Insight into the Folding Inhibition of the HIV-1 Protease by a Small Peptide

Massimiliano Bonomi,* Francesco L. Gervasio,* Guido Tiana,† Davide Provasi,† Ricardo A. Broglia,†‡ and Michele Parrinello*

*Computational Science, Department of Chemistry and Applied Biosciences, ETH Zürich, Lugano, Switzerland; †Department of Physics, University of Milan, Milan, Italy; and ‡The Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

ABSTRACT It has recently been shown that the highly protected segments 24–34 (S_2) and 83–93 (S_8) of each of the two 99-mers of human immunodeficiency virus type 1 protease play an essential role in the folding of the monomers, giving rise to the so-called (postcritical) folding nucleus ((FN) minimum condensation unit ensuring folding) when they dock. This scenario received further support from model calculations that demonstrated that the peptide p-S_8, displaying an amino acid sequence identical to the corresponding (83–93) segment of the monomer, can be used to interfere with the formation of the FN and eventually to inhibit folding by docking the fragment 24–34. Experiments in vitro and in cells infected with ex vivo wild-type and multiresistant HIV isolates confirm that the inhibition power of p-S_8 is robust. On the other hand, there is no direct evidence demonstrating the validity of the proposed mechanism of inhibition associated with p-S_8. To shed light on this question and to provide the basis for the design of a molecule mimetic to p-S_8, to be used as lead of an eventual drug against AIDS, we study, in this paper, with the help of all-atom simulations in explicit solvent and the novel method of metadynamics combined with parallel tempering: a), the free energy and the equilibrium structure of each of the peptides p-S_2 and p-S_8; b), the details of the docking mechanism of the two peptides and the free energy associated with this process. Whereas p-S_8 is found to be well structured, p-S_2 is rather flexible, wrapping itself around p-S_8 to give rise to the FN, which is stabilized by three particular hydrogen bonds.

INTRODUCTION

The protease of the human immunodeficiency virus type 1 (HIV-1-PR) is a 99-amino-acid-long homodimer (Fig. 1) that plays an essential role in the HIV life cycle by processing the viral polyprotein. During viral replication this protease cuts the long polypeptide producing the components needed for the assembly of the mature virus. For this reason the protease is one of the main targets of acquired immunodeficiency syndrome therapies.

One way to inhibit the cleaving function of HIV-1-PR is to cap its active site. Indeed all the drugs approved so far by FDA rely on this mechanism of competitive inhibition (1). The main disadvantage of these inhibitors is that, under their pressure, HIV-1-PR mutates at the active site or at a site controlling its conformation in such a way that the drug can no longer bind its target and the cleaving function is restored, albeit with reduced efficiency. Early signs of resistance usually appear after 6–8 months from the beginning of the treatment (1).

Folding of the HIV-1-PR in neutral solution is a three-state process (see, e.g., Tiana and Broglia (2)), which means that the two monomers first fold independently ($2U \rightarrow 2N$) and then dock in the dimer native state ($2N \rightarrow N_2$). This scenario is supported by sedimentation equilibrium experiments that demonstrate a consistent population of the native monomeric phase N (3), by NMR analysis of the protease mutated to weaken the interactions across the monomers (4), by all-atom simulations of the monomer in implicit solvent (5), and by Go-model simulations of the dimer (6).

A different strategy of blocking the activity of HIV-1-PR is based on the inhibition of the dimerization (2), relying on the fact that the monomer is inactive, or of the folding of the monomer itself. Broglia et al. (7) pioneered this idea of inhibiting the folding of the protease monomers, rather than capping the active site of the homodimer, exploiting the knowledge of the folding mechanism of single domain proteins.

Several experiments support the idea that a few fragments of these proteins act as folding initiation sites, becoming structured at an early stage in the folding process (for a review, see Fersht (8)). These include hydrogen exchange measurements, which have shown that the folding of cytochrome *c* proceeds by assembling five folding units (9), and nuclear magnetic resonance experiments under denaturating conditions, which have detected partially formed native structure, suggestive of folding initiation sites, in lysozime (10), α-spectrin SH3 (11), staphilococcal nuclease (12), and protein L (13).

Lattice model studies of protein folding done by Shaknovich and co-workers (14) and more recently by Tiana and Broglia (15) have rationalized the scenario of folding of single domain proteins as a hierarchic process occurring in the following order:

1. Local elementary structures (LES) are formed. These are short stretches of the protein containing a few hydrophobic, strongly interacting, and highly conserved (''hot'') amino acids;

Submitted February 9, 2007, and accepted for publication June 8, 2007.

Address reprint requests to Massimiliano Bonomi, Computational Science, Dept. of Chemistry and Applied Biosciences, ETH Zürich, c/o USI Campus, via Buffi 13, CH-6900 Lugano, Switzerland. E-mail: massimiliano.bonomi@phys.chem.ethz.ch.

Editor: Jose Onuchic.

0006-3495/07/10/2813/09 $2.00

doi: 10.1529/biophysj.107.106369

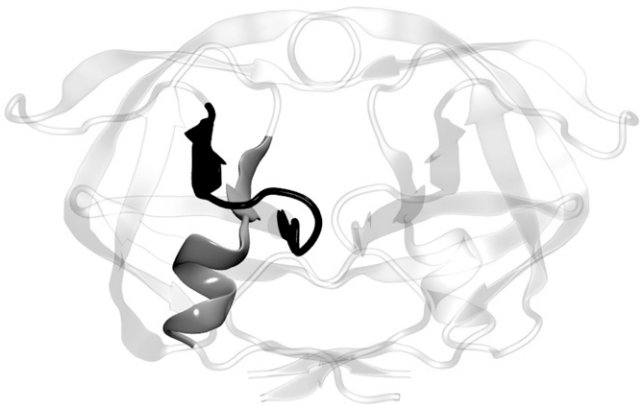

FIGURE 1 The two disjoint identical chains of 99 amino acids that build the HIV-1 protease. Coordinates are taken from the sequence of the HIV-1-PR wild-type (PDB code, 1BVG). The two local elementary structures are colored in black (S_2, the fragment 24–34) and gray (S_8, the fragment 83–93). All the figures in the article are made with the VMD package (54).

2. LES dock into the (postcritical) folding nucleus (FN) of the protein overcoming the major free energy barriers of the folding process;
3. The remaining amino acids relax into the native conformation shortly after the formation of the FN.

