbitrary conformation, $\mathbf{R}$, is defined as the projection of $\mathbf{R}$ on the nearest segment, $\mathbf{R}_{i+1} - \mathbf{R}_i$, of the reaction path,

$$Q(\mathbf{R}) = Q_i + (\mathbf{R} - \mathbf{R}_i) \cdot \mathrm{e}_i / Q_{tot} \tag{3}$$

with $\mathbf{e}_i$ the unit vector in the direction of the path at point i,[19,28]

$$\mathrm{e}_i = (\mathbf{R}_{i+1} - \mathbf{R}_i) / |\mathbf{R}_{i+1} - \mathbf{R}_i|.$$

If the path turns sharply, then the nearest part of the path may be the vertex at $\mathbf{R}_j$ rather than either of the two segments joining at $\mathbf{R}_j$; in this case, $Q(\mathbf{R})$ is set equal to Q_j. In calculating $Q(\mathbf{R})$ for the simulation data restrained to the vicinity of a particular replica, $\mathbf{R}_j$, the start of the nearest segment (i) is assumed to be either j-1 or j, depending on whether the point is behind or ahead of $\mathbf{R}_j$ on the path. Notice that with this definition of $Q(\mathbf{R})$, if $\mathbf{R}$ goes to $\mathbf{R}_{j\pm1}$, then $Q(\mathbf{R})$ goes to $Q_{j\pm1}$.

Unbiasing the Data

The biasing potential can be written as

$$E_{rest} = \alpha((\mathbf{R} - \mathbf{R}_j)^2 = \alpha(Q(\mathbf{R}) - Q_j)^2 Q_{tot}^2 + \alpha P^2 \tag{4}$$

where $\mathbf{P}$ is everything in $\mathbf{R}$ orthogonal to $\mathbf{e}_i$,

$$\mathbf{P} = (\mathbf{R} - \mathbf{R}_j) - (Q(\mathbf{R}) - Q_j) Q_{tot} \mathrm{e}_i. \tag{5}$$

Then to calculate the potential of mean force, for each replica point a probability function is generated, which is biased with respect to Q but unbiased with respect to $\mathbf{P}$,

$$\langle N(Q) \rangle_j = \frac{\langle \delta(Q - Q(\mathbf{R})) e^{\beta\alpha P^2} \rangle_j}{\langle e^{\beta\alpha P^2} \rangle_j} \tag{6}$$

where

$$\langle \cdots \rangle_j = \frac{\int d\mathbf{R} (\cdots) e^{-\beta H_0} e^{-\beta E_{rest}}}{\int d\mathbf{R} e^{-\beta H_0} e^{-\beta E_{rest}}},$$

β is $1/k_BT$, k_B is Boltzmann's constant and H_0 is the unbiased part of the Hamiltonian. The potential of mean force, W(Q), is calculated from the set of $<N(Q)>_j$ distributions using the weighted histogram method.[29,30]

Revising the Reaction Path

The reaction path, with $M = 85$, connecting minimized structures corresponding to the closed structure[8] and the open structure, 1HHP,[5] was found using the program MOIL.[31] In order to include some of the aqueous environment, 75 crystal waters were retained in these structures. The calculation of the potential of mean force along the path was calculated using the AMBER4.1 software package[32] and the Amber 1994 force field.[33] This force field and program were used because simulations of HIV-1 protease using the Amber force field and Ewald sums for the long-ranged electrostatic interactions showed good agreement with X-ray crystal structure data over 200 ps trajectories.[34] Since the force field used in the PMF calculations is different from the force field used to generate the path, and also since the PMF calculations include bulk solvent, the AMBER-generated trajectories deviate from the MOIL-generated path. To correct for this and find the path that the AMBER force field will follow, the trajectories generated by umbrella sampling along the path are used to define a new path. This procedure was iterated a few times until the new path deviated from the old path by no more than 0.7 Å.

Simulation Details

The path is defined in terms of the cartesian coordinates of the M replicas and not in a molecular reference frame. Therefore, to use the cartesian representation of the path in umbrella sampling, corrections must be made for the rigid-body translations and rotations. Rather than eliminating the rigid-body motions of the system, as done elsewhere,[19,28] we chose to allow rigid-body motion but make the reference coordinates follow the system by minimizing the root-mean-square distance between the restrained coordinates of the reference and the system before enforcing the restraints each time

step. This modification to AMBER 4.1 added no noticeable run-time to the program.

