

# Multiple Routes and Milestones in the Folding of HIV-1 Protease Monomer

Massimiliano Bonomi<sup>1\*</sup>, Alessandro Barducci<sup>1</sup>, Francesco L. Gervasio<sup>2</sup>, Michele Parrinello<sup>1</sup>

**1** Computational Science, Department of Chemistry and Applied Biosciences, ETH Zurich, Lugano, Switzerland, **2** Computational Biophysics Group, Structural Biology and Biocomputing Programme, Spanish National Research Centre, Madrid, Spain

## Abstract

Proteins fold on a time scale incompatible with a mechanism of random search in conformational space thus indicating that somehow they are guided to the native state through a funneled energetic landscape. At the same time the heterogeneous kinetics suggests the existence of several different folding routes. Here we propose a scenario for the folding mechanism of the monomer of HIV-1 protease in which multiple pathways and milestone events coexist. A variety of computational approaches supports this picture. These include very long all-atom molecular dynamics simulations in explicit solvent, an analysis of the network of clusters found in multiple high-temperature unfolding simulations and a complete characterization of free-energy surfaces carried out using a structure-based potential at atomistic resolution and a combination of metadynamics and parallel tempering. Our results confirm that the monomer in solution is stable toward unfolding and show that at least two unfolding pathways exist. In our scenario, the formation of a hydrophobic core is a milestone in the folding process which must occur along all the routes that lead this protein towards its native state. Furthermore, the ensemble of folding pathways proposed here substantiates a rational drug design strategy based on inhibiting the folding of HIV-1 protease.

**Citation:** Bonomi M, Barducci A, Gervasio FL, Parrinello M (2010) Multiple Routes and Milestones in the Folding of HIV-1 Protease Monomer. PLoS ONE 5(10): e13208. doi:10.1371/journal.pone.0013208

**Editor:** Markus J. Buehler, Massachusetts Institute of Technology, United States of America

**Received:** August 5, 2010; **Accepted:** September 11, 2010; **Published:** October 13, 2010

**Copyright:** © 2010 Bonomi et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The authors have no support or funding to report.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: mbonomi@ethz.ch

## Introduction

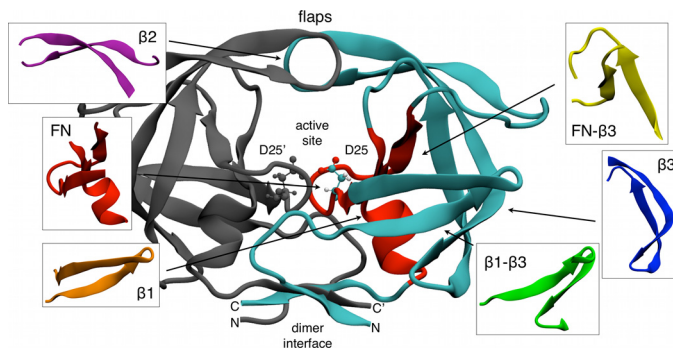
The protease of Human Immunodeficiency Virus type 1 (HIV-1-PR) is a dimer in its catalytic competent form (Fig. 1). Each of the two identical monomers has a single domain composed of 99 amino acids. Several experimental [1–3] and computational studies [4,5] suggest that the folding of this enzyme is a three-state process in which first two monomers fold independently and then dock in the dimer native state. Studying the folding of the HIV-1-PR monomer is therefore the first step in the comprehension of the whole enzyme formation.

Understanding protein folding at an atomic resolution is a fundamental yet challenging task for both experiments and simulations. In the case of HIV-1-PR, characterizing the ensemble of folding pathways would also provide precious information for the rational design of novel anti-HIV drugs. HIV-1-PR is one of the main targets of Acquired Immuno-Deficiency Syndrome therapies as it performs an essential function in the HIV life cycle by cleaving the viral poly-protein and producing the components that are needed for the mature virus assembly. The virus very high mutation rate is capable of eluding the effects of competitive inhibition based drugs in a very short time [6]. An alternative strategy for neutralizing the HIV-1-PR function consists in inhibiting the formation of the protease by interfering either with the folding process of the monomer or with its dimerization [7,8]. Several lines of evidence show that the unfolded protein, the monomer and the dimer although separated by large barriers have comparable energies [3]. Furthermore, it

has been suggested that the dimer is stabilized by the substrate. In such a scenario it would be more advantageous to target the monomer folding.

In light of these considerations, a deep understanding of the monomer folding mechanism would have broader significance. To this end many theoretical and experimental studies have been performed [9–11]. Among the first group, a possible scenario has been suggested by off-lattice  $G\bar{O}$  models simulations in which first the fragments 24–34 ( $S_2$ ) and 83–93 ( $S_8$ ) fold to form the so-called Local Elementary Structures (LES) and subsequently these LES dock in the folding nucleus (FN) [12,13]. Recently we have shown by fully atomistic molecular dynamics (MD) and metadynamics [14] simulations the stability of the LES in solution and calculated the strength of their interaction, adding further evidence to their central role in monomer folding [15]. It has also been proposed that a peptide mimicking the sequence of one of the two LES (p- $S_8$ ) could be used as efficient folding inhibitor [12]. Remarkably, these theoretical predictions have been confirmed by *in-vitro* and *ex-vivo* experiments [16,17]. Standard enzymatic assays indicate that p- $S_8$  inhibits the HIV-1-PR with an inhibition constant  $K_I = 2.58 \pm 0.78 \mu\text{M}$  [16], while results on infected cells indicate that this peptide is not cytotoxic and inhibits the maturation of the virus at a micromolar concentration [17]. These successes further increase the need for a thorough investigation of the folding intermediates and of all the possible folding pathways.

Here we use state-of-the-art computational techniques to formulate a plausible scenario for the folding mechanism of HIV-1-PR



**Figure 1. Structure of the HIV-1-PR dimer (PDB code 1BVG).** The active site is situated at the interface between the two identical subunits of the homodimer. It is characterized by the sequence Asp-Thr-Gly (Asp25, Thr26 and Gly27), common to aspartic proteases. The two Asp25 residues act as the catalytic residues. At the top, in evidence the flexible flaps region. In the small boxes, the six main motifs in which we classified the native contact map in Fig. 2.

doi:10.1371/journal.pone.0013208.g001

monomer. First we performed 0.5  $\mu$ s long all-atom MD simulations in explicit solvent at room temperature. Then we run multiple high-temperature unfolding simulations, clustering the structures visited and analyzing the network of transitions between clusters. Finally, we used a recently developed structure-based potential at atomistic resolution [18] together with a combined parallel tempering [19] and metadynamics [20] (PTMetaD) [21] to obtain well converged multidimensional free-energy surfaces (FES). Especially effective has proven the application of the very recently developed well-tempered ensemble (WTE) [22].

