doi:10.1016/j.jmb.2004.04.028 *J. Mol. Biol.* (2004) **340**, 67–79

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

The Folding and Dimerization of HIV-1 Protease: Evidence for a Stable Monomer from Simulations

Yaakov Levy[1,2]*, Amedeo Caflisch[3], Jose N. Onuchic[1] and Peter G. Wolynes[1,2]

[1]*Center for Theoretical Biological Physics and Department of Physics University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA*

[2]*Department of Chemistry and Biochemistry, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA*

[3]*Department of Biochemistry University of Zurich Winterthurerstrasse 190 CH-8057 Zurich, Switzerland*

**Corresponding author*

HIV-1 protease (PR) is a major drug target in combating AIDS, as it plays a key role in maturation and replication of the virus. Six FDA-approved drugs are currently in clinical use, all designed to inhibit enzyme activity by blocking the active site, which exists only in the dimer. An alternative inhibition mode would be required to overcome the emergence of drug-resistance through the accumulation of mutations. This might involve inhibiting the formation of the dimer itself. Here, the folding of HIV-1 PR dimer is studied with several simulation models appropriate for folding mechanism studies. Simulations with an off-lattice Gō-model, which corresponds to a perfectly funneled energy landscape, indicate that the enzyme is formed by association of structured monomers. All-atom molecular dynamics simulations strongly support the stability of an isolated monomer. The conjunction of results from a model that focuses on the protein topology and a detailed all-atom force-field model suggests, in contradiction to some reported equilibrium denaturation experiments, that monomer folding and dimerization are decoupled. The simulation result is, however, in agreement with the recent NMR detection of folded monomers of HIV-1 PR mutants with a destabilized interface. Accordingly, the design of dimerization inhibitors should not focus only on the flexible N and C termini that constitute most of the dimer interface, but also on other structured regions of the monomer. In particular, the relatively high ϕ values for residues 23–35 and 79–87 in both the folding and binding transition states, together with their proximity to the interface, highlight them as good targets for inhibitor design.

Keywords: HIV-1 protease; dimerization; ϕ values; Gō model; MD simulations

Introduction

The homodimeric human immunodeficiency virus type 1 (HIV-1) protease (PR) is crucial for the processing of the viral polyprotein and the maturation of the virus that causes the acquired immunodeficiency syndrome (AIDS). Therefore, the protease is an attractive target for designing pharmacological inhibitors in the fight against AIDS, making it one of the best studied enzymes. The HIV-1 PR is a homodimer with C_2 symmetry in the absence of ligands. The enzyme is an aspartic protease that consists of two identical 99 residue subunits (Figure 1A). Each subunit contributes one catalytic aspartic acid residue (Asp25) to form the active site.[1,2] The active-site region is capped by two identical β-hairpin loops, the flaps (residues 45–55 in each monomer), which regulate substrate entry into the active site. While the flap β-hairpins in the ligand-bound protease are well ordered and interact with the substrate, in the free protease the flaps are very flexible and adopt closed and open conformations.[3–7] The two subunits interact mostly through a four-stranded antiparallel β-sheet (which contributes about 75% of

Abbreviations used: HIV-1, human immunodeficiency virus type 1; PR, protease; AIDS, acquired immunodeficiency syndrome; TSE, transition state ensemble.