Model calculations (16) indicate that knowledge of the LES of a protein can be used to interfere with its folding: short peptides with the same sequence as a local elementary structure (peptidic LES or p-LES) can compete with the native LES in the formation of the FN. The protein finds it entropically favorable to bind the p-LES instead of the native LES because the loss of translational entropy of the p-LES due to the docking is much less than the amount lost in the folding of the whole protein. On the other hand calculations suggest that most of the enthalpic contribution to the folding process is due to the docking of the LES (15,17). The suggestion of Broglia et al. (16) is that the docking of a p-LES with its complementary LES is favored and inhibits the formation of the FN, thus blocking the protein folding and its biological activity.

This type of inhibition is not likely to be affected by mutations of the protein: any mutation of an amino acid that is not hot, i.e., not essential for the stability or docking of LES, would not prevent the p-LES from binding to its complementary LES, whereas a hot mutation would destabilize the protein and make it unstable (18).

The LES of the monomer of HIV-1-PR have been identified (Fig. 1) as the fragments 24–34 (S_2), 75–78 (S_7), and 83–93 (S_8) on the basis of a wide variety of evidence: hierarchy of formation (7), formation temperatures of native contacts in Go-model simulations (19), evolutionary data (20), normal modes (21), ϕ-value analysis (6), study of unfolding penalties (22), and protease mutations in infected patients (23). Go-model simulations of the HIV-1-PR together with p-S_8 have confirmed the inhibitory effect of the small peptide even in the presence of mutations of the protease (7).

In vitro experiments with standard enzymatic assays indicate that p-S_8 inhibits the HIV-1-PR with an inhibition constant $K_I = 2.58 \pm 0.78$ μM. Circular dichroism suggests that the protein denatures upon binding with the p-LES (24). Results on infected cells indicate that the p-LES 83–93 (or better, a more soluble variant of it not containing the last isoleuceine (C-terminal site 93) and denoted BRU in Rusconi et al. (25) and Cicero (26)) is not cytotoxic and inhibits the maturation of the virus at a micromolar concentration.

Results in the literature (7,24–26) indicate that p-S_8 constitutes a promising lead for an eventual drug to fight HIV, unlikely to allow for escape mutants. Obvious disadvantages of the (nonconventional) inhibitor is the value of its inhibition constant (in the micromolar level as compared to nanomolar observed in connection with conventional, active site-centered drugs), its length, its hydrophobicity, and its peptidic character. A clear task lying ahead is to develop molecules mimetic to BRU which, aside from not being recognized by the organism, are more rigid than the original peptide and thus display a stronger affinity with the complementary LES.

Detailed knowledge of the associated docking could prove instrumental in the design of an effective mimetic molecule. For this purpose in this work we have performed all-atom simulations in explicit water on p-S_2, p-S_8, and their complex. In particular, we have gained insight into the folding of p-S_8, the HIV-1-PR folding inhibitor, by reconstructing its free energy surface (FES) using metadynamics. Finally, we have used the combined method of metadynamics and parallel tempering, recently proposed by Bussi et al. (27), to obtain the free energy profile associated with the docking of the two peptides.

METHODS

Metadynamics is a method for efficiently computing free energies and for accelerating rare events (28). It has already been used successfully in the description of a number of different phenomena, including crystal structure prediction (9,40,41), chemical reactions (29–35), protein/peptide complexes, and ligand docking (36–39).

In this method, a set of a few collective variables (CVs) s, which are a function of the system coordinates x, must first be identified as suitable to describe the process of interest: they must be able to distinguish between the initial and final state and they should include the main ''slow'' variable of the system.

The dynamic in the space of the CVs is guided by the free energy of the system plus a history-dependent potential

$$F_G(s,t) = \int_0^t dt' \, W \, exp\left(-\frac{(s - s(x(t')))^2}{2\delta^2}\right), \qquad (1)$$

sum of Gaussians of width δ and weight W centered along the CVs' trajectory up to time t. Metadynamics is able to:

1. Accelerate rare events; the accumulation of the hills pushes the system out of the local minima and drives it to visit new regions of phase space.
2. Provide an unbiased estimate for the free energy surface through the biasing potential; because $\lim_{t\to\infty} -F_G(s,t) \sim F(s)$, an estimate of the FES underlying a region explored by the time t is simply the sum of all the accumulated hills, changed in sign (42).

As already pointed out, metadynamics requires an a priori knowledge of the slow collective variable of the process studied and may fail when reconstructing a free energy surface if one of these slow variables is neglected. To overcome these limitations, metadynamics has recently been combined with parallel tempering (27).

In the parallel tempering method (PT), M replicas of a system of N particles at different temperatures T_i evolve independently until an exchange of configurations between two replicas is attempted (43,44). The partition function of the extended example of noninteracting replicas can be written as:

$$Q = \prod_{i=1}^{M} \frac{q_i}{N!} \int d\mathbf{r}_i^N \, e^{-\beta_i U(\mathbf{r}_i^N)}, \tag{2}$$

where $\mathbf{r}_i^N$ indicates the positions of the N particles of system i at inverse temperature $K_B\beta_i$, U the potential energy, and the factor q_i comes from integrating out the kinetic part of the Hamiltonian.

The probability of a swap between the ensembles j and k is:

$$p(j\to k) = min\{1, (\beta_j - \beta_k)(U(\mathbf{r}_j^N) - U(\mathbf{r}_k^N))\}. \tag{3}$$

High temperature systems are more diffusive and they are able to sample larger areas in phase space by overcoming barriers: by exchanging configurations with replicas at lower temperatures, PT avoids letting these ''cold'' systems be trapped in local energy minima and improves their configurational sampling.

Parallel tempering does not need any a priori knowledge of the system, but on the other hand its ability to overcome large free energy barriers is limited and its scaling with system size is poor.