The all-atom simulations and structure-based potential FES revealed a complex scenario respectively for the high-temperature unfolding and folding mechanism of this protein. Both processes were characterized by the simultaneous presence of multiple pathways and milestone events. In either cases, heterogeneity could be ascribed to the behavior of particular  $\beta$ -hairpin subunits, while the milestone events corresponded to the disruption/formation of an extended folding nucleus composed by the two LES  $S_8$  and  $S_2$  plus another hydrophobic fragment (residues 73–80). This remarkable agreement between the nature and the fine details of the high-temperature unfolding process and folding mechanism prompted us to formulate valuable hypothesis about the actual folding routes of the HIV-1 monomer.

These results obtained here for the HIV-1-PR monomer might have a more general valence. In fact one finds that folding is guided by a milestone event which occurs rather rapidly reducing the conformational space that needs to be sampled. On the other end, the heterogeneous nature of the overall process is consistent with a body of experimental evidence on single domain protein folding.

## Results

While models of different complexity were used, a common description of the protein in terms of native contacts was adopted throughout all this study (see *Methods* section). In Fig. 2 we show the contact map of HIV-1-PR in its native state. After extensive analysis of both structure and dynamics, we classified these contacts into six main groups. The three  $\beta$ -hairpin structures are labeled as  $\beta 1$  (residues 10–23),  $\beta 2$  (residues 41–58) and  $\beta 3$  (residues 55–75). We also label different set of interactions, namely  $\beta 1$ - $\beta 3$  stands for the interaction between  $\beta 1$  and  $\beta 3$ , while FN refers to the interaction between fragments 24–34 and 83–92. This group corresponds to the folding nucleus [12,15]. Finally the interaction FN- $\beta 3$  will be referred to below. The fragment 83–93,

which is a very stable  $\alpha$ -helix [15], was not studied as a separate entity but in relation to its role in the FN structure.

In the following sections we analyze separately the results of our all-atom and coarse-grained simulations. In the *Discussion*, on the basis on these results we propose a plausible scenario for the folding mechanism of the monomer.

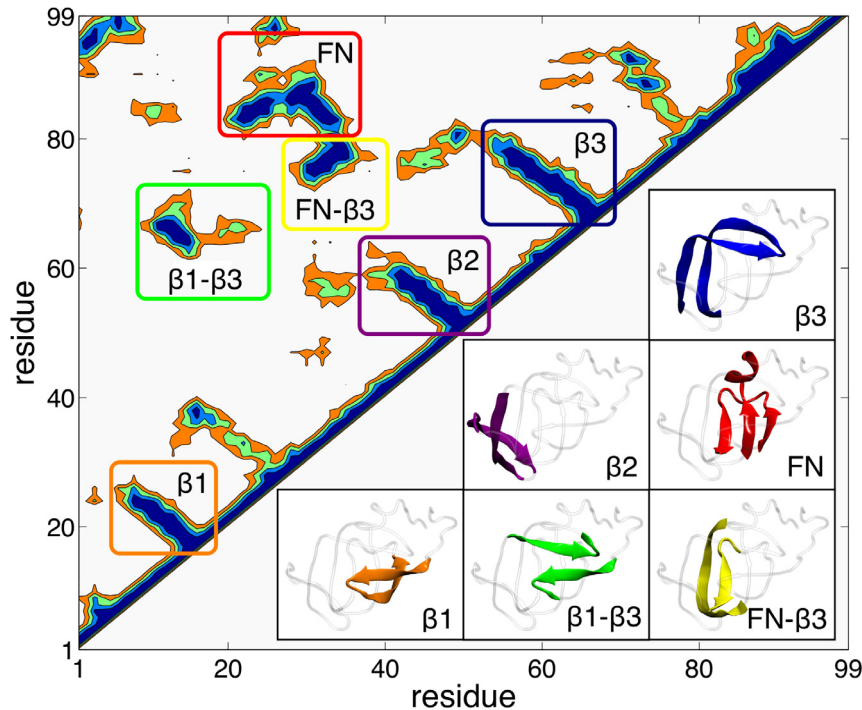
## All-atom simulations

Starting from the crystallographic structure, we performed a 512 ns long NVT simulation of HIV-1-PR monomer at 300K using an all-atom representation for the protein and for the solvent degrees of freedom. The simulation was performed to confirm the stability of the native state of the protein and study the relative fluctuations of the various motifs of the fold. Details of the simulation protocol are reported in the *Methods* section. The structure of the monomer was stable during the whole NVT simulation. The root mean square fluctuations of the  $C_\alpha$  atoms were within 0.5 and 1.5  $\text{\AA}$  along most of the chain (Figure S1). Three regions displayed a greater mobility: the C and N termini and the  $\beta 2$  fragment.

The terminal regions displayed fluctuations of the order of 3 to 4  $\text{\AA}$ . These values are much smaller than those reported in Ref. [5]. There the authors found values ranging from 5 to 10  $\text{\AA}$  in a much shorter (5 ns) simulation. The difference can be explained either by the different force field used, AMBER99SB here and CHARMM in Ref. [5], or by the longer time-scale explored by our simulation. We incline to the second hypothesis. After several tens of nanoseconds we observed a partial assembly of the N and C termini into a  $\beta$ -sheet structure (Figure S2). This rearrangement most likely made the structure more rigid. What is more, the formation of such  $\beta$  structure was also observed in a longer simulation that used the GROMOS force field [23], adding weight to our suggestion.

The flexibility of  $\beta 2$  is not surprising. This motif corresponds in the dimer structure to one of the two flaps (Fig. 1), a region that has been shown both experimentally [24,25] and computationally [26–28] to be extremely flexible. This flexibility is indeed functional to the enzymatic activity. The structures we focused on remained stable during the simulation with the exception of  $\beta 1$ - $\beta 3$  which underwent a fluctuation in which this interaction was broken and reformed in about 10 ns (Figure S2 and Figure S3). The most stable behavior was instead shown by  $\beta 1$ , which exhibited the smallest fluctuations.

We then performed multiple high-temperature unfolding simulations starting from the native conformation. It has been



**Figure 2. Contact map of HIV-1-PR in its native state.** We classified the contacts in six groups: in orange  $\beta 1$  (residues 10–23), in violet  $\beta 2$  (residues 41–58), in blue  $\beta 3$  (residues 55–75), in red FN (residues 24–34 and 83–92), in green  $\beta 1$ – $\beta 3$  and in yellow FN– $\beta 3$ . doi:10.1371/journal.pone.0013208.g002

suggested that high-temperature unfolding MD simulations can be used to formulate useful hypothesis for experimental studies on folding [29–31].