E-mail address of the corresponding author: klevy@physics.ucsd.edu

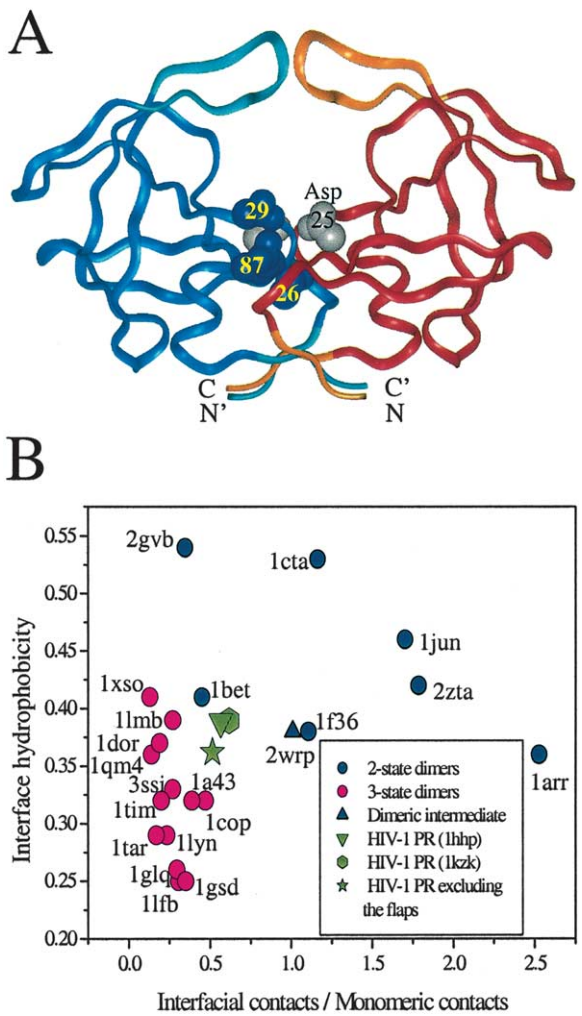

Figure 1. A, The structure of homodimeric HIV-1 PR. The two identical subunits are colored red and blue whereas their flap (residues 45–55) and the N and C termini (residues 1–4 and 96–99, respectively) are colored orange and cyan. The catalytic Asp25 is represented by gray van der Waals spheres. Residues T26, D29, and R87 are represented by blue van der Waals spheres. Mutating these residues to Ala, Asn, and Lys, respectively, yields a stable folded monomer.[18–20] B, A "phase diagram" that correlates the association mechanism of the homodimers with their structural properties. Each homodimer is classified structurally on the basis of the number of intramonomeric and interfacial native contacts as well as the interface hydrophobicity. The phase diagram includes homodimers that are classified experimentally as two-state or three-state dimers. The location of HIV-1 PR in the diagram does not indicate a marked preference. When the flaps, which are intrinsically flexible in the ligand-free form of the enzyme, are excluded, the HIV-1 PR adopts properties that are more typical of three-state homodimers.

the dimerization stabilization energy[8]) consisting of the N and C termini of the protease monomers (residues 1–4 and 96–99, respectively, see Figure 1A).

Two strategies are used to develop inhibitors of HIV-1 PR. The first approach is to design inhibitors that compete with natural substrates for the active site, on the basis of structural and chemical complementarities,[9,10] as well as thermodynamic stabilities.[11] However, while several drugs inhibit the HIV-1 PR by blocking the active site, their efficacy is restricted due to the natural selection of protease variants that are still catalytically competent but have lower affinity for the drugs than the wild-type enzyme. A second strategy takes advantage of the fact that the monomers are inactive, and suggests developing compounds that destabilize the dimeric structure of the protease by binding at the subunit interface. This approach was applied to design peptides derived from the N and C termini of the HIV-1 PR (linked by various linkers) with the aim to inhibit enzymatic activity by blocking its interface and thus preventing dimerization.[12–15] Although dimerization inhibitors were found to be effective against the protease activity, they are not in clinical use. A third possibility would be to inhibit the folding of the monomer itself.

Understanding the binding mechanism of HIV-1 PR is crucial, as it may help to suggest new routes to disrupt the association of the enzyme subunits and therefore destroy its activity. Experimental denaturation studies suggest that folding and binding of HIV-1 PR occur simultaneously.[8,16,17] Accordingly, these studies suggest that the folded dimer is in equilibrium with unfolded monomers, and that individual folded monomers do not exist at appreciable concentration, as they are intrinsically unstable. However, in a recent molecular dynamics simulation study using an all-atom representation, a monomeric HIV-1 PR appeared to be relatively stable.[6] The isolated monomer has secondary and tertiary structure very similar to that of the bound monomer, yet it exhibits an enhanced flexibility of its N and C termini, which, in the presence of another monomer, constitute most of the dimer interface. Some recent NMR studies have reported folded monomeric HIV-1 PR for several mutants.[18–20] The introduced mutations (T26A, D29N, and R87K) destabilize the interface by affecting the network of intermonomeric interactions. Moreover, a native-like single subunit fold was detected in an intrachain-linked monomer by preventing the interface formation by engineering cysteine residues at the N and C termini (at positions 2 and 97 or 98).[19]

Here, we study the thermodynamics and kinetics of folding and binding of HIV-1 PR using three simulation models that address different aspects of the dimer formation. We used a Gō model to study the binding processes and to characterize the transition state ensemble of folding and binding. This model, which corresponds to a perfectly funneled energy landscape, has been used recently to study the association mechanisms of more than ten homodimers that are classified experimentally as to whether intermediates accumulate during their binding.[21] The Gō model simulations successfully reproduce the experimental classification of the homodimers and indicate that binding, like

folding,[22,23] is governed by funneled energy landscape. In addition, good agreement was found between the simulation and experimental ϕ values for dimerization of the Arc repressor and the tetrameric protein domain from tumor suppressor p53 (Y.L. *et al*., unpublished results). The success of the simplified model in predicting both the gross and finer structure aspects of the binding mechanisms of various protein complexes makes this model appropriate to re-examine the binding mechanism of HIV-1 PR and to query the existence of monomeric intermediates during association.

To complement the Gō simulations, an isolated monomer was simulated with an all-atom forcefield using explicit as well as implicit models for the water molecules. While the all-atom simulations are too computationally demanding for studying folding (and the dimerization), these simulations may provide an indication of the structural stability of the unbound folded monomer. The combination of Gō and all-atom simulations gives us information on whether the folding of the HIV-1 PR monomers is coupled to the dimerization (i.e. two-state mechanism) or there is an association of already folded monomers (i.e. three-state mechanism).[24–26] As a first step to infer the binding mechanism of HIV-1 PR, its structural properties were compared to those of other homodimers that have been classified experimentally as either two-state or three-state. More specifically, the HIV-1 PR dimer is placed on a phase diagram that describes the ratio between the interfacial and monomeric contacts as well as the interface hydrophobicity.[6] This phase diagram (Figure 1B) has already distinguished successfully between homodimers that are formed by two-state and three-state mechanisms. The HIV-1 PR was placed in the phase diagram on the basis of two ligand-free forms of the enzyme (PDB codes 1hhp and 1kzk) that differ mainly in the flap conformation. We can locate it in the diagram when the flaps, which are known to be very flexible in the absence of bound ligand,[3–6] are excluded. The location in the phase diagram of HIV-1 PR with or without the flaps does not support unambiguously either mechanism, since both structures are on the boundary between the two dimer types; however, when the flaps are excluded, the protease shifts slightly to a region in the phase diagram that is more typical of a three-state binding mechanism.