The present simulations were done in an environment of 3,349 TIP3P water molecules[35] with five chloride counter ions for charge neutrality, a 1 fs time step, rigid bond lengths enforced with SHAKE,[36] particle mesh Ewald sums,[37] α equal to 1 kcal/mol/Å^2 and in the canonical (constant T,V,N) ensemble using the Berendsen coupling algorithm with T = 300K.[38] The constant for the biasing potential, α, was chosen so that there is sufficient overlap of $(N(Q)($ with neighboring points. However, for some (13) of the points there was insufficient overlap, so additional simulations at points halfway between the neighboring replicas were performed. For each of the points on the path, 45 ps trajectories were run, with 15 ps for equilibration, starting with equilibrated coordinates from a neighboring point, and 30 ps for data collection.

RESULTS

The two structures differ in a number of ways, as illustrated in Figure 1. The tip of the left flap is in front of the right flap in the closed (gray) and in back in the open structure (orange). So the flaps cannot move as rigid bodies from one dimeric structure to the other, since they would collide with each other. Internal rearrangement of the flaps must occur during the transition. Figure 2 shows the final structures and four intermediates along the path. In order to make a good interflap hydrogen bond in the closed structure, one of the flap tips is a β-I turn and the other is β-II (in Fig. 2 at Q = 0, the top flap is β-I and the bottom is β-II). In the open structure (Q = 1), this hydrogen bond does not exist and both turns are β-II, so the transition between the two structures involves this change in the turn of one of the flaps. Consistent with these observations about the beginning and final structures, the path shows a lot of flexibility in the internal coordinates. There are 12 changes in the backbone torsion angles of more than 110° along the path, involving residues 48–52 for both monomers. The torsion angles of hairpin termini residues 45–47 and 54–56 stay near the values for an anti-parallel β-sheet. The torsional angles for Phe53, which has been mutated to a tryptophan for use as a fluorescent probe for inhibitor binding,[24] change by about 60° during the transition, indicating that the fluorescence of Trp53 might be sensitive to the conformational changes of the ligand-free enzyme. From Figure 2 it is apparent that the flap tip residues show a lot of flexibility along the path, while the residues near the ends of the flap maintain their β-sheet structure. The flexibility of the flap tips and the loss of β-sheet structure was also observed in a previous simulation of flap movement.[27]

From Equation 2, Q depends on how many atoms in the system define the coordinate set, **R**. For large systems, the number of atoms necessary to define the path is a difficult question. Is the path defined by the entire protein, including sidechains, or by only the mainchain atoms, or by only part of the protein? What about strongly bound water molecules? Most potential of mean force calculations for large systems use a coordinate that only depends on a few degrees of freedom, such as a torsional angle[14] or an end-to-end distance,[15] and so avoid this problem. The potential of mean force and the free-energy barrier may strongly depend on how the path is defined for two separate reasons. One concerns the familiar difficulties with dividing surfaces in transition state theory. If the path depends on coordinates which involve conformational changes with high free-energy barriers (such as the β-I to β-II turn transition), then the free energy of that path will be higher than that of a path which is not dependent on those degrees of freedom. This can be corrected by calculating the transmission coefficient, which will be very different for the above two path definitions. (For a further discussion of this point, see reference 28.) Second, the transition may take place through not just one but a number of thermally accessible paths. This will be especially important if the transition state is more disordered than the product or reactant state. Therefore, restraining the coordinates around a particular value will create an artificially high free energy for the more disordered transition state. One can get around this problem by including all thermodynamically accessible paths; however, if many of these paths are different only in the position of a few degrees of freedom, then a single path without dependence on these degrees of freedom will do just as well. One might hope to have a sufficiently loose definition of the reaction coordinate which covers all paths and yet still involves the rate-limiting conformational changes, although there is no rigorous way to do this. These two effects may either underestimate the barrier by being independent of the free-energy barriers associated with important conformational changes or overestimate the barrier by only including a subset of the larger region of conformational space which describe the transition state. Clearly, estimating rates from thermodynamic barriers should be done carefully. Three alternate definitions of the path will be examined to see how that influences the potential of mean force. These paths are defined as:

1. Mainchain heavy atoms (C, C_α, N, and O) for residues 45–56 and 45′–56′ (96 atoms);
2. C_α atoms for residues 45–56 and 45′–56′ (24 atoms); or
3. Mainchain heavy atoms for the residues at the flap tips, 49–51 and 49′–51′ (24 atoms).

We can analyze the same set of simulation data, with the same restraints, for these three definitions of Q. The different definitions will just change which coordinates contribute to Q and which are the orthogonal coordinates, **P** (see Eq. 5).