The configurations visited in the unfolding simulations were clustered on the basis of the number of native contacts formed using the k-means algorithm [32]. Details of the unfolding simulations, the clustering algorithm and the network analysis can be found in the *Methods* section.

The analysis of the clusters network reveals the existence of at least two distinct unfolding pathways (Fig. 3). The main difference between the two routes is determined by  $\beta 1$  (Fig. 3, panel  $\beta 1$ ). In the most populated pathway, this hairpin is the last secondary structure to unfold, while in the other pathway it unfolds in the early stages. The folding of  $\beta 2$  (Fig. 3, panel  $\beta 2$ ) is uncorrelated to the overall unfolding of the protein. Folded and unfolded  $\beta 2$  configurations can be found at different stages along the two main pathways. The hairpin  $\beta 3$  (Fig. 3, panel  $\beta 3$ ) unfolds later than the other motifs in both the dominant pathways, only when almost 70% of the protein is unfolded. The interaction  $\beta 1$ – $\beta 3$  is quite weak and breaks at the early stages of unfolding (Fig. 3, panel  $\beta 1$ – $\beta 3$ ). If we take this information together with that of panels  $\beta 1$  and  $\beta 3$ , we can conclude that in the main pathway, when the contacts between the two hairpins are broken, each of the two  $\beta$ -strands remains folded. In the alternative route, the inter-hairpin contacts are lost almost in sync with  $\beta 1$  unfolding, while  $\beta 3$  remains structured for longer times. The sets of contacts FN and FN– $\beta 3$ , which involve a large number of hydrophobic residues buried inside the enzyme, are the last to unfold (Fig. 3, panel HYDRO).

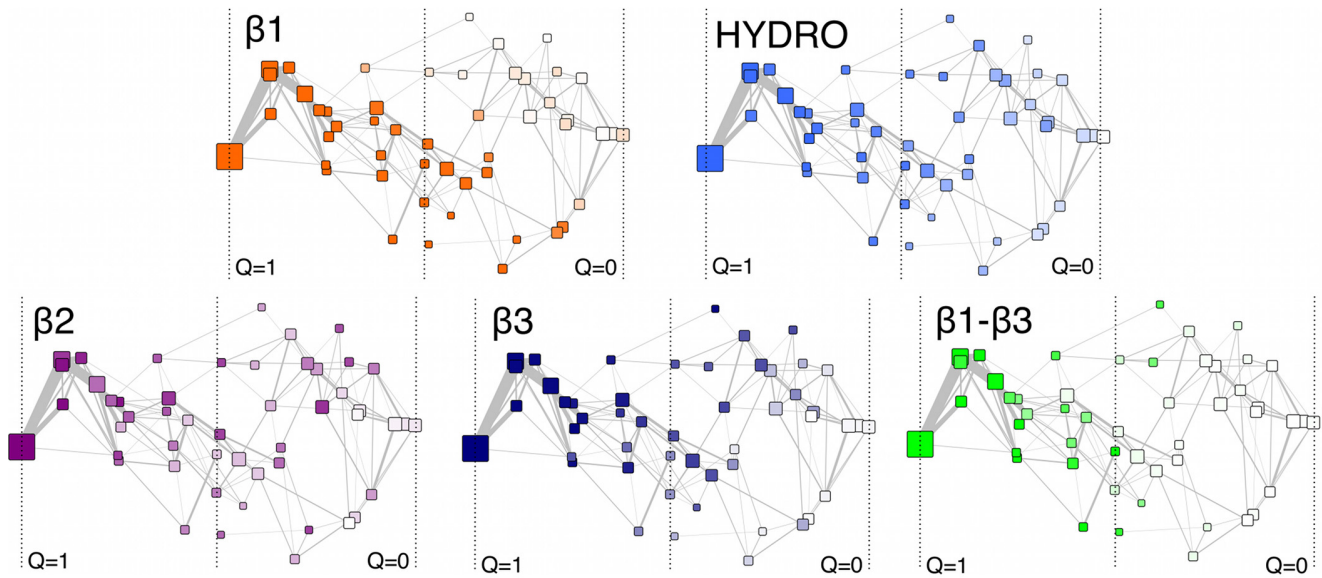
### Structure-based potential simulations

The calculation of all-atom folding FES in explicit solvent is still impractical for this system, despite the increasing computational power [33] and the variety of enhanced sampling methods

available [34,35]. Not only for this reason but also to have an independent view on the folding mechanism, we chose to adopt a simplified potential energy function that is consistent with our representation of the protein in terms of native contacts. We used the structure-based potential recently introduced by Whitford *et al.* [18]. Despite its simplified nature, this potential has been shown to predict the folding mechanism of the B domain of Protein A, of the SH3 domain of C-Src Kinase and of Chymotrypsin Inhibitor 2, in agreement with other  $C_\alpha$  G $\bar{O}$  models and all-atom force fields. The success of this potential is probably connected to the fact that evolution has led proteins to display funneled energy landscapes with small degrees of ruggedness. This means that evolution has optimized the protein sequence in order to ensure a robust folding such that the native state does not have to compete with denatured conformations [36–40]. Thus a model based on the topology of the native state can be very effective in predicting the folding mechanism of proteins. What is more, a funneled landscape does not preclude the presence of multiple kinetically relevant folding routes, as it has been shown in various studies [41–44].

In the spirit of Ref. [45], in order to understand the folding mechanism we calculated multiple two-dimensional (2-D) FES as a function of the fraction of native contacts of our six motifs and at several temperatures across the model folding temperature ( $T_f \approx 113.5\text{K}$ ). To this effect we used PTMetaD boosted by WTE [22] and in combination with the reweighting technique of Ref. [46] (see *Methods* section).