Combining metadynamics with parallel tempering means adding a bias potential $F_G^{(i)}(s,t)$ to each replica: the partition function (2) becomes

$$Q = \prod_{i=1}^{M} \frac{q_i}{N!} \int d\mathbf{r}_i^N \, e^{-\beta_i[U(\mathbf{r}_i^N) + F_G^{(i)}(s(\mathbf{r}_i^N),t)]}, \tag{4}$$

and the probability of a swap from replica j to k becomes $p(j\to k) = min\{1, \Delta\}$ with:

$$\Delta = (\beta_j - \beta_k)[U(\mathbf{r}_j^N) - U(\mathbf{r}_k^N)] + \beta_j[F_G^{(j)}(s(\mathbf{r}_j^N),t) - F_G^{(j)}(s(\mathbf{r}_k^N),t] + \beta_k[F_G^{(k)}(s(\mathbf{r}_k^N),t) - F_G^{(k)}(s(\mathbf{r}_j^N),t)]. \tag{5}$$

The two methods are to some extent complementary: parallel tempering leads the system to cross moderate energy barriers on all degrees of freedom whereas metadynamics allows high barriers on a few selected collective variables to be overcome. The combination of the two (PTMetaD) helps in solving the problem of ''hidden'' degrees of freedom, i.e., of identifying and including all the slow variables of a physical phenomenon in the CVs of metadynamics. PTMetaD has already been successfully applied to the study of the folding of the 16-residue C-terminal fragment of protein G-B1 (β-hairpin). In the same article it has been shown that, in this case, PTMetaD explores the phase space at least three times faster than standard parallel tempering (27).

Recently, it has also been demonstrated that the combination of metadynamics with umbrella sampling (34,45) could help in improving the accuracy of the free energy landscape calculated with metadynamics alone.

RESULTS AND DISCUSSIONS

In this section we first present the results of our simulations of p-S_8 and p-S_2 in explicit water. These two peptides play a dominant role in the formation of the FN, whereas the small fragment p-S_7, corresponding to residues 75–78 in the sequence of HIV–1–PR, is too short to be specific and most likely only stabilizes further the nucleus already formed by the other two LES (see Broglia et al. (7) and the definition of open LES in Tiana and Broglia (46)). The peptides p-S_2 and p-S_8 behave rather differently, as suggested by preliminary molecular dynamics (MD) simulations. Because the former looks poorly structured, we simply carried out a cluster analysis on a long MD simulation at room temperature. On the other hand, the latter has an α-helix shape that suggests the need to employ metadynamics to cross higher barriers and reconstruct the free energy profile of folding.

At the end we present an analysis of the mechanism of interaction between p-S_2 and p-S_8. To obtain the free energy landscape of this docking process, we have used, because of the advantages described in the Methods section, the combined method of metadynamics and parallel tempering.

Whenever we have reconstructed a free energy surface as a function of a few selected CVs, we have checked the convergence of our calculations and the choice of the CVs by analyzing the distribution of the commitment probability (47).

S_2: the fragment 24–34 of HIV-1-PR

The small hydrophobic peptide p-S_2 is built up of 11 residues: 24–LDTGADDTVLE–34. All the simulations were started from the native geometry and performed in a cubic box of 37.5 Å side containing 1676 water molecules. The simulations were performed using the ORAC code (48) and the OPLS-AA force field, in explicit water (TIP3P model), with multiple time step RESPA integrator (0.5 fs the innermost shell time step) and treating electrostatic with the particle mesh Ewald method.

The thermalization of the system was performed in steps of 200 ps: starting from 0 K and keeping the solute fixed, we gradually heated the system using a Berendsen thermostat and gradually removed the constraints on the peptide. In the last steps of the 400-ps simulation carried out at 300 K, we left the solute free and coupled the system to a Berendsen barostat at 1 atm to complete equilibration.

After thermalization, we ran a preliminary 5-ns-long MD in the NVT ensemble. The peptide turned out to be unstructured. This result was later confirmed by performing a cluster analysis on a 55-ns-long NVT MD trajectory at 300 K. We considered a structure to belong to a cluster when its distance from any element of the cluster was $<$1.5 Å. With this definition we found 38 clusters.

The most populated cluster is characterized by a β-turn shape stabilized by two hydrogen bonds involving the

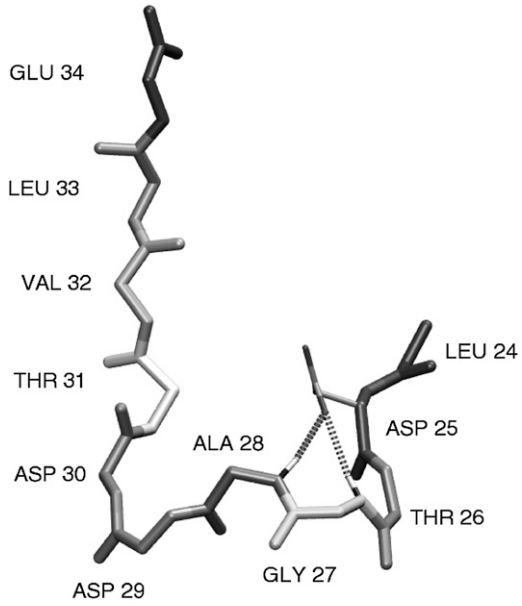

FIGURE 2 A representative of the most populated cluster from the analysis of an NVT trajectory of the peptide p-S_2. In evidence the hydrogen bonds set among Asp-25, Ala-28, and Gly-27.

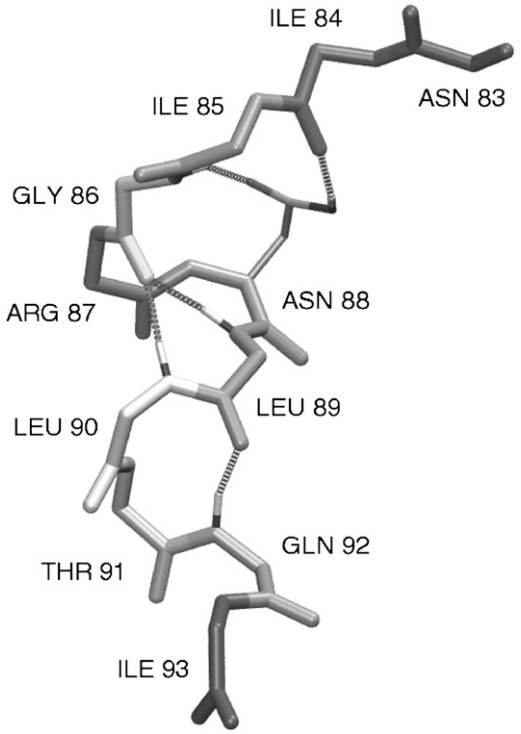

FIGURE 3 The peptide p-S_8, corresponding to the fragment 83–93 of the sequence of HIV-1 protease. In evidence the hydrogen bonds that stabilized the native state structure.

carboxyl oxygen in Asp-25 and NH hydrogens of Ala-28 and Gly-26, as shown in Fig. 2.