TABLE I. Thermodynamic Properties for the Closed to Open Transition in HIV-1 Protease

Path definition	ΔA (kcal/mol)	−TΔS (kcal/mol)	ΔE (kcal/mol)	$\Delta A^{\dagger}$ (kcal/mol)
1	−13 ± 5	−17 ± 17	4 ± 22	25 ± 6
2	−2 ± 2	−10 ± 15	8 ± 20	11 ± 3
3	−6 ± 6	−6 ± 21	0 ± 22	20 ± 6
Average	−7 ± 3	−11 ± 10	4 ± 12	

The equilibrium constant and free energy change for the process can be calculated from

$$\Delta A = -kT \ln\left(\int_{Q^{\dagger}}^{1} e^{-W(Q)/kT}\, dQ \Big/ \int_{0}^{Q^{\dagger}} e^{-W(Q)/kT}\, dQ\right) \quad (7)$$

where $Q^{\dagger}$ is the value of Q for which the free energy is maximum. The value of ΔA should be independent of the path between the two states and a comparison of the values for the different path definitions provides an estimate of the errors of the calculations (see Table I). The sampling errors can be found by breaking the 30 ps of data into three parts and calculating the thermodynamic properties from the first 10 ps, the middle 10 ps, and the last 10 ps. The resulting standard deviations are the reported error bars shown in Table I. The entropy can be calculated from[29]

$$-TS(Q) = W(Q) - \sum_{j} \langle E(Q)\rangle_j \big/ \langle N(Q)\rangle_j \quad (8)$$

where the sum is over all the umbrella sampling points j and $<E(Q)>_j$ is the average potential energy at j as a function of Q,

$$\langle E(Q)\rangle_j = \frac{\langle \delta(Q - Q(\mathbf{R}))H_0(\mathbf{R})e^{\alpha P^2/kT}\rangle_j}{\langle e^{\alpha P^2/kT}\rangle_j} \quad (9)$$

The entropy change for the transition is then given by

$$\Delta S = \frac{\int_{Q^{\dagger}}^{1} S(Q)e^{-W(Q)/kT}\, dQ}{\int_{Q^{\dagger}}^{1} e^{-W(Q)/kT}\, dQ} - \frac{\int_{0}^{Q^{\dagger}} S(Q)e^{-W(Q)/kT}\, dQ}{\int_{0}^{Q^{\dagger}} e^{-W(Q)/kT}\, dQ} \quad (10)$$

The energy change is given by ΔE=ΔA + TΔS. The calculations predict that the open structure is more stable than the closed structure. While the error bars are large for ΔE and ΔS, the calculations suggest that the transition is entropically favorable and energetically unfavorable (Table I).

Figure 3 shows W(Q) for the three different path definitions. The barrier depends strongly on how the path is defined. This is also apparent from the activation free energy,

$$\Delta A^{\dagger} = -kT \ln\left(e^{-W(Q^{\dagger})/kT} \Big/ \int_{0}^{Q^{\dagger}} e^{-W(Q)/kT}\, dQ\right) \quad (11)$$

shown in Table I. If the path is only defined in terms of C_α atoms, then the barrier is lower than if it also depends on C, N, and O (and so on the ϕ and ψ angles). Because the β-sheet structure is lost for the residues in the flap tip, allowing the path to be independent of the conformationally flexible hydrogen bonding groups decreases the free energy substantially for the transition state structures relative to the initial and final states.

It is helpful to consider the interaction energy between the residues on one flap and the residues on the other flap. In the Amber 4.1 force field, this energy involves Lennard-Jones and Coulombic interactions. The interflap interaction energy as a function of Q is shown in Figure 4A. The interflap energy for the closed structure is less by about 10 kcal/mol than for the open structure, and much of this energy is lost during the transition, although some loss in the interflap energy will be made up in increases in the interactions with the solvent. Figure 4B plots the torsion energy for residues 45–55 and 45′–55′ in the flap. From Figures 2–4, a sequence of steps for the transition can be determined. First, the flap tips pull away from each other in a direction perpendicular to the interflap hydrogen bond between the carbonyl oxygen of residue 50 and the amide hydrogen of residue 51′, disrupting the β-sheet structure of the first monomer. By Q = 0.2, the hydrogen bond is lost, the interflap energy increases, and W(Q) increases as well. Around Q = 0.38, the flap tip of the second monomer folds back toward itself and both flaps have a bent-L structure with high torsion energy and are only weakly interacting. This is the point with the highest free energy. The β-II to β-I turn transition of the second monomer takes place around Q = 0.5. Then the flap tips pass to the other side of each other, by Q = 0.75, and finally expand and reform their β-sheet structure, decreasing the torsional and interflap energies.