An atomic-level description of the folding and dimerization of HIV-1 PR may provide a more accurate knowledge of the stability of the isolated monomer as well as of the dimer. Characterizing the transition state ensemble for folding and binding will indicate the mechanism of recognition and the key residues for these processes. More specifically, an atomic-level description of the monomer dimerization suggests an alternative approach for design of more powerful dimerization inhibitors.

Results and Discussion

Folding of HIV-1 PR: association of structured monomers

We simulated the formation of dimeric HIV-1 PR using a simple Gō model. During constant-temperature simulations, we monitored the number of native contacts, Q_{Total}, which is a sum of monomeric native contacts (Q_{A} and Q_{B}) and interfacial native contacts ($Q_{\text{Interface}}$). Folding and binding rate coefficients are strongly temperature-dependent. It is thus crucial to determine accurately the transition temperature. For this purpose, we apply the multiple histogram method to compute the heat capacity:

$$Cv = (\langle E^2\rangle - \langle E\rangle^2)/kT^2$$

as a function of temperature.[27] The heat capacity plot for the association of the monomers is shown in Figure 2A and exhibits two peaks at temperatures T_1 and T_2. The two transition temperatures of the protease indicate the existence of at least two decoupled processes. Two trajectories that were carried out at the two transition temperatures are shown in Figure 2B and C. These trajectories indicate that temperature $T_1(=1.41\varepsilon)$ is the transition temperature for binding between two folded monomers. However, at T_1 folding/unfolding of a single monomer can occur and there is an equilibrium between one or two structured monomers and the folded dimer (Figure 2B). Temperature T_2 $(=1.47\varepsilon)$ is the transition temperature for monomer folding and the equilibrium is between two unstructured monomers and a single folded monomer (Figure 2C). The existence of only two peaks (for some three-state homodimers three peaks were found[28]) reflects the fact that monomer folding and binding are not completely separated thermodynamically. However, the binding events of HIV-1 PR in Figure 2B indicate clearly that the dimer is formed by association of already folded monomers and no coupling between folding and binding is observed. For Arc-repressor, factor for inversion stimulation, troponin site III, gene V protein, and β nerve growth factor, which were found experimentally to fold by binding, a single peak in the specific heat plot was observed. For λ repressor, superoxide dismutase, LFB1 transcription factor, λ cro repressor, and streptomyces subtilisin inhibitor, which have a stable folded monomer, two or three peaks in the specific heat profiles were detected.[21,28] The fact that two peaks were detected in the corresponding specific heat capacity profile of HIV-1 PR supports the suggestion that dimerization takes place after monomer folding.

The free energy profiles for the folding and dimerization of the protease are shown in Figure 3A. The free energy for the dimer was calculated at T_1 (Figure 2A) and that of the monomer at the T_{f} for monomer folding $(=1.46\varepsilon)$. The free energy profile for monomer folding was calculated