S_8: the fragment 83–93 of HIV-1-PR

The highly hydrophobic peptide p-S_8 is made up of 11 residues: 83–NIIGRNLLTQI–93. The initial configuration was taken from the atomic positions of the 83–93 fragment in the native state of the HIV-1-PR (Protein Data Bank code 1BVG (49)). Its native conformation (from N- to C-terminal) is made up of a piece of β-strand (stretch 83–85), an α-helix (87–91) and a turn (92–93). It is stabilized by five hydrogen bonds of which only the one linking the oxygen of the carbonyl group of Gly-86 and the NH hydrogen of Leu-90 is intrabackbone, as shown in Fig. 3. The simulation box is a cubic cell of 37.3 Å lattice and contains 1649 water molecules. The thermalization protocol and all the other details of the simulations are the same as for p-S_2.

To describe the folding of this small peptide we chose two CVs: the radius of gyration (R_G) and the distance from the native contact map (*CMdist*).

The first CV is calculated over the N_α α-carbon of the peptide, whose coordinates are $\vec{r}_i$:

$$R_G = \frac{1}{N_\alpha}\sqrt{\frac{1}{2}\sum_{k=1}^{N_\alpha} \|\vec{r}_i - \vec{r}_j\|^2}. \quad (6)$$

Because it gives a rough measure of the size of the molecule, R_G is able to discriminate between a globular state and an extended conformation.

The contact map is a $N_\alpha \times N_\alpha$ symmetric matrix (M) whose elements are defined as $M_{ij} = f(r_{ij})$, where $r_{ij} = \|\vec{r}_i - \vec{r}_j\|$ is the distance between two C_α atoms and

$$f(r) = \begin{cases} \dfrac{1-\left(\dfrac{r}{r_0}\right)^p}{1-\left(\dfrac{r}{r_0}\right)^q} & r \le r_{cut} \\ 0 & r > r_{cut} \end{cases}, \quad (7)$$

is a switching function. The parameters used in this simulation are: $r_0 = 8.5$ Å, $r_{cut} = 10.5$ Å, $p = 8$ and $q = 10$. The second CV is the distance from the native contact map, i.e., from the contact map of the protein in its native configuration ($\tilde{M}$), defined as the Frobenius distance between the two matrices:

$$CMdist = \|M - \tilde{M}\|_F = \left(\sum_{i,j}^{N_\alpha}(M_{ij} - \tilde{M}_{ij})^2\right)^{1/2}. \quad (8)$$

This collective variable gives a measure of the number of native contacts already formed.

Using these two CVs we have been able to unfold and refold the peptide to a root mean square deviation of 0.3 Å from the native configuration in a 6-ns simulation of pure metadynamics. The free energy surface obtained is shown in Fig. 4. Three regions can be distinguished.

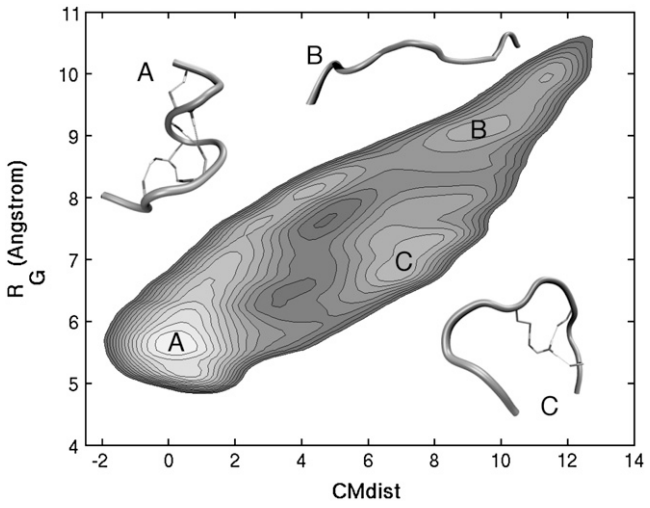

FIGURE 4 Free energy surface of p-S_8 as a function of the distance between the contact map of the native state and of the gyration radius, evaluated only on the C_α. Isoenergy lines are drawn every 1 kcal/mol.

The basin labeled as A ($CMdist \simeq 0, R_G \simeq 5.6$ Å) includes the global minimum of the FES: it describes the native states of the peptide in which the α-helix is formed.

Region B ($CMdist > 8.0$, $R_G > 9.0$ Å) corresponds to the unfolded coil state of the peptide: the α-helix turns are completely extended, all the backbone hydrogen bonds are broken. Basin A is 7 kcal/mol more stable than basin B. The free energy barrier for going from A to B is 9 kcal/mol.

Region C ($6.0 < CMdist < 8.0$, 6.5 Å $< R_G <$ 7.5 Å) corresponds to a local minimum 6 kcal/mol higher in energy than the global one. In these configurations the helix is broken, the peptide has the shape of a U-turn, stabilized by a set of hydrogen bonds between the side chain of Arg-87 and the carbonyl oxygens of Leu-89 and Thr-91.

To assess whether the collective variables chosen provide an accurate characterization of the transition state between the folded and unfolded conformations, we have analyzed the distribution of the associated commitment probability (47). We started from a guess path going from basin A to B and, considering it as a set of N beads connected by springs, we relaxed it on our free energy surface as in a ''nudged elastic band'' (NEB) calculation (50,51), keeping the extremes fixed. In this way we obtained a set of N points i in the collective variables' space along a minimum free energy path going from the folded to the unfolded state of the peptide. The commitment probability P_i^A (or P_i^B) is the probability of reaching basin A (or B) in short MD trajectories starting from a configuration in which the value of CV is in the neighborhood of the point i.

In particular, the transition state ensemble is made up of all the configurations for which $P^A \simeq P^B \simeq 1/2$. In Fig. 5 we have plotted the commitment probability as a function of the position along the minima free energy path from A to B and the relative free energy value according to metadynamic calculations.

Because the point of the path where $P_i^A \simeq P_i^B \simeq 1/2$ is consistent within the accuracy of the calculation with the saddle point in energy, we conclude that our choice of collective variable describes the transition state correctly.

The transition state ensemble is made up of coil configurations with just one of the hydrogen bonds of the native state formed, the one between the oxygen of the carbonyl group of Leu-89 and the NH of Gln-92. This point of the FES is 2 kcal/mol higher in energy than the coil state of the peptide, in agreement with the energy cost of building the first loop of a helix as predicted by the Zimm and Bragg theory (52). The energy difference comes from the balance between the entropic cost of freezing the conformations of three residues and the gain in energy arising from the formation of one hydrogen bond.