Figure 5 shows the RMS overlap of the C_α atoms with the closed structure[8] and the open structure[5] as a function of Q. At each endpoint, the overlap with the more stable structure is about 1.2 ± 0.2 Å, about the same RMS as seen in other simulations of free HIV protease.[34,39] The RMS difference between the two crystal structures is 2.0 Å. The changes in the RMS with Q are roughly monotonic, indicating that overlap with either structure correlates well with the transition coordinate. Also, it does not appear that any large structural differences from either crystal structure take place during the transition, since the RMS functions do not go through a maximum. The smoothness of the RMS overlaps suggests that no regions other than the flaps go through any large-scale motion during the transition.

Movement of nonflap residues during the transition can be examined by calculating a difference–distance matrix. For a particular window, centered at Q, the average distance between two residues is

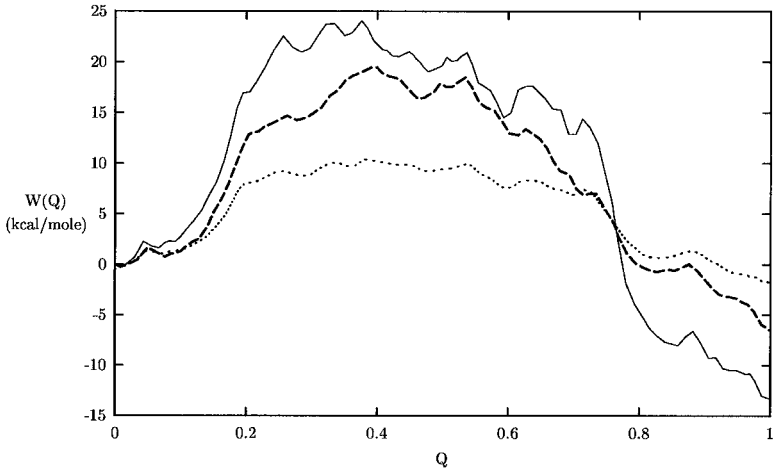

Fig. 3. Potential of mean force, W(Q), for three different definitions of the reaction coordinate, Q: mainchain heavy atoms for residues 45–56 and 45′–56′ (solid line); C_α?for residues 45–56 and 45′–56′ (dotted line); and mainchain heavy atoms for residues 49–51 and 49′–51′ (dashed line).

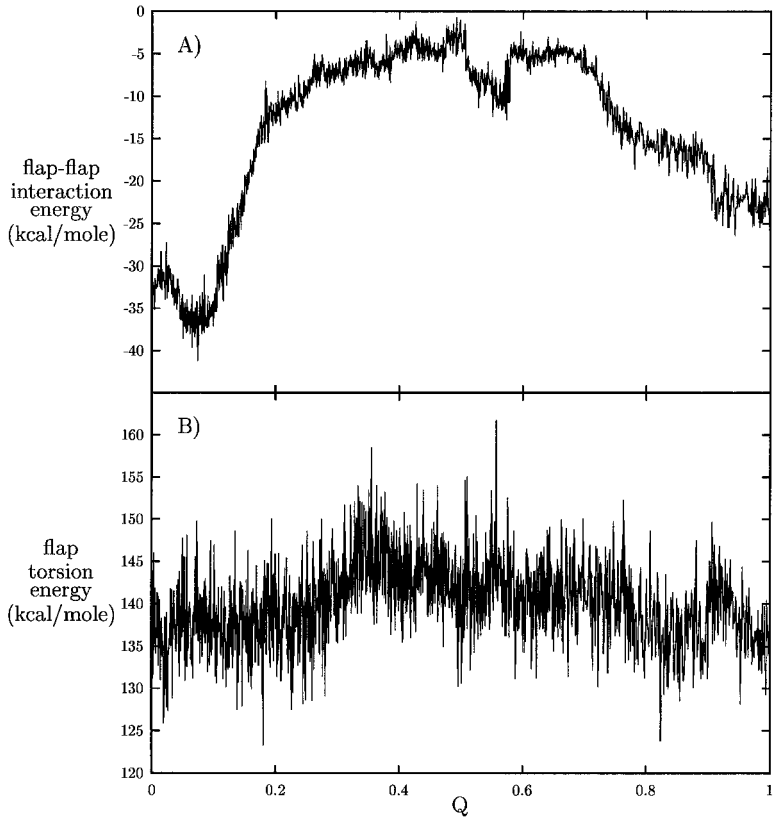

Fig. 4. Interaction energy between residues on different flaps (**A**) and torsion energy of the flap residues (**B**).