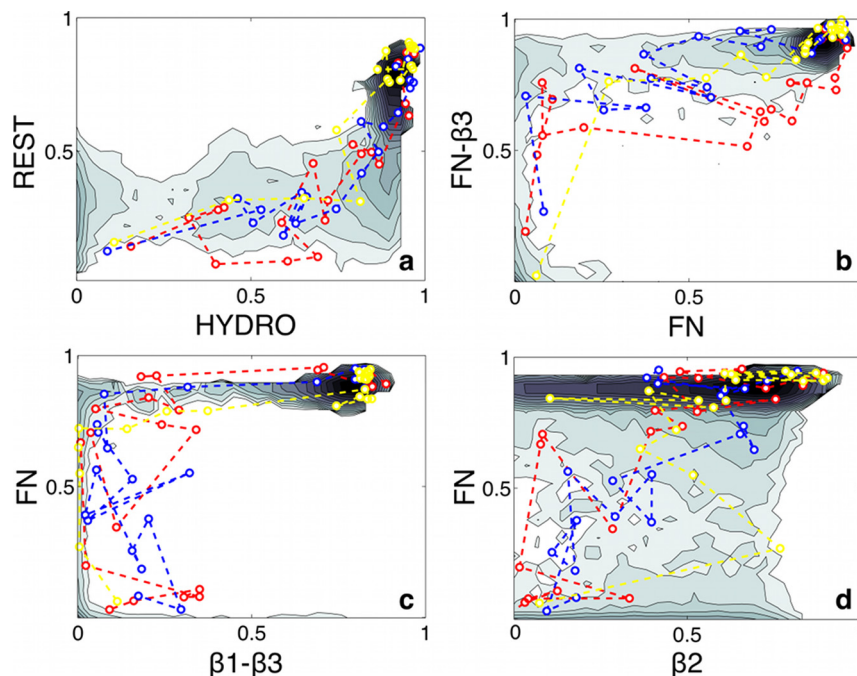
Before analyzing the 2-D FES, let us discuss briefly the stability of the three  $\beta$ -strand structures (Figure S6). Throughout the range of all temperatures,  $\beta 1$  appears the most stable among the  $\beta$ -strands. For temperatures lower than  $T_f$ , the FES of the hairpin  $\beta 1$  and  $\beta 3$  display bimodal distributions corresponding to folded and unfolded states that are sharper than  $\beta 2$ . This suggests that, even at low temperature,  $\beta 2$  is more flexible than the other hairpin-like subunits of the monomer.



**Figure 3. Network of clusters from all-atom high-temperature unfolding trajectories.** Clusters are positioned according to the metric multidimensional scaling algorithm [59]. The size of the area of each cluster is proportional to its population, the link size to the number of trajectories connecting two clusters.  $Q$  is the total number of contacts. From left ( $Q=1$ ) to right ( $Q=0$ ), clusters contain conformations more and more unfolded. The vertical dotted lines are drawn at  $Q=1$ ,  $Q=0.5$  and  $Q=0$  as a guide for the eye. In each panel, clusters are colored with an intensity which is (linearly) proportional to the number of contacts that belong to a certain motif and that are still formed at that particular stage of the overall unfolding of the protein. The set HYDRO is composed by the group of contacts FN and FN- $\beta 3$ , which were grouped together since they displayed a similar behavior along the unfolding pathways. doi:10.1371/journal.pone.0013208.g003

The set of 2-D FES provides a clear explanation for the sequence of events that characterize the folding mechanism at the different temperatures. The hydrophobic core of the monomer composed by the set of contacts FN and FN- $\beta 3$  is the first structure formed during

the folding process. In fact, if we analyze the FES as a function of the hydrophobic contacts and all the other variables (Fig. 4a), we see a clear L-shaped landscape. This indicates that first the hydrophobic collapse takes place and only after the rest of the structure is formed.



**Figure 4. FES as a function of the fraction of native contacts of the six motifs at  $T = 0.969$  obtained by reweighting the structure-based potential PTMetaD simulation.** a. HYDRO vs. all the other contacts. b. FN vs. FN- $\beta 3$ . c.  $\beta 1-\beta 3$  vs. FN. d.  $\beta 2$  vs. FN. Isoenergy lines are drawn every  $k_B T$ . In dashed lines, projection of the relevant part of the all-atom explicit-solvent unfolding simulations at 700K (blue and red) and 500K (yellow). doi:10.1371/journal.pone.0013208.g004

The L-shape, which is very sharp at low temperature, becomes less definite as the temperature increases, but it is still clearly recognizable (Figure S7). This is a further proof that the hydrophobic collapse is a fundamental milestone in the folding process. If we analyze the single motifs contributing to the hydrophobic core, we notice that the contacts FN- $\beta$ 3 are formed before FN (Fig. 4b) independently of temperature (Figure S8).

The remaining steps of the folding process are more complex. The contacts  $\beta$ 1- $\beta$ 3 are formed after the hydrophobic collapse (Fig. 4c). When this happens,  $\beta$ 1 and  $\beta$ 3 can be either folded or unfolded, depending on T (Fig. 5). At low T,  $\beta$ 1 and  $\beta$ 3 are formed independently and only after they dock forming the contacts  $\beta$ 1- $\beta$ 3. At higher temperatures, the sequence of events can vary and  $\beta$ 1- $\beta$ 3 tends to be formed before  $\beta$ 1 and  $\beta$ 3 are fully structured.

Lastly, we discuss  $\beta$ 2 whose behavior is less correlated with the overall folding process. If we examine the 2-D FES involving  $\beta$ 2, we notice that this part of the protein can be folded independently of the other subunits. For instance, in the FES as function of  $\beta$ 2 and FN (Fig. 4d) the shallow basins correspond to  $\beta$ 2 folded whatever FN is formed or not.