Docking of the two p-LES

The two p-LES of the HIV-1-PR are small and quite flexible peptides, especially p-S_2, which is far less structured than p-S_8. They are both very hydrophobic and have opposite charges: p-S_8 has a positive charge (+1) due to arginine, p-S_2 is negatively charged (−4) because of the presence of three aspartic acids and one glutamic acid.

Molecular dynamic simulations of the docked configuration, corresponding to most of the folding nucleus of HIV-1 protease (missing S_7; 75–78), have shown that this structure

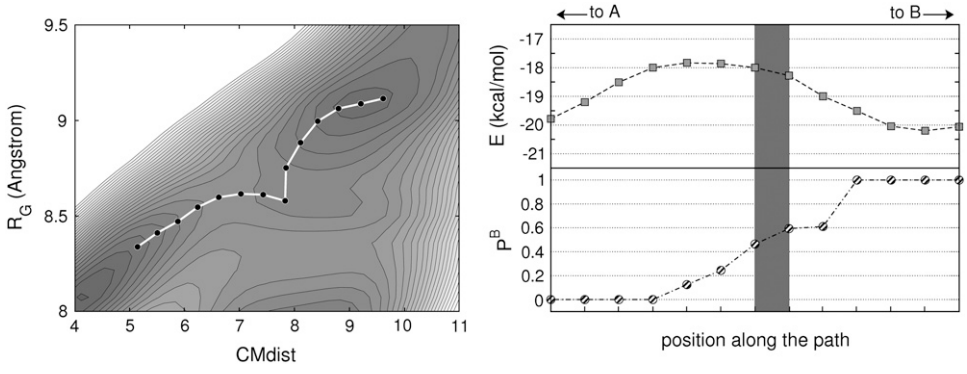

FIGURE 5 Committor analysis. On the left the minima free energy path going from basin A (native state) to basin B (unfolded coil state), as resulting from NEB. Only the portion across the transition state is shown. On the right top panel the free energy value on this portion of the path, as resulting from metadynamics; on the bottom the corresponding commitment probability, calculated as explained in the text. The gray stripe includes the points on the path whose commitment probability is close to 0.5.

is stabilized by the three hydrogen bonds across the peptides shown in Fig. 6. Therefore, the choice of the collective variables for describing the interaction of these two peptides should take into account both the hydrophobic interaction between the peptides, which leads to the elimination of water from the interface, and the formation of the three hydrogen bonds *hb*1, *hb*2, and *hb*3, which eventually stabilize the docked structure.

We used as our first collective variable the distance from the intermolecular contact map of the docked state. It is quite similar to the one used for the study of p-S_8 folding, where instead of the α-carbons we have taken the atoms involved in the formation of the three hydrogen bonds and the switching function parameters have been regulated on the hydrogen bonding distances ($r_0 = 2.5$ Å, $r_{cut} = 3.5$ Å).

This CV ranges from zero, for configurations with *hb*1, *hb*2, and *hb*3 formed, to a maximum value corresponding to a null instantaneous contact map: this maximum must be reached when the peptides are far enough not only for the hydrogen bonds to be broken but also for the water to be inserted in the interface. Therefore, the parameters p and q of the switching function in Eq. 7 must be chosen accordingly.

The second collective variable is the coordination with water molecules of the atoms forming the three selected hydrogen bonds and of a few others involved in the formation of secondary hydrogen bonds during the simulation. The coordination is defined as

$$Z = \sum_{i,j} \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^p}{1 - \left(\frac{r_{ij}}{r_0}\right)^q}, \qquad (9)$$

where i runs over the coordinating atoms, j over all the water molecules, r_{ij} is the distance between atom i and water molecule j, $r_0 = 4.0$ Å, $p = 8$, and $q = 14$ are the parameters of the switching function.

With these collective variables we used the combined method of metadynamics and parallel tempering described in the Methods section. We prepared 64 replicas of the system at different temperatures, starting from 300 to 600 K and distributed in such a way as to result in an homogeneous acceptance probability for the swaps of $\simeq 20\%$.

Exchange between the replicas was attempted every 120 fs, whereas a Gaussian hill of 0.1 kcal/mol height was deposited every 600 fs: after $\simeq 2$ ns of simulation, we obtained the free energy profile shown in Fig. 7. We can distinguish four regions in the free energy surface as a function of these collective variables. Basin A ($CMdist < 0.2$, $8 < Z < 14$) corresponds to the docked configuration, the global minimum where all the three hydrogen bonds *hb*1, *hb*2, and *hb*3 are formed. The second region B ($0.6 < CMdist < 0.8$, $13 < Z < 18$), separated from the former by a barrier of $\simeq 4$ kcal/mol, is a local minimum characterized by the breaking of *hb*1 and water insertion between the atoms involved in this bond.

Configurations where the value of collective variables belongs to basin C ($1.4 < CMdist < 1.7$, $15 < Z < 18$) have two hydrogen bonds broken (*hb*1 and *hb*2): they are separated from region B by a barrier of $\simeq 4$ kcal/mol and from region D by $\simeq 2$ kcal/mol, which correspond to the breaking of the last hydrogen bond *hb*3. Such configurations have all the hydrogen bonds broken and water molecules inserted in between but still the two peptides are not separated by a layer of solvent. This basin is separated by a barrier of $\simeq 4$ kcal/mol from the region where the two LES are fully solvated.

As in the case of the study of p-S_8 folding, we have analyzed the distribution of commitment probability to

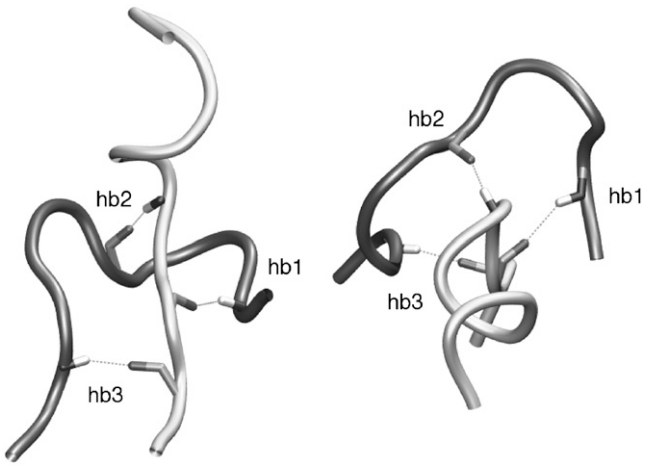

FIGURE 6 The two local elementary structures docked in the folding nucleus of HIV-1 Protease from a side view (*left*) and a top view (*right*). The hydrogen bonds that stabilize the structure are labeled as *hb*1 (Ile-85–Asp-25), *hb*2 (Arg-87–Ala-28), and *hb*3 (Ile-84–Val-32).