calculated, using the average of the residue's four mainchain heavy atoms to define the residue's position. The difference–distance matrix is the magnitude of the difference of each interresidue distance between two positions along the path. The matrix has large values for elements corresponding to residue distances, which change from one part of the path to another. The distance–difference matrix was calculated for the transition state region (Q = 0.38) versus both the starting and final states. The residues which show the greatest changes in the interresidue distances are the flap residues, centered around 50 and 149, which show large charges relative to almost all other residues (Fig. 6). The next largest feature is around 80 and 178 (Fig. 6B) and to a lesser extent around 180 (Fig. 6A). The region around residue 80 is close to flap region (there is contact between the sidechains Pro79 and Ile54 and a mainchain hydrogen bond between residues 77 and 57) and this region moves between the open and closed structures (see Fig. 1). Another region which moves from the open to the closed structure around residue 39 does not show any change in the difference–distance plot greater than 2.5 Å. If the open 3HVP[4]

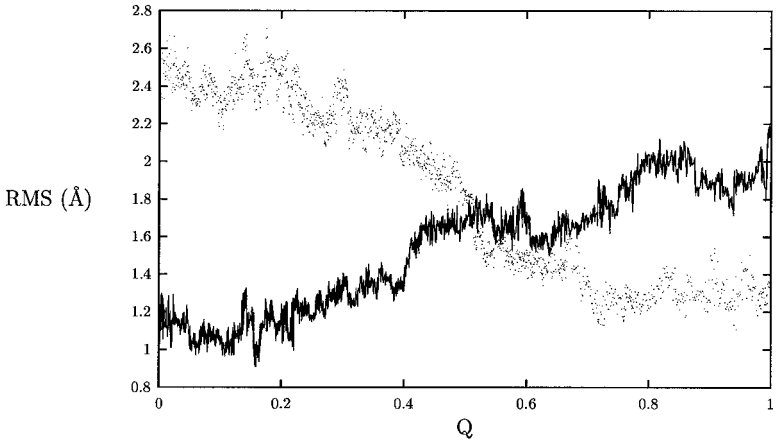

Fig. 5. The root-mean-square (RMS) deviation of C_α atoms as a function of Q, from the closed structure (solid line)[8] and the open structure (dashed line).[5]