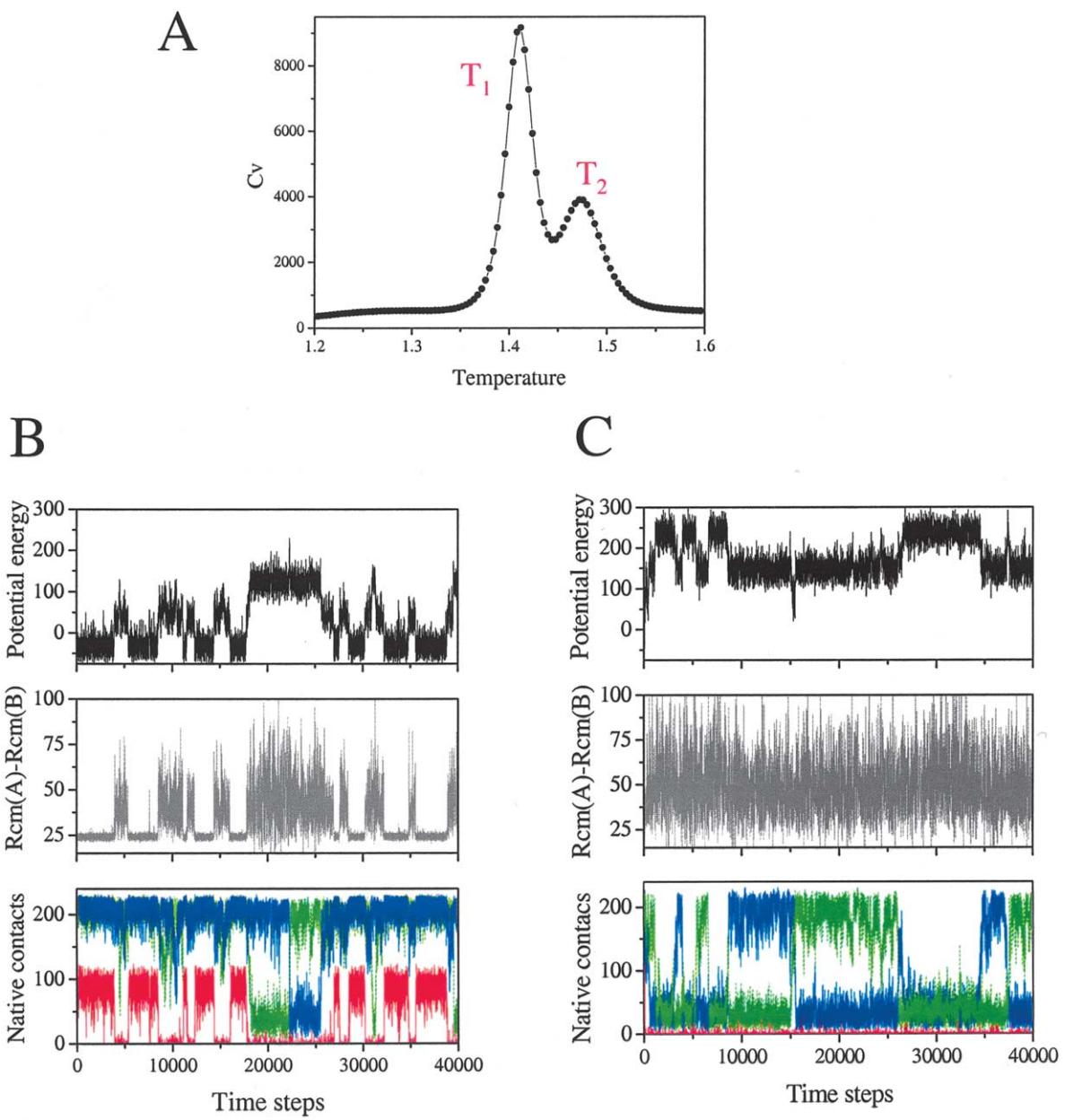

Figure 2. The heat capacity of HIV-1 PR as a function of temperature in units of ε which is the energy gain of native contact formation (A). The heat capacity profile includes two peaks (designated T_1 and T_2), which indicate the existence of more than a single transition. Typical trajectories of folding and association at T_1 (B) and T_2 (C). The time evolution of the potential energy, the separation distance (the distance between the center of mass of the two chains, Rcm(A)−Rcm(B)), as well as, Q_A (green), Q_B (blue), and $Q_{Interface}$ (red), illustrate the decoupling between folding and binding.

from simulations of an isolated monomer. The main barrier for monomer folding (TSE_1) is at $85 < Q < 125$ and an additional barrier but much lower (TSE_2) is detected at larger Q. The dimer free energy profile contains four states and not two, as expected from a two-state dimer. These states include: two unfolded chains (2U), a single folded monomer (while the other is unfolded), two unbound folded monomers (2M), and a folded dimer (D). The free energy profile for the dimer shows three main barriers. Two barriers correspond to the folding of a single monomer (the secondary small folding barriers, TSE_2, are hardly seen at T_1) and a barrier for binding.

The free energy surface of the binding process of HIV-1 PR is projected onto several progress variables for folding and binding: the monomeric native contacts (Q_A and Q_B), interfacial native contacts ($Q_{Interface}$), the total number of native contacts (Q_{Total}), and the distance between the center of mass of the two subunits (Figure 3). These provide a detailed investigation on the binding mechanism. The free energy surfaces projected along Q_{Total} and the separation distance between the two subunits