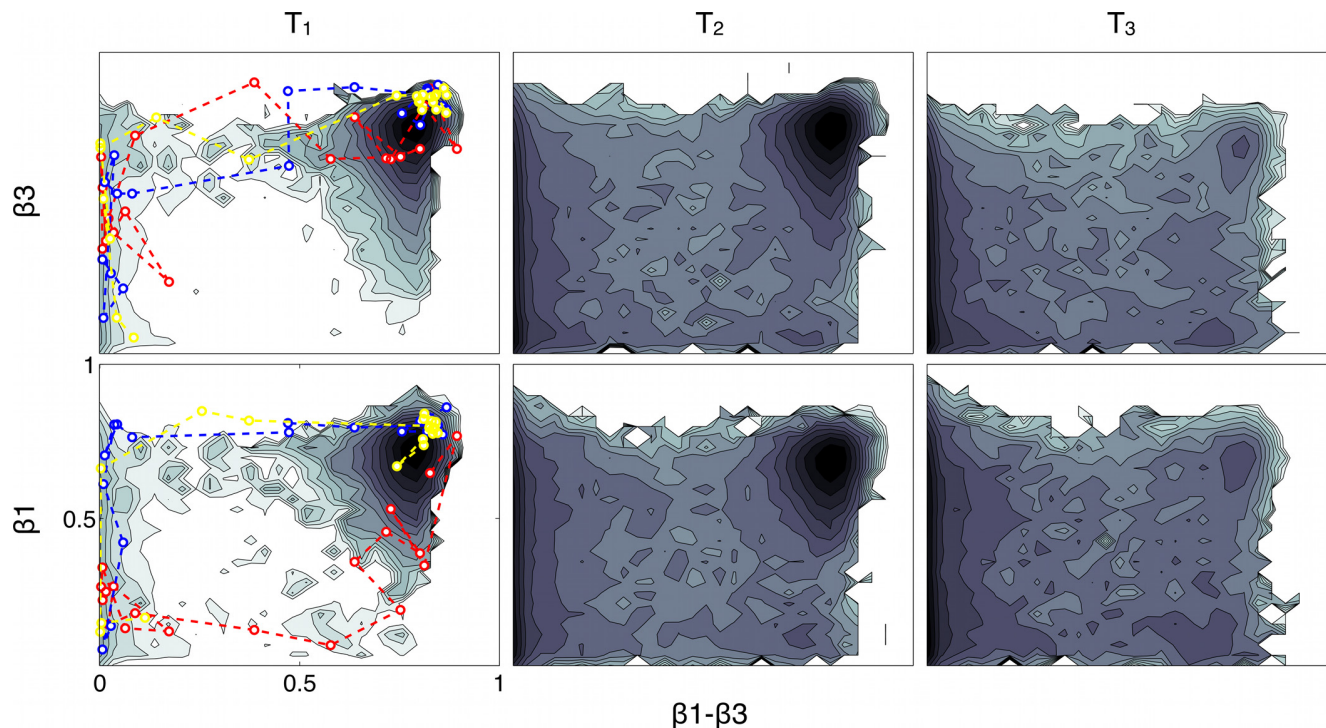
## Discussion

The simulations of the HIV-1-PR monomer performed with theoretical models of different complexity suggested that the high-temperature unfolding and the folding mechanism of this protein are heterogeneous processes. The remarkable analogies found in the nature and in the fine details of these two processes stimulated us to make mechanistic predictions about the actual folding routes of the monomer. To better visualize these analogies, we have projected the all-atom explicit-solvent unfolding trajectories onto the structure-based potential FES (Fig. 4 and 5).

In our scenario, folding is characterized by a milestone event followed by multiple different pathways leading to the native state. The milestone event corresponds to the assembling of an hydrophobic core rich in Valine, Leucine and Isoleucine residues. This collapse can be represented as a two steps process. First, the FN- $\beta$ 3 is formed, and then the folding nucleus made by the two LES S<sub>8</sub> and S<sub>2</sub> is assembled. This result is suggested both by the simulations with the structure-based potential and by the set of high-temperature unfolding trajectories in explicit solvent. The analysis of the latter clearly shows that the unfolding of the hydrophobic core is the last step before the complete denaturation of the monomer in all pathways. This collapsed structure might be related with the possible presence of an intermediate in the folding process, as proposed in Ref. [3].

Several theoretical works indicate that these fragments play an essential role in the folding process. The hierarchy of contacts formed during a G $\ddot{o}$ -model simulation [12]; a study of cooperative folding units that exhibit a stronger protection against unfolding than other parts of the monomer [47]; a Gaussian network model study of the normal modes around the native conformation [48]; the study of structured fragments in the transition state region detected by  $\phi$ -values analysis [5]; the stability temperature associated with each contact of HIV-1-PR [49]; the residues most conserved in those sequences that can successfully fold during a simulation of HIV-1-PR evolution [50].

Further support to these theoretical findings comes from different experimental facts. As common to retrovirus, the mechanism of reproduction of HIV is very fast and at the same time extremely prone to errors. The intrinsic nature of this mechanism together with the pressure induced by drugs led to the appearance of many mutated HIV-1-PR with conserved folding features. An analysis of mutations in 28417 isolates coming from



**Figure 5. FES as a function of  $\beta$ 1- $\beta$ 3 vs.  $\beta$ 3 contacts (top panels) and of  $\beta$ 1- $\beta$ 3 vs.  $\beta$ 1 contacts (bottom panels) at  $T_1 = 0.969$ ,  $T_2 = 0.998$  and  $T_3 = 1.021$ .** FES are obtained by reweighting the structure-based potential PTMetaD simulation. Isoenergy lines are drawn every  $k_B T$ . In dashed lines, projection of the relevant part of the all-atom explicit-solvent unfolding simulations at 700K (blue and red) and 500K (yellow). doi:10.1371/journal.pone.0013208.g005

treated and untreated patients infected with HIV showed that only conservative mutations, *i.e.* substitutions with an amino acid with similar chemical properties, affected those residues belonging to the hydrophobic core [51]. Moreover, a study of the degrees of conservation of residues in a family of proteins structurally similar to HIV-1-PR demonstrated that fragments 22–33 and 81–90 were the most conserved regions [52]. These facts support our theoretical findings as evolution is likely to pay a greater attention in conserving those parts of the protein that play a central role in the folding process.

A strong indication that fragment S<sub>8</sub> is crucial for folding comes also from the ability of the mimetic peptide p-S<sub>8</sub> to inhibit the folding of HIV-1-PR. This was demonstrated in standard enzymatic essays that measured the inhibition constant of p-S<sub>8</sub> ( $K_I = 2.58 \pm 0.78 \mu\text{M}$ ). Circular dichroism spectroscopy showed that inhibition was accompanied by a strong decrease of the  $\beta$ -sheet content suggesting that the enzyme was at least partially unfolded [16]. *Ex-vivo* experiments with infected cells demonstrated that p-S<sub>8</sub> was able to cross the cell membrane, was not toxic to the peripheral blood mononuclear cells and had an antiviral levels below toxic concentration. Consistently with the enzymatic essays, this peptide inhibited the maturation of the virus at a micromolar concentration [17].

The rationale behind the use of p-S<sub>8</sub> as folding inhibitor is that this peptide may interact with its counterpart S<sub>2</sub> and prevent the formation of FN and eventually the folding of the whole monomer. Our results open the way to another possible interpretation. Instead of interacting with its natural partner S<sub>2</sub>, the highly hydrophobic peptide may also disturb the assembling of the entire hydrophobic core. Further experiments on the complex between the protease and the inhibitor could shed light on the details of this process.

After hydrophobic collapse, multiple pathways lead the protein to its native state. Our results suggest that, at physiological condition, the main pathway proceeds through the formation of the two  $\beta$ -strands  $\beta 1$  and  $\beta 3$  and then their docking into the native conformation. This fact is supported also by our long unbiased simulation. Here the breaking and reforming of the hydrogen bonds pattern between  $\beta 1$  and  $\beta 3$  was observed in the time scale of 0.5  $\mu\text{s}$ , while the two subunits remained individually formed. This suggests that the interaction between these two fragments is only marginally stable at room temperature. Moreover, if we examine the main route in the network of high-temperature unfolding pathways (Fig. 3, panel  $\beta 1$ – $\beta 3$ ), the contacts  $\beta 1$ – $\beta 3$  are broken while  $\beta 1$  and  $\beta 3$  retain their native structure.