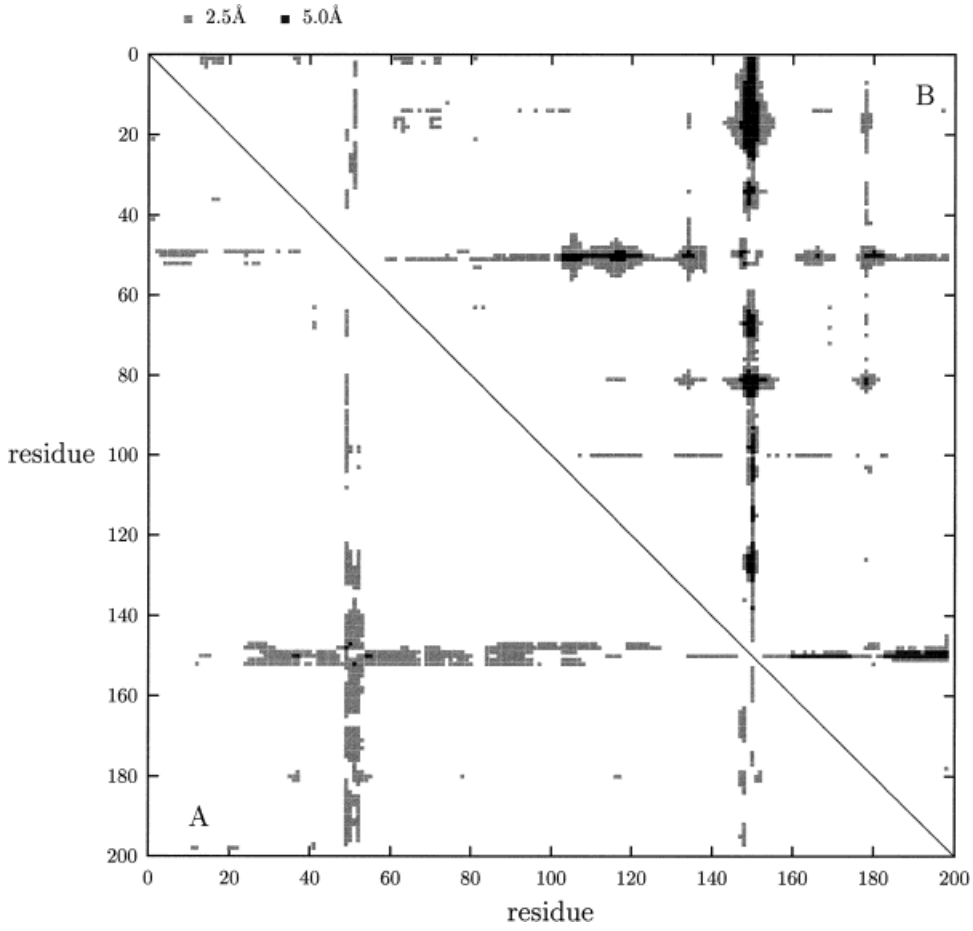

Fig. 6. The difference–distance matrix showing the magnitude of the difference in the interresidue distances for (**A**) Q = 0.38 − Q = O and (**B**) Q = 1 − Q = 0.38. Differences greater than 2.5 Å are shown in the gray squares and difference greater than 5 Å are shown in the black squares. Residues 1–99 of the second monomer are numbered as 100–198.

and the closed 4HVP[2] structures are superimposed by minimizing the RMS distance of all the C_α atoms, then the distances between the C_α atoms are 6.7 Å for residue 51, 2.4 Å for residue 39, and 1.8 Å for residue 39. So the magnitudes of the difference–distance matrix (flap residues being largest followed by the residues around 79) are compatible with the distances between the starting and final structures and it appears that the closed-to-open transition mostly involves the residues in the flap regions, and those residues in direct contact with the other regions not strongly coupled to the transition. There

are some values above 2.5 Å for the terminal residues (1–3,100 and 197–198).

DISCUSSION

The free energy difference for the closed to open transition is small enough so that factors such as mutations, crystal contacts, temperature, and substrate binding could shift the relative stability (see Table I). Crystal contacts involving residue 45 of the flap do stabilize the open structure.[3,4,34] Other flap resides interact with residues not on the flap but on the same monomer, involving van der Waals interactions between the sidechains of 32 and 47 and between 54 and 79. Residues 32, 45, 47, and 79 are among those that are different for SIV and HIV-1, so these sidechain differences may explain the different observed crystal structures.[6]

The error bars on these free energy changes are statistical estimates of sampling errors. Harder to estimate are systematic errors from the force field. In terms of the relative stabilities of the two structures, an important property of the force field is the relative strength of protein–protein relative to protein–water interactions. Additionally, it has been shown in studies of conformational changes in smaller molecules that the neglect of the conformational dependence of the molecular partial charges leads to errors of about 2 kcal/mol in the relative free energy changes.[40,41,42] The Amber 4.1 force field[32] was used because previous simulations starting from the open crystal structure did not deviate greatly from the starting structure.[34] From Figure 5 it is evident that both endpoints have a low RMS overlap with the relevant crystal structure. The larger RMS overlap with the open structure (about 1.3 Å) is due to the lack of crystal contacts, which stabilize the flaps in a slightly different configuration in the crystal form than in the aqueous form.[34] So at the very least, the potential is accurate enough to have stable structures corresponding to both the open and the closed structures.

A question remains as to the biological relevance of having two nearly equally stable structures of the same enzyme. The closed flap configuration positions the substrate with the correct orientation in the active site.[2] The open structure has weaker interflap interactions (see Fig. 4A) and has a higher potential energy than the closed structure. The weaker flap–flap interactions of the open structure may allow the flaps to open to bind ligands more easily and more rapidly than the closed structure. The flaps in HIV-2 and SIV protease, which are 90% identical in sequence, crystallize in the open and closed forms, respectively, in the absence of a ligand,[6,7] while the flaps in crystalline RSV protease are disordered.[9] An additional flap conformation, intermediate between the closed and open forms, has been observed for the HIV-1 protease complexed with a haloperidol derivative.[43] Thus, the flap structure of retroviral proteases displays a wide range of conformational flexibility that is presumably required for the specific polyprotein recognition and processing functions of these enzymes.