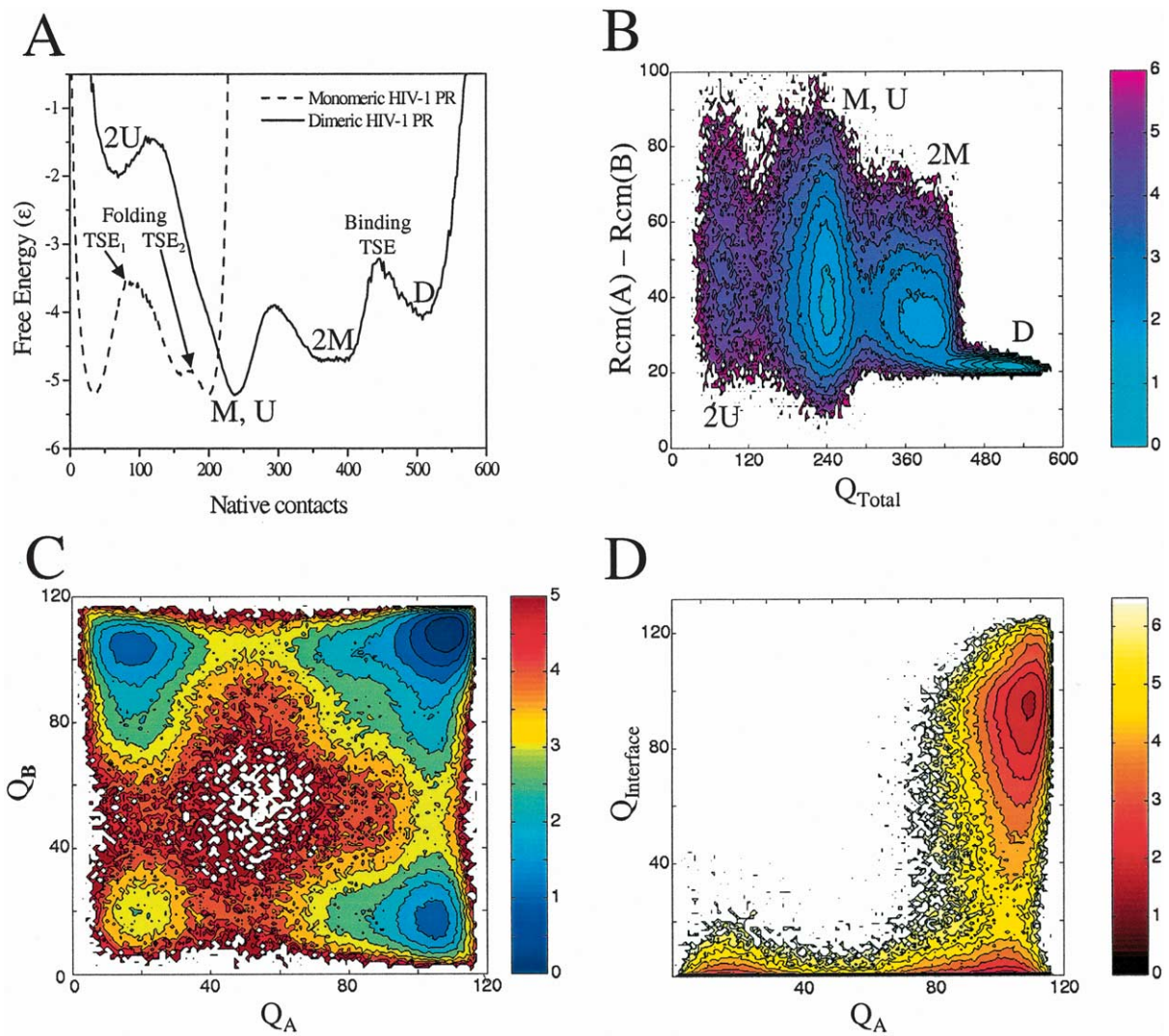

Figure 3. A, Free energy as a function of the reaction coordinate Q_{Total} for monomeric (broken line) and dimeric (continuous line) HIV-1 PR. The free energies for the monomer and dimer were calculated at the monomer T_f and T_1, respectively. B–D, Free energy surfaces for folding and binding. Free energy surfaces of HIV-1 PR are plotted as a function of the total number of native contacts (Q_{Total}), the separation distance between the two chains (Rcm(A)-Rcm(B)), the intra-subunit native contacts (Q_A and Q_B), and the inter-subunit native contacts ($Q_{Interface}$). The free energy surfaces clearly indicate the decoupling between the folding of the two monomers as well as the decoupling between folding and binding.

(Figure 3B) shows that starting from two unfolded chains, which can be far away from each other (the distance between their center of mass is up to 100 Å), folded monomers have to be formed for dimerization. Namely, the structured monomers constitute intermediates during the folding process. The surface projected along the Q_A and Q_B coordinates (Figure 3C) demonstrates the lack of coupling between the folding of the two subunits as each monomer folds irrespective of the other monomer. The projection on the monomeric contacts (Q_A or Q_B) and the interfacial contacts ($Q_{Interface}$) (Figure 3D) shows the decoupling between folding and binding reflected by the low free energy of a state where a monomer is folded with no interfacial contacts. The Gō simulations indicate strongly that the two chains are autonomous entities and can fold regardless of the presence of the other subunit, and that binding is not conditional for their stability. A similar dimerization mechanism was obtained when the set of native contacts obtained from the CSU software was included in the Gō potential.

The nature of the transition state ensemble of folding and dimerization

An understanding of the folding and binding mechanisms requires a microscopic characterization of the transition state ensemble (TSE) of each process. Experimentally, the nature of the transition state ensemble is inferred from the ϕ value,[29] which is the ratio of the effect of a mutation at a given position along the chain on

the stability of the TSE over its effect on the stability of the folded state. A ϕ value close to 1 means that the mutation similarly affects the TSE and the folded state, indicating that the mutated residue is structured in the TSE as in the folded state. Inversely, a ϕ value close to 0 means that the mutation does not affect the stability of the TSE (relative to its unfolded state), indicating that the mutated residue is unstructured at the TSE. While ϕ value analysis has been used widely to decipher folding mechanisms, it has been applied recently to characterize the TSE of binding;[30–32] namely, to study reactions that are higher than unimolecular reactions.

The ϕ value of each of the native contacts was calculated for the two transition state ensembles observed in the free energy profile for monomer folding (TSE_1 and TSE_2) as well as the transition state for binding (see Figure 3A). The contacts ϕ values, ϕ_{ij}, which were calculated on the basis of the probability to form the contact under consideration at the native and unfolded states as well as at the transition state (equation (2)), were used to estimate ϕ_i, the ϕ value at position i. The ϕ_i for folding and binding describe the extent to which residue i is structured at the corresponding free energy barrier.