The FES obtained from our structure-based potential simulations suggest also that, at temperature lower than  $T_f$ , there is no preferential order in which  $\beta 1$  and  $\beta 3$  fold (Figure S9). However,  $\beta 1$  turns out to be the most stable between the two. The stability of  $\beta 1$  was confirmed by our long unbiased all-atom simulation in explicit solvent and by the sequence of events in the unfolding network. Here this  $\beta$ -strand remained structured along the main route toward the unfolded state almost until the complete unfolding of the protein. The predominance of this folding route with respect to others is very sensitive to external conditions. As temperature increases, others pathways become more and more populated. In particular the interaction  $\beta 1$ – $\beta 3$  appears to be formed before each subunits folds, and  $\beta 1$  seems to get structured before  $\beta 3$ .

Our results suggest also that the fragment  $\beta 2$  that corresponds to the flap region in the dimer structure retains a behavior almost uncorrelated with the overall folding process. This can be seen in our structure-based potential FES (Fig. 4) and in the network of unfolding pathways (Fig. 3, panel  $\beta 2$ ). This reflects itself in the

flexibility of this fragment in the native state of the monomer (Figure S1). This flexibility is required to accommodate the substrate inside the active site cavity [24–28,53].

Finally, the native state of the monomer appears stable with respect to thermal unfolding. In our 0.5  $\mu\text{s}$  long unbiased simulation at room temperature the protein retained a very compact structure with root mean square fluctuations on average lower than 1.5 Å, except for the termini and the flap region. These results are compatible with the high unfolding barrier measured experimentally [3].

In conclusion, we have formulated, in a biologically and pharmacologically relevant case, a folding scenario in which multiple pathways and milestones coexist. In this picture, the formation of an hydrophobic core is a milestone event, while the rest of the protein can reach its native state following different pathways and order of assembling. The insight obtained from our simulations, which is supported by several lines of theoretical and experimental evidence, can guide a more rational design of folding-inhibitor drugs as the residues that play a key role in the folding process have been identified. Targeting the formation of the hydrophobic core, being a process common to all the folding pathways, could prove a successful strategy in the fight against AIDS.

## Methods

In the following paragraphs, we provide the details of our all-atom explicit-solvent simulations, of the clustering and network analysis and of the structure-based potential runs. Additional technical information can be found in Text S1.

Throughout all the simulations and analysis, we used a common definition of contact maps. A contact between the  $i$ th and  $j$ th C $_{\alpha}$  atom of the protein is defined as  $C_{ij} = [1 - (r_{ij}/r_0)^6] [1 - (r_{ij}/r_0)^{10}]^{-1}$  where  $r_{ij}$  is the distance between the two atoms and  $r_0 = 8.5$  Å [54]. This definition of contact map is different from the one commonly used in literature [54], which is discrete and where  $r_0$  is a sharp cutoff. In order to be used as collective variables in a metadynamics simulation, contacts must be defined in terms of a function with continuous derivatives.

## All-atom simulations

All-atom simulations were carried out using AMBER99SB force field [55] and NAMD 2.7b1 code [56]. The initial configuration was taken from the structure of the HIV-1-PR dimer (PDB code 1BVG). The monomer was solvated in a periodic cubic box of 84 Å using 18957 TIP3P water molecules [57]. The system was pressurized at 1 atm at 300K using a Langevin thermostat and piston for 500 ps. The NVT run was carried out for 512 ns at 300K using a Langevin thermostat. The unfolding analysis was performed on a set of 30 trajectories generated starting from the same equilibrated structure with different initial velocities. The temperature of 700K was enforced using a Langevin thermostat. All the thermal unfolding runs were simulated for 8 ns. The final configurations had a RMSD calculated on the C $_{\alpha}$  atoms that ranged from 11 to 22 Å from the native structure. An additional simulation at 500K was performed for 1.1  $\mu\text{s}$ .

## Clustering and network analysis

The ensemble of configurations produced in the unfolding simulations was clusterized using the k-means algorithm [32], using as distance between two configurations a properly defined distance in the space of contact maps. Two clusters were connected by a link if a transition between them was observed during the unfolding simulations. To visualize the connectivity

among clusters, we used Visone [58]. The method used to display the network of clusters was the metric multidimensional scaling [59].

The clustering algorithm used here is based on a choice *a priori* of the number of clusters in which data are organized. We explicitly checked that the sequence of events and the different unfolding pathways found by our analysis were robust with respect to this choice (Figure S4). The final data reported in Fig. 3 were generated using 50 clusters.

### Structure-based potential simulations

Coarse-grained simulations were carried out using the all-atom structure-based potential introduced in Ref. [18] and GROMACS 4 [60] equipped with PLUMED [61]. For the PTMetaD simulation, 16 replicas were distributed with a geometric progression in a temperature range between 0.969 and 1.057 in unit of  $T_f = 113.5\text{K}$ . To keep the target temperature, the stochastic thermostat of Bussi *et al.* [62] was used. Exchanges between configurations were attempted every 200 steps. The total simulation time for each replica was  $2 \cdot 10^7$  steps. As collective variable, we used the total number of native contacts without any discrimination among our six subsets. Gaussians of 1.0 kjoule/mol height and 5.0 width were deposited every 1000 steps. We monitored the convergence by calculating at different times the free-energy difference between folded and unfolded states (Figure S5). Convergence was accelerated by orders of magnitude with respect to standard PT [22]. To calculate from the biased simulations the multiple FES as a function of the fraction of native contacts of our six descriptors, we used the reweighting algorithm of Ref. [46].

### Supporting Information

**Text S1** Additional technical details about the all-atom explicit solvent and coarse-grained simulations, and the reweighting algorithm.

Found at: doi:10.1371/journal.pone.0013208.s001 (0.04 MB PDF)

**Figure S1** Root mean square fluctuations (RMSF) of C $\alpha$  atoms during a simulation at room temperature initiated from the crystallographic structure. The RMSF has been calculated using the tool *g\_rmsf* included in the GROMACS 4 package. In the insert, residues with RMSF between 1.0 Å and 1.5 Å are colored in green while residues with RMSF larger than 1.5 Å are colored in red.