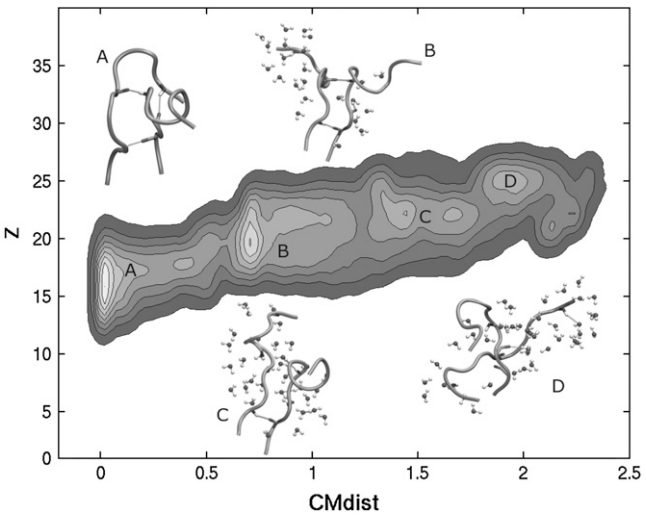

FIGURE 7 Free energy surface of docking between p-S_8 and p-S_2 as a function of the distance from the intermolecular contact map of the docked structure and of the coordination with water of few interfacial atoms. Isoenergy lines are drawn every 1 kcal/mol.

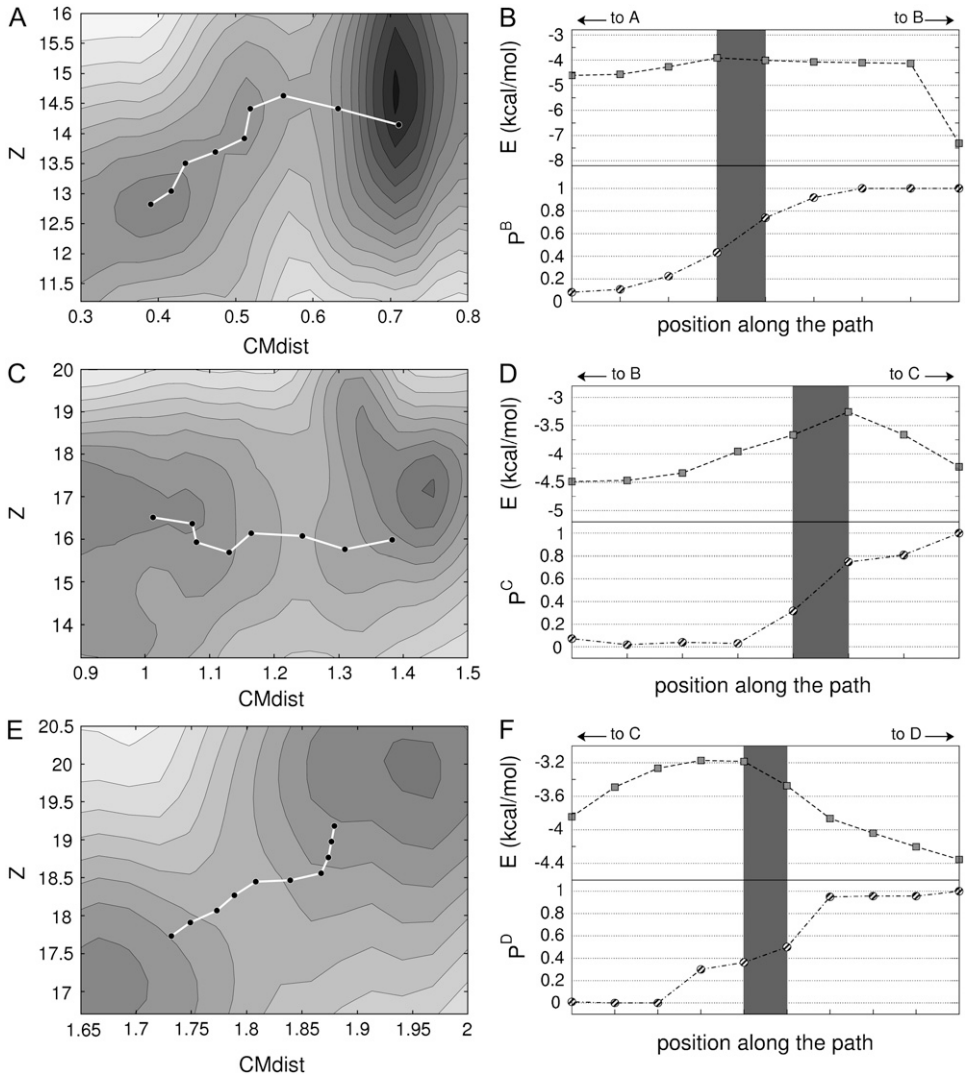

FIGURE 8 Committor analysis. In the first line the transition from basin A to B, in the second from B to C, and in the last line the transition from C to D. On the left panel in each line is the portion of the minima free energy path across the transition state; (*right top panel*) the free energy value on this portion of the path, as resulting from metadynamics, (*bottom*) the corresponding commitment probability. The gray stripe includes the points on the path whose commitment probability is close to 0.5.

verify the correct description of the transition states across different basins. Our choice of collective variables provides a good description of these transition states, as shown in Fig. 8.

Because the residues 88–92 are not involved in the interactions with p-S_2, one could imagine designing a shorter inhibitor, mimicking only the stretch 83–87 of p-S_8. However, taking into account that shorter p-LES are as a rule less specific and effective (16,46), the possibility of shortening this peptide without compromising the interactions with the complementary local elementary structure S_2 remains an open question. An experiment of activity of the shorter peptide, similar to the one reported in Broglia et al. (24), could shed light into this important question and serve as a test of the proposed mechanism of inhibition.