Among the ϕ_i at the major transition state ensemble for folding (TSE_1) only the ϕ values of residues 27–35 and 79–87 are above 0.5 (Figure 4A). These two regions, which are far apart along the sequence, fold cooperatively and define the kinetic bottleneck for monomer folding (Figure 4C). This is consistent with the recent report on the importance of the triplet Asp-Thr-Gly (residues 25–27) for the initial phases of dimer formation.[33] Our observation of the critical role of these residues in the monomer folding is in agreement with Gō simulations of the protease where the crystallographic symmetry was enforced during the simulations.[34] Interestingly, these regions include mutating sites responsible for drug resistance.[34,35] Moreover, the catalytic Asp25 is close to these key residues for folding. The ϕ

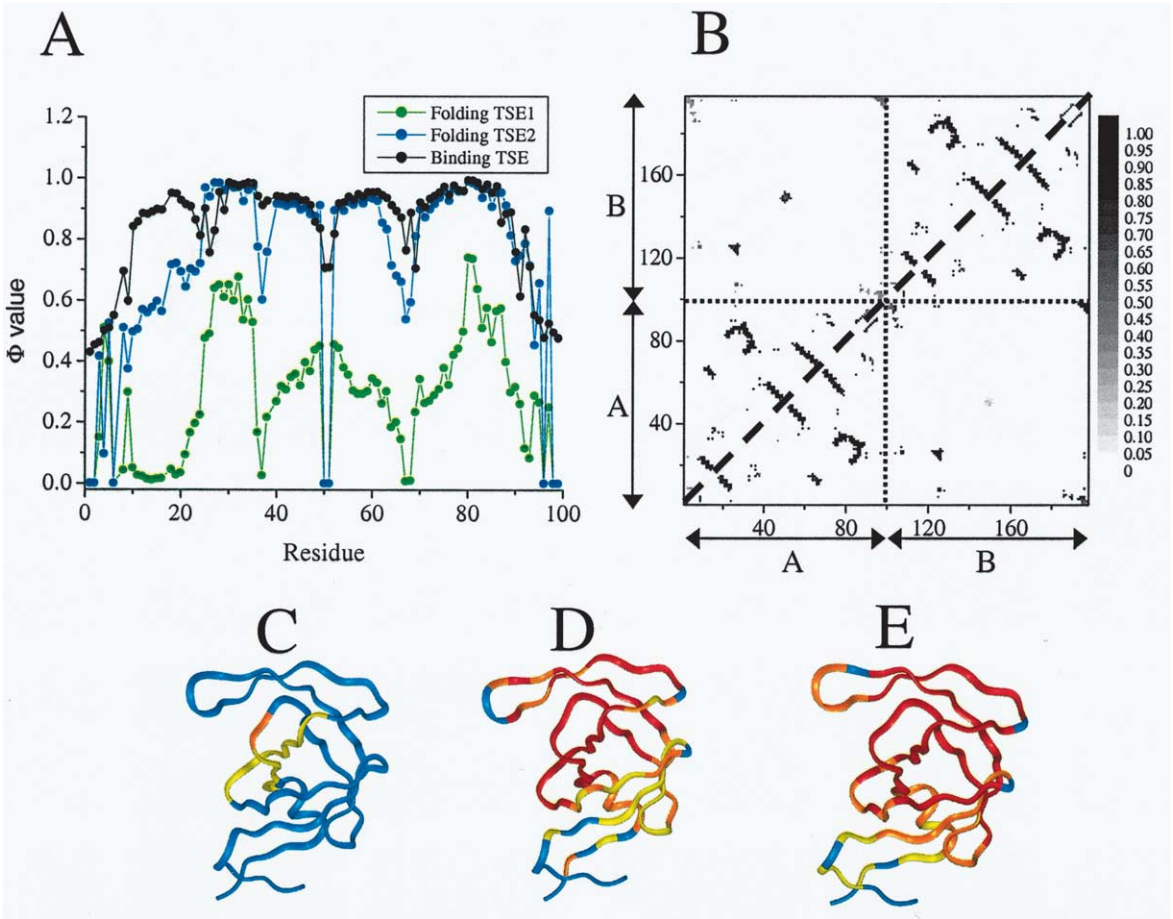

Figure 4. The ϕ value analysis for folding and binding of HIV-1 PR. A, The ϕ value at the folding transition state ensembles (TSE_1 and TSE_2) and the binding transition state ensemble. The ϕ values at TSE_1 and TSE_2 were calculated on the basis of the folding simulations of a single subunit and defined by $85 < Q < 125$ and $150 < Q < 170$, respectively (see Figure 3A). The binding ϕ values were calculated on the basis of the dimer simulation and the binding TSE is defined by $425 < Q < 475$. B, The ϕ value for all native contacts at the binding TSE (upper triangle) and the probability of native contact formation at the binding TSE (bottom triangle). C–E, Color representation of the ϕ values at TSE_1, TSE_2, and the binding TSE: $\phi_i < 0.5$ blue, $0.5 < \phi_i < 0.7$ yellow, $0.7 < \phi_i < 0.9$ orange, $\phi_i > 0.9$ red.

values at the second transition state ensemble for monomer folding (TSE_2) reflect a much more folded monomer. Relatively low ϕ values are found for the termini, which become folded only upon binding and the flap tip, which is intrinsically unfolded. From the ϕ values at TSE_2 it is clear that the N terminus is less folded than the C terminus. Also, part of the N terminus together with residues 36–38 and 64–68 that form a β-sheet in the folded state are partially folded at TSE_2 (Figure 4D).