Found at: doi:10.1371/journal.pone.0013208.s002 (7.06 MB EPS)

**Figure S2** Analysis of the all-atom explicit-solvent simulation of HIV-1-PR monomer at room temperature. Top. Time evolution of hydrogen bond numbers between N-terminus (residues 3–7) and

C-terminus (residues 95–98). Bottom. Time series of the six sets of native contacts of HIV-1-PR monomer.

Found at: doi:10.1371/journal.pone.0013208.s003 (4.56 MB EPS)

**Figure S3** Atomistic detail of  $\beta 1$ – $\beta 3$  interaction. During the 0.5  $\mu\text{s}$  NVT simulation starting from the native state, the hydrogen bonds GLU65:H-LYS14:O and GLU65:OE2-LYS14:2HE are broken and reformed.

Found at: doi:10.1371/journal.pone.0013208.s004 (4.04 MB EPS)

**Figure S4** Network analysis performed with 100 clusters (top) and 200 clusters (bottom). Coloring is done according to the formation of  $\beta 1$ .

Found at: doi:10.1371/journal.pone.0013208.s005 (3.64 MB EPS)

**Figure S5** FES convergence in the PTMetaD run measured as the free-energy difference between folded and unfolded states as a function of time.

Found at: doi:10.1371/journal.pone.0013208.s006 (1.94 MB EPS)

**Figure S6** FES as a function of the fraction of native contacts of three  $\beta$ -strand subunits of HIV-1-PR at  $T_1 = 0.969$ ,  $T_2 = 0.998$  and  $T_3 = 1.021$ . FES are obtained by reweighting the structure-based potential PTMetaD simulation.

Found at: doi:10.1371/journal.pone.0013208.s007 (1.62 MB EPS)

**Figure S7** FES as a function of HYDRO and all the other contacts at  $T_1 = 0.969$ ,  $T_2 = 0.998$  and  $T_3 = 1.021$ . FES are obtained by reweighting the structure-based potential PTMetaD simulation. Isoenergy lines are drawn every kBT.

Found at: doi:10.1371/journal.pone.0013208.s008 (3.84 MB EPS)

**Figure S8** FES as a function of FN and FN- $\beta 3$  at  $T_1 = 0.969$ ,  $T_2 = 0.998$  and  $T_3 = 1.021$ . FES are obtained by reweighting the structure-based potential PTMetaD simulation. Isoenergy lines are drawn every kBT.

Found at: doi:10.1371/journal.pone.0013208.s009 (3.74 MB EPS)

**Figure S9** FES as a function of  $\beta 1$  and  $\beta 3$  at  $T_1 = 0.969$ ,  $T_2 = 0.998$  and  $T_3 = 1.021$ . FES are obtained by reweighting the structure-based potential PTMetaD simulation. Isoenergy lines are drawn every kBT.

Found at: doi:10.1371/journal.pone.0013208.s010 (3.70 MB EPS)

### Acknowledgments

Computational time for this work was provided by the Swiss National Supercomputing Centre-CSCS.

### Author Contributions

Conceived and designed the experiments: MP. Performed the experiments: MB. Analyzed the data: MB AB FLG. Contributed reagents/materials/analysis tools: AB. Wrote the paper: MB FLG MP.

### References

- Xie D, Gulnik S, Gustchina E, Yu B, Shao W, et al. (1999) Drug resistance mutations can affect dimer stability of HIV-1 protease at neutral pH. *Protein Sci* 8: 1702–1713.
- Ishima R, Ghirlando R, Todtzer J, Gronenborn AM, Torchia DA, et al. (2001) Folded monomer of HIV-1 protease. *J Biol Chem* 276: 49110–49116.
- Noel AF, Bilsel O, Kundu A, Wu Y, Zitzewitz JA, et al. (2009) The folding free-energy surface of HIV-1 protease: Insights into the thermodynamic basis for resistance to inhibitors. *J Mol Biol* 387: 1002–1016.
- Levy Y, Caffish A (2003) The flexibility of monomeric and dimeric HIV-1 PR. *J Phys Chem B* 107: 3068–3079.
- Levy Y, Caffish A, Onuchic JN, Wolynes PG (2004) The folding and dimerization of HIV-1 protease: Evidence for a stable monomer from simulations. *J Mol Biol* 340: 67–79.
- Tomasselli AG, Heinrichson RL (2000) Targeting the HIV-protease in AIDS therapy: a current clinical perspective. *Biochim Biophys Acta* 1477: 189–214.
- Bowman MJ, Chmielewski J (2002) Novel strategies for targeting the dimerization interface of HIV protease with cross-linked interfacial peptides. *Biopolymers* 66: 126–133.
- Bannwarth L, Reboud-Ravaux M (2007) An alternative strategy for inhibiting multidrug-resistant mutants of the dimeric HIV-1 protease by targeting the subunit interface. *Biochem Soc Trans* 35: 551–554.
- Rout MK, Hosur RV (2009) Fluctuating partially native-like topologies in the acid denatured ensemble of autolysis resistant HIV-1 protease. *Archives of Biochemistry and Biophysics* 482: 33–41.
- Ishima R (2003) Solution structure of the mature HIV-1 protease monomer: insight into the tertiary fold and stability of a precursor. *Journal of Biological Chemistry* 278: 43311–43319.
- Kogo H, Takeuchi K, Inoue H, Kihara H, Kojima M, et al. (2009) Urea-dependent unfolding of HIV-1 protease studied by circular dichroism and small-angle x-ray scattering. *Biochim Biophys Acta* 1794: 70–74.