CONCLUSION

For ligand-free HIV-1 protease, these calculations predict that the free energy of the open structure is less than the closed structure by -7 ± 3 kcal/mol. The open form is the structure observed by X-ray crystallography. Interestingly, the more stable structure does not appear to have a lower potential energy, but rather is stabilized by entropy. The entropy for the transition, $-T\Delta S$, is -11 ± 10 kcal/mol and the enthalpy change is 4 ± 12 kcal/mol. The increased conformational freedom of the flap tips in the open structure is evident from X-ray crystallography experiments as well. For the open structures, the B-factors of the C_α atoms of the flap tips (residues 49–51) are roughly twice as large as the average for all the C_α atoms.[4,5] For the closed structures, including those with no substrate, the B-factors of the flap C_α atoms are not different from the average.[2,6,8] The closed structure is stabilized by stronger interactions between the flaps, most importantly, a hydrogen bond between the two flap tips. The transition occurs through a collapse and reformation of the β-sheet structure of the conformationally flexible, glycine-rich flap tips, rather than moving as rigid domains (Fig. 2). The disordered nature of the transition state region leads to a potential of mean force that depends strongly on which atoms are used to define the reaction coordinate, Q (Fig. 3). For this reason, a free-energy barrier for the transition cannot be unambiguously assigned. The nature of conformational transitions of proteins and the applicability of a reaction path formalism to study the dynamics of these transitions are areas which need to be examined further. However, we demonstrated that the formalism described here can be used to calculate the thermodynamics of structural transitions in proteins which involve complex rearrangements of many atoms.

REFERENCES

1. Wlodawer, A., Erickson, J.W. Structure-based inhibitors of HIV-1 protease. Annu. Rev. Biochem. 62:543–585, 1993.
2. Miller, M., Schneider, J., Sathyanarayana, B.K. et al. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. Science 246: 1149–1152, 1989.
3. Navia, M.A., Fitzgerald, P.M.D., McKeeve, B.M. et al. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. Nature (London) 337:615–620, 1989.
4. Wlodawer, A., Miller, M., Jaskolski, M. et al. Conserved folding in retroviral proteases: Crystal structure of a synthetic HIV-1 proteinase. Science 245:616–621, 1989.
5. Spinelli, S., Liu, Q.Z., Alzari, P.M., Hirel, P.H., Poljak, R.J. The three dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. Biochimie 73:1391–1396, 1991.
6. Wilderspin, A.F., Sugrue, R.J. Alternate native flap conformation revealed by 2.3 Å resolution structure of SIV protease. J. Mol. Biol. 239:97–103, 1994.