The ϕ values for dimerization are generally higher than those for folding (Figure 4A and E). The ϕ values for binding were calculated on the basis of the probabilities of all contacts of a specific residue to be formed. For most residues, the ϕ values at the TSE for binding are above 0.9, indicating structured positions. The regions with ϕ values smaller than 0.5 correspond to the termini. This indicates that the monomers are folded at the transition state ensemble of binding and only regions that participate in the interface contact network are not fully folded. The binding ϕ value of the native contacts in the dimeric form as well as the probability of contact formation at the binding TSE are shown in Figure 4B. The monomeric native contacts have both high ϕ values and high formation probabilities. The ϕ value analysis of the binding TSE strongly supports our previous observation that monomer folding and binding of HIV-1 PR are decoupled processes.

Stability of an isolated monomer

To assess the stability of a folded monomer, an isolated monomeric HIV-1 PR was studied using molecular dynamic simulations. These were performed with an explicit water model for 5 ns at 300, 330, and 350 K. In addition, simulations were performed using the EEF1 implicit solvent model[36] for 100 ns at the same temperatures. The RMSD

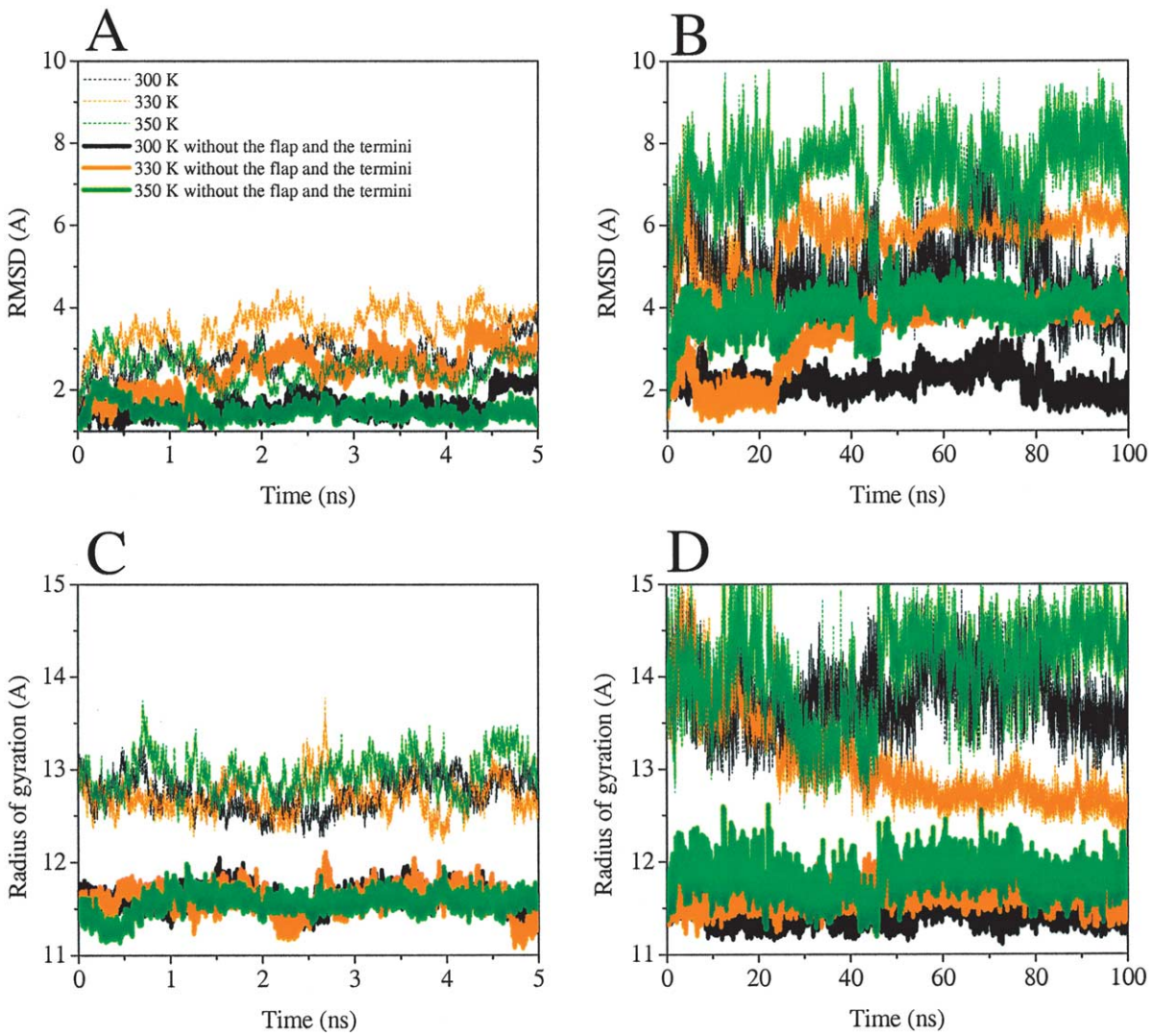

Figure 5. The flexibility of monomeric HIV-1 PR. Time evolution of the RMSD and the radius of gyration in the explicit (A and C) and implicit (B and D) solvent simulations of the monomeric enzyme. The RMSD and radius of gyration were calculated for all residues and when the flap and the termini residues were excluded.