12. Broglio RA, Tiana G, Sutto L, Provasi D, Simona F (2005) Design of HIV-1-PR inhibitors that do not create resistance: Blocking the folding of single monomers. *Protein Sci* 14: 2668–2681.
13. Broglio RA, Levy Y, Tiana G (2008) HIV-1 protease folding and the design of drugs which do not create resistance. *Curr Opin Struct Biol* 18: 60–6.
14. Laio A, Parrinello M (2002) Escaping free energy minima. *Proc Natl Acad Sci USA* 99: 12562–12566.
15. Bonomi M, Gervasio FL, Tiana G, Provasi D, Broglio RA, et al. (2007) Insight into the folding inhibition of the HIV-1 protease by a small peptide. *Biophys J* 93: 2813–21.
16. Broglio RA, Provasi D, Vasile F, Ottolina G, Longhi R, et al. (2006) A folding inhibitor of the HIV-1 protease. *Proteins: Structure, Function, and Bioinformatics* 62: 928–933.
17. Rusconi S, Cicero ML, Laface AE, Ferramosca S, Siriani F, et al. (2007) Susceptibility to a non-conventional (folding) protease inhibitor of human immunodeficiency virus type 1 isolates in vitro. *Proceedings of the International School of Physics “Enrico Fermi” Protein Folding and drug design* 165: 293–299.
18. Whitford PC, Noel JK, Gosavi S, Schug A, Sanbonmatsu KY, et al. (2008) An all-atom structure-based potential for proteins: Bridging minimal models with all-atom all-atom empirical forcefields. *Proteins* 75: 430–441.
19. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314: 141–151.
20. Barducci A, Bussi G, Parrinello M (2008) Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys Rev Lett* 100: 020603.
21. Bussi G, Gervasio FL, Laio A, Parrinello M (2006) Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics. *J Am Chem Soc* 128: 13435–41.
22. Bonomi M, Parrinello M (2010) Enhanced sampling in the well-tempered ensemble. *Phys Rev Lett* 104: 190601.
23. Yan M, Sha Y, Wang J, Xiong X, Ren J, et al. (2008) Molecular dynamics simulations of HIV-1 protease monomer: Assembly of n-terminus and c-terminus into  $\beta$ -sheet in water solution. *Proteins* 70: 731–738.
24. Galiano L, Bonora M, Fanucci GE (2007) Interflap distances in HIV-1 protease determined by pulsed epr measurements. *J Am Chem Soc* 129: 11004–11005.
25. Nicholson LK, Yamazaki T, Torchia DA, Grzesiek S, Bax A, et al. (1995) Flexibility and function in HIV-1 protease. *Nature Struct Biol* 2: 274–280.
26. Hornak V, Okur A, Rizzo RC, Simmerling C (2006) HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc Natl Acad Sci USA* 103: 915–920.
27. Ding F, Layten M, Simmerling C (2008) Solution structure of HIV-1 protease flaps probed by comparison of molecular dynamics simulation ensembles and EPR experiments. *J Am Chem Soc* 130: 7184–7185.
28. Piana S, Carloni P, Parrinello M (2002) Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease. *J Mol Biol* 319: 567–583.
29. Fersht A, Daggett V (2002) Protein folding and unfolding at atomic resolution. *Cell* 108: 573–582.
30. Daggett V (2006) Protein folding-simulation. *Chem Rev* 106: 1898–1916.
31. Settanni G, Fersht AR (2008) High temperature unfolding simulations of the TRPZ1 peptide. *Biophys J* 94: 4444–4453.
32. Lloyd SP (1982) Least-squares quantization in PCM. *Ieee T Inform Theory* 28: 129–137.
33. Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE (2009) Long-timescale molecular dynamics simulations of protein structure and function. *Curr Opin Struct Biol* 19: 120–127.
34. Chipot C, Pohorille A (2007) *Free Energy Calculations. Theory and Applications in Chemistry and Biology*. Berlin/Heidelberg: Springer. 518 p.
35. Dellago C, Bolhuis PG (2009) Transition path sampling and other advanced simulation techniques for rare events. *Adv Polym Sci* 221: 167–233.
36. Leopold PE, Montal M, Onuchic JN (1992) Protein folding funnels - a kinetic approach to the sequence structure relationship. *Proc Natl Acad Sci USA* 89: 8721–8725.
37. Wolynes PG, Onuchic JN, Thirumalai D (1995) Navigating the folding routes. *Science* 267: 1619–1620.
38. Dill KA, Chan HS (1997) From Levinthal to pathways to funnels. *Nat Struct Biol* 4: 10–19.
39. Wolynes PG (1997) Folding funnels and energy landscapes of larger proteins within the capillarity approximation. *Proc Natl Acad Sci USA* 94: 6170–6175.
40. Shakhnovich E, Gutin AM (1993) Engineering of stable and fast-folding sequences of model proteins. *Proc Natl Acad Sci USA* 90: 7195–7199.
41. Munoz V, Eaton WA (1999) A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc Natl Acad Sci USA* 96: 11311–11316.
42. Galzitskaya OV, Finkelstein AV (1999) A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proc Natl Acad Sci USA* 96: 11299–11304.
43. Alm E, Baker D (1999) Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc Natl Acad Sci USA* 96: 11305–11310.
44. Clementi C, Jennings PA, Onuchic JN (2000) How native-state topology affects the folding of dihydrofolate reductase and interleukin-1 beta. *Proc Natl Acad Sci USA* 97: 5871–5876.
45. Weinkam P, Zong CH, Wolynes PG (2005) A funneled energy landscape for cytochrome c directly predicts the sequential folding route inferred from hydrogen exchange experiments. *Proc Natl Acad Sci USA* 102: 12401–12406.
46. Bonomi M, Barducci A, Parrinello M (2009) Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *J Comput Chem* 30: 1615–1621.
47. Wallqvist A, Smythers GW, Covell DG (1998) A cooperative folding unit in HIV-1 protease. implications for protein stability and occurrence of drug-induced mutations. *Protein Eng* 11: 999–1005.
48. Bahar I, Atilgan AR, Demirel MC, Erman B (1998) Vibrational dynamics of folded proteins: Significance of slow and fast motions in relation to function and stability. *Phys Rev Lett* 80: 2733–2736.
49. Cecconi F, Micheletti C, Carloni P, Maritan A (2001) Molecular dynamics studies on HIV-1 protease drug resistance and folding pathways. *Proteins* 43: 365–372.
50. Tiana G, Broglio RA (2009) The molecular evolution of HIV-1 protease simulated at atomic detail. *Proteins* 76: 895–910.
51. Shafer RW, Hsu P, Patik AK, Craig C, Brendel V (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.
52. Holm L, Sander C (1996) Mapping the protein universe. *Science* 273: 595–602.
53. Pietrucci F, Marinelli F, Carloni P, Laio A (2009) Substrate binding mechanism of HIV-1 protease from explicit-solvent atomistic simulations. *J Am Chem Soc* 131: 11811–11818.
54. Vendruscolo M, Najmanovich R, Domany E (1999) Protein folding in contact map space. *Phys Rev Lett* 82: 656–659.
55. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, et al. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* 65: 712–725.
56. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26: 1781–802.
57. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79: 926–935.
58. Brandes U, Wagner D (2003) Visone: Analysis and visualization of social networks. In: Jünger M, Mutzel P, eds. *Graph Drawing Software*. Berlin, Germany: Springer. pp 321–340.
59. Brandes U, Pich C (2007) Eigensolver methods for progressive multidimensional scaling of large data. In: *Proc. 14th Intl. Symp. Graph Drawing (GD '06)*. pp 42–53.
60. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4: 435–447.
61. Bonomi M, Branduardi D, Bussi G, Camilloni C, Provasi D, et al. (2009) PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comp Phys Comm* 180: 1961–1972.
62. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126: 014101.