CONCLUSIONS

In this work we have studied the properties of the small peptide p-S_8, which is thought to substitute the LES S_8 in the formation of the folding nucleus of the HIV-1 protease, inhibiting the folding of the monomers and thus of the whole protein.

The fragment p-S_8 turned out to be very stable, with an α-helix shape, its stabilization energy being $\simeq 9$ kcal/mol. This result suggests that this segment of the protease, so stable in isolation, could form independently of the rest of the protein: this is a crucial requirement for folding to be a hierarchical process (53). On the other hand, we have shown that the peptide p-S_2, corresponding to the fragments 24–34 of HIV-1-PR, is much more flexible and substantially unstructured.

We have studied the docking of the two peptides by means of the combined method of metadynamics and parallel tempering, using a measure of the water present in the interface and the formation of hydrogen bonds across the molecules as order parameters. This process does not configure as a docking between two rigid molecules, but rather as a wrapping of a flexible peptide (p-S_2) around a rigid structure (p-S_8).

Although this picture can be somewhat altered in the case in which p-S_8 acts on the whole protein, and the segment S_2 belongs to the HIV-1-PR monomer, pilot calculations of such a situation do not indicate much change from the p-S_2 and p-S_8 situation. We have observed that p-S_2 and p-S_8 recognize each other forming a rather stable complex, a fact that lends strong support to the (postcritical) FN found in Go-model simulations. We have demonstrated that this complex is stabilized by three particular hydrogen bonds, none of them involving the terminal α-helix of p-S_8.

Furthermore, the insight obtained from the results of the calculations reported above is likely to help the design of a synthetic molecule that inhibits the folding of HIV-1-PR by mimicking the activity of the peptide p-S_8.

The authors thank Davide Branduardi and Giovanni Bussi for useful discussions, the Swiss National Supercomputing Centre (CSCS), and ''Enrico Fermi'' Center for providing computational resources.

G.T. acknowledges the financial support of the 2003 Fondo per gli della Ricerca di Base (FIRB) program of the Italian Ministry for University and Scientific Research.

REFERENCES

1. Tomasselli, A. G., and R. L. Heinrikson. 2000. Targeting the HIV-protease in AIDS therapy: a current clinical perspective. *Biochim. Biophys. Acta.* 1477:189–214.
2. Tiana, G., and R. A. Broglia. 2002. Folding and design of dimeric proteins. *Proteins.* 49:82–94.
3. Xie, D., S. Gulnik, E. Gustchina, B. Yu, W. Shao, W. Qoroneh, A. Nathan, and J. W. Erickson. 1999. Drug resistance mutations can affect dimer stability of HIV-1 protease at neutral pH. *Protein Sci.* 8:1702–1713.
4. Ishima, R., R. Ghirlando, J. Todzser, A. M. Gronenborn, D. A. Torchia, and J. M. Louis. 2001. Folded monomer of HIV-1 protease. *J. Biol. Chem.* 276:49110–49116.
5. Levy, Y., and A. Caflish. 2003. The flexibility of monomeric and dimeric HIV-1 PR. *J. Phys. Chem. B.* 107:3068–3079.
6. Levy, Y., A. Caflish, J. N. Onuchic, and P. G. Wolynes. 2004. The folding and dimerization of HIV-1 protease: evidence for a stable monomer from simulations. *J. Mol. Biol.* 340:67–79.
7. Broglia, R. A., G. Tiana, L. Sutto, D. Provasi, and F. Simona. 2005. Design of HIV-1-PR inhibitors that do not create resistance: blocking the folding of single monomers. *Protein Sci.* 14:2668–2681.
8. Fersht, A. R. 1997. Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* 7:3–9.
9. Martonak, R., A. Laio, M. Bernasconi, C. Ceriani, P. Raiteri, F. Zipoli, and M. Parrinello. 2005. Simulation of structural phase transitions by metadynamics. *Zeitschrift fur Kristallographie.* 220:489–498.
10. Miranker, A., and S. E. Radford. 1991. Demonstration by NMR of folding domains in lysozyme. *Nature.* 349:633–636.
11. Kortemme, T., M. J. S. Kelly, L. E. Kay, J. Forman-Key, and L. Serrano. 2000. Similarities between the spectrin SH3 domain denatured state and its folding transition state. *J. Mol. Biol.* 297:1217–1229.
12. Shortle, D., and M. S. Ackerman. 2001. Persistence of native-like topology in a denatured protein in 8 M urea. *Science.* 293:487–489.
13. Yi, Q., M. L. Scalley-Kim, E. J. Alm, and D. Baker. 2000. NMR characterization of residual structure in the denatured state of protein L. *J. Mol. Biol.* 299:1341–1351.
14. Abkevich, V. I., A. M. Gutin, and E. I. Shakhnovich. 1994. Specific nucleus as the transition state for protein folding. *Biochemistry.* 33: 10026–10036.
15. Tiana, G., and R. A. Broglia. 2001. Statistical analysis of native contact formation in the folding of designed model proteins. *J. Chem. Phys.* 114:2503–2507.
16. Broglia, R. A., G. Tiana, and R. Berera. 2003. Resistance proof, folding-inhibitor drugs. *J. Chem. Phys.* 118:4754–4758.
17. Broglia, R. A., and G. Tiana. 2001. Hierarchy of events in the folding of model proteins. *J. Chem. Phys.* 114:7267–7273.
18. Tiana, G., R. A. Broglia, H. E. Roman, E. Vigezzi, and E. I. Shakhnovich. 1998. Folding and misfolding of designed protein-like chains with mutations. *J. Chem. Phys.* 108:757–761.
19. Cecconi, F., C. Micheletti, P. Carloni, and A. Maritan. 2001. Molecular dynamics studies on HIV-1 protease drug resistance and folding pathways. *Proteins.* 43:365–372.
20. Holm, L., and C. Sander. 1996. Mapping the protein universe. *Science.* 273:595–602.
21. Bahar, I., A. R. Atilgan, M. C. Demirel, and B. Erman. 1998. Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. *Phys. Rev. Lett.* 80:2733–2736.
22. Wallqvist, A., G. W. Smythers, and D. G. Covell. 1998. A cooperative folding unit in HIV-1 protease. Implications for protein stability and occurrence of drug-induced mutations. *Protein Eng.* 11:999–1005.
23. Shafer, R. W., P. Hsu, A. K. Patick, C. Craig, and V. Brendel. 1999. Identification of biased amino acid substitution patterns in human immunodeficiency virus type 1 isolates from patients treated with protease inhibitors. *J. Virol.* 73:6197–6202.
24. Broglia, R. A., D. Provasi, F. Vasile, G. Ottolina, R. Longhi, and G. Tiana. 2006. A folding inhibitor of the HIV-1 protease. *Proteins.* 62:928–933.
25. Rusconi, S., M. L. Cicero, A. E. Laface, S. Ferramosca, F. Siriani, E. Cesana, D. Provasi, G. Tiana, M. Galli, M. Moroni, A. Clivio, and R. A. Broglia. 2007. Susceptibility to a non-conventional (folding) protease inhibitor of human immunodeficiency virus type 1 isolates in vitro. *In* 14th Conference on Retroviruses and Opportunistic Infections (CROI). Los Angeles, CA. Foundation for Retrovirology and Human Health, Alexandria, VA.
26. Cicero, M. L. 2007. Efficacia *in vitro* di un piccolo peptide nell'inibizione del folding conformazionale della proteasi di HIV-1. PhD thesis. University of Milano, Milan, Italy.
27. Bussi, G., F. L. Gervasio, A. Laio, and M. Parrinello. 2006. Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.* 128:13435–13441.
28. Laio, A., and M. Parrinello. 2002. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA.* 99:12562–12566.
29. Iannuzzi, M., A. Laio, and M. Parrinello. 2003. Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics. *Phys. Rev. Lett.* 90:238302.
30. Stirling, A., M. Iannuzzi, A. Laio, and M. Parrinello. 2004. Azulene-to-naphthalene rearrangement: the Car-Parrinello metadynamics method explores various reaction mechanisms. *ChemPhysChem.* 5:1558–1568.
31. Churakov, S., M. Iannuzzi, and M. Parrinello. 2004. Ab initio study of dehydroxylation-carbonation reaction on brucite surface. *J. Phys. Chem. B.* 108:11567–11574.
32. Gervasio, F. L., A. Laio, M. Iannuzzi, and M. Parrinello. 2004. Influence of DNA structure on the reactivity of the guanine radical cation. *Chem. Eur. J.* 10:4846–4852.
33. Ensing, B., A. Laio, F. L. Gervasio, M. Parrinello, and M. L. Klein. 2004. A minimum free energy reaction path for the E2 reaction between fluoro ethane and a fluoride ion. *J. Am. Chem. Soc.* 126:9492–9493.
34. Ensing, B., and M. L. Klein. 2005. Chemical theory and computation special feature: perspective on the reactions between F^- and CH_3CH_2F: the free energy landscape of the $E2$ and S_N2 reaction channels. *Proc. Natl. Acad. Sci. USA.* 102:6755–6759.
35. Wu, Y., J. Schmitt, and R. Car. 2004. Mapping potential energy surfaces. *J. Chem. Phys.* 121:1193–1200.
36. Ceccarelli, M., C. Danelon, A. Laio, and M. Parrinello. 2004. Microscopic mechanism of antibiotics translocation through a porin. *Biophys. J.* 87:58–64.