7. Chen, Z., Li, Y., Chen, E. et al. Crystal structure at 1.9-Å resolution of human immunodeficiency virus (HIV) II protease completed with L-735,524, an orally bioavailable inhibitor of the HIV proteases. J. Biol. Chem. 42:26344–26348, 1994.
8. Bhat, T.N., Erickson, J.W., private communication.
9. Miller, M., Jaskolski, M., Rao, J.K., Leis, J., Wlodawer, A. Crystal structure of a retroviral protease proves relationship to aspartic acid family. Nature (London) 337:576–579, 1989.
10. Schulz, G.E. Domain motions in proteins. Curr. Opin. Struct. Biol. 1:883–888, 1991.
11. Schiffer, M., Ainsworth, C., Xu, Z.-B. et al. Structure of a second crystal form of Bence-Jones protein loc: Strikingly different domain associations in two crystal forms of a single protein. Biochemistry 28:4066–4072, 1989.
12. Eigenbrot, C., Randal, M., Kossiakoff, A.A. Structural effects induced by mutagenesis affected by crystal packing factors: The structure of a 30–51 disulfide mutant of basic pancreatic trypsin inhibitor. Proteins 14:75–87, 1992.
13. McCammon, J.A., Harvey, S.C. "Dynamics of Proteins and Nucleic Acids." Cambridge, UK: Cambridge University Press, 1987.
14. Northrup, S.H., Pear, M.R., Lee, C.-H., McCammon, J.A., Karplus, M. Dynamical theory of activated processes in globular proteins. Proc. Natl. Acad. Sci. USA 79:4035–40397 1982.
15. Huston, S.E., Marshall, G.R. $\alpha/3_{10}$-Helix transitions in α-methylalanine homopeptides: Conformational transition pathway and potential of mean force. Biopolymers 34:75–90, 1994.
16. Boczko, E.M., Brooks, C.L. First-principles calculation of the folding free energy of a three-helix bundle protein. Science 269:393–396, 1995.
17. Pratt, L. A statistical method for identifying transition states in high dimensional problems. J. Chem. Phys. 85:5045–5048, 1986.
18. Elber, R., Karplus, M. A method for determining reaction paths in large molecules: Application to myoglobin. Chem. Phys. Lett. 139:375–380, 1987.
19. Czerminski, R., Elber, R. Self-avoiding walk between two fixed points as a tool to calculate reaction paths in large molecular systems. Int. J. Quant. Chem. 24:167–1867 1990.
20. Gillilan, R.E., Wilson, K.R. Shadowing, rare events, and rubber bands. A varational Verlet algorithm for molecular dynamics. J. Chem. Phys. 97:1757–1772, 1992.
21. Cho, A.E., Doll, J.D., Freeman, D.L. The construction of double-ended classical trajectories. Chem. Phys. Lett. 229: 218–224, 1994.
22. Smart, O.S. A new method to calculate reaction paths for conformational transitions of large molecules. Chem. Phys. Lett. 222:503–512, 1994.
23. Furfine, E.S., D'Souza, E., Ingold, K.J. et al. Two-step binding mechanism for HIV protease inhibitors. Biochemistry 31:7886–78911 1992.
24. Rodriguez, E.J., Deckman, C.D., Abu-Soud, H., Raushel, F. M., Meek, T.D. Inhibitor binding to the phe53trp mutant of HIV-1 protease promotes conformation changes detectable by spectroflourometry. Biochemistry 32:3557–3563, 1993.
25. Maschera, B., Palfi, G.D.G., Wright, L.L. et al. 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.
26. Nicholson, L.K., Yamazaki, T., Torchia, D.A. et al. Flexibility and function in HIV-1 protease. Struct. Biol. 2:274–280, 1995.
27. Collins, J.R., Burt, S.K., Erickson, J.W. Flap opening in HIV-1 protease simulated by "activated" molecular dynamics. Struct. Biol. 2:334–338, 1995.
28. Elber, R., Chen, D.P., Rojewska, D., Eisenberg, R. Sodium in gramicidin: An example of a permion. Biophys. J. 66:906–924, 1995.
29. Ferrenberg, A.M., Swendsen, R.H. Optimized Monte Carlo data analysis. Phys. Rev. Lett. 63:1195–1198, 1989.
30. Kumar, S., Bouzida, D., Swendsen, R.H., Kollman, P.A., Rosenberg, J.M. The weighted histrogram analysis method for free-energy calculations on biomolecules. I. The method. J. Comp. Chem. 13:1011–1021, 1992.
31. Elber, R., Roitberg, A., Simmerling, C. et al. "NATO Conference Proceeding on 'Statistical Mechanics and Protein Structure'." Doniach S. (ed.). New York: Plenum Press, 1994:165–191.
32. Pearlman, D.A., Case, D.A., Caldwell, J.W. "Amber 4.1." San Francisco: University of California, 1995.
33. Cornell, W.D., Cieplak, P., Bayly, C.I. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc. 117:5179–5197, 1995.
34. York, D.M., Darden, T.A., Pedersen, L.G., Anderson, M.W. Molecular dynamics simulation of HIV-1 protease in a crystalline environment and in solution. Biochemistry 32:1443–1453, 1993.
35. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79:926–9357 1983.
36. Allen, M.P., Tildesley, D.J. "Computer Simulation of Liquids." Oxford, UK: Oxford University Press, 1987.
37. York, D.M., Darden, T.A., Pedersen, L.G. The effect of long-range electrostatic interactions in simulations of macromolecular crystals: A comparison of the Ewald and truncated list methods. J. Chem. Phys. 99:8345–8348, 1993.
38. Berendsen, H.J.C., Postma, J.P.M., Gunsteren, W.F.V., DiNola, A., Haak, J.R. Molecular dynamics with coupling to to an external bath. J. Chem. Phys. 81:3684–36901 1984.
39. Harte, W.E., Jr., Swaminathan, S., Mansuri, M.M., Martin, J.C., Rosenberg, I.E., Beveridge, D.L. Domain communication in the dynamical structure of human immunodeficiency virus I protease. Proc. Natl. Acad. Sci. USA 87:8864–8868, 1990.
40. Jorgensen, W.L., Gao, J. Cis-trans energy difference for the peptide bond in the gas phase and in aqueous solution. J. Am. Chem. Soc. 110:4212–4216, 1988.
41. Cieplak, P., Kollman, P. On the use of electrostatic potential derived charges in molecular mechanics force fields. The relative salvation free energy of *Cis*- and *Trans*-*N*-methyl-acetamide. J. Comp. Chem. 12:1232,1991.
42. Rick, S.W., Berne, B.J. Dynamical fluctuating force fields: The aqueous salvation of amides. J. Am. Chem. Soc. 118:672–679, 1996.
43. Rutenber, E., Fauman, E.B., Keenan, R.J. Structure of a non-peptide inhibitor completed with HIV-1 protease. J. Biol. Chem. 268:15343–15346, 1993.
44. Kraulis, P.J. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. J. Appl. Cryst. 24:946–950, 1991.