37. Gervasio, F. L., A. Laio, and M. Parrinello. 2005. Flexible docking in solution using metadynamics. *J. Am. Chem. Soc.* 127:2600–2607.

38. Branduardi, D., F. L. Gervasio, A. Cavalli, M. Recantini, and M. Parrinello. 2005. The role of the peripheral anionic site and cation-interactions in the ligand penetration of the human AChE gorge. *J. Am. Chem. Soc.* 127:9147–9155.

39. Fiorin, G., A. Pastore, P. Carloni, and M. Parrinello. 2006. Using metadynamics to understand the mechanism of calmodulin/target recognition at atomic detail. *Biophys. J.* 91:2768–2777.

40. Martonak, R., A. Laio, and M. Parrinello. 2003. A molecular dynamics method for simulations in the canonical ensemble. *Phys. Rev. Lett.* 90: 75503.

41. Raiteri, P., R. Martonak, and M. Parrinello. 2005. Exploring polymorphism: the case of benzene. *Angew. Chem. Int. Ed. Engl.* 44:3769–3773.

42. Bussi, G., A. Laio, and M. Parrinello. 2006. Equilibrium free energies from nonequilibrium metadynamics. *Phys. Rev. Lett.* 96:90601.

43. Hansmann, U. H. E. 1997. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* 281:140–150.

44. Sugita, Y., and Y. Okamoto. 1999. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314:141–151.

45. Babin, V., C. Roland, T. A. Darden, and C. Sagui. 2006. The free energy landscape of small peptides as obtained from metadynamics with umbrella sampling corrections. *J. Chem. Phys.* 125:204909.

46. Tiana, G., and R. A. Broglia. 2001. Reading the three-dimensional structure of a protein from its amino acid sequence. *Proteins.* 45:421–427.

47. Geissler, P. L., C. Dellago, and D. Chandler. 1999. Kinetic pathways of ion pair dissociation in water. *J. Phys. Chem. B.* 103:3706–3710.

48. Procacci, P., E. Paci, T. Darden, and M. Marchi. 1997. ORAC: a molecular dynamics program to simulate complex molecular systems with realistic electrostatic interactions. *J. Comp. Chem.* 18:1848–1862.

49. Yamazaki, T., A. P. Hinck, Y. X. Wang, L. K. Nicholson, D. A. Torchia, P. Wingfield, S. J. Stahl, J. D. Kaufman, C. H. Chang, P. J. Domaille, and P. Y. S. Lam. 1996. Three-dimensional solution structure of the HIV-1 protease complexed with DMP323, a novel cyclic urea-type inhibitor, determined by nuclear magnetic resonance spectroscopy. *Protein Sci.* 5:495–506.

50. Jonsson, H., G. Mills, and K. W. Jacobsen. 1998. Nudged elastic band method for finding minimum energy paths of transitions. *In* Classical and Quantum Dynamics in Condensed Phase Simulations. World Scientific, Singapore. 385–404.

51. Henkelman, G., and H. Jonsson. 2000. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* 113:9978–9985.

52. Zimm, B. H., and J. K. Bragg. 1959. Theory of the phase transition between helix and random coil in polypeptide chains. *J. Chem. Phys.* 31:526–535.

53. Fersht, A. 1998. Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding. W. H. Freeman and Company, New York, NY.

54. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD – Visual Molecular Dynamics. *J. Mol. Graph.* 14:33–38.