from the X-ray structure during the final nanosecond (from 4 ns to 5 ns) of the explicit water simulations is 3.02(±0.47) Å, 3.77(±0.25) Å, and 2.51(±0.34) Å at 300, 330, and 350 K, respectively. This result is significant, since a recent study on 34 proteins has shown that several are unstable and reach RMSD larger than 5 Å within 5 ns at 300 K.[37] The RMSD from the X-ray structure during the final nanosecond of the explicit water simulations where the flap and the termini are excluded from the calculations are 1.86(±0.33) Å, 2.96(±0.96) Å, and 1.40(±0.15) Å at 300, 330, and

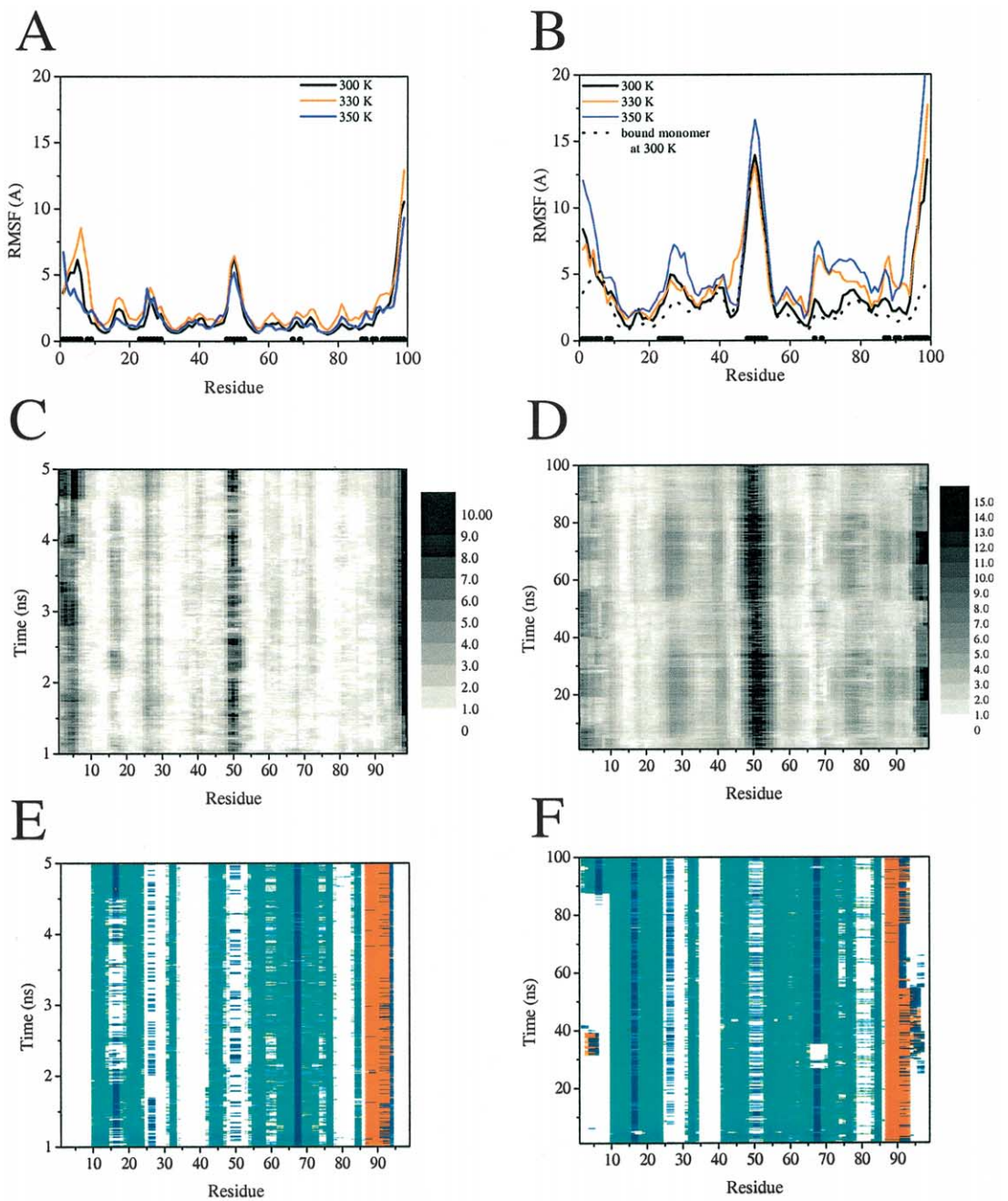

Figure 6. Structural stability of monomeric HIV-1 PR. RMS fluctuations of the C^{α} atoms from the crystal structure as a function of residue number. The RMS fluctuations are based on 5 ns all-atom simulations with explicit (A) and 100 ns simulations with implicit (B) solvent models at 300, 330, and 350 K. The dots on the *x*-axis indicate the residues that have interfacial contacts. The RMS fluctuations of a bound monomer were calculated based on simulation of the dimer with the implicit solvent model at 300 K. The RMS fluctuations indicate great flexibility of the termini and the flap (residues 45–55). The time-dependence of the residual RMSD (after least-squares deviation) is plotted for the 300 K simulations with the explicit (C) and implicit (D) solvent models. Secondary structure of isolated monomeric HIV-1 PR as a function of simulation time based on the explicit (E) and implicit (F) solvent model at 300 K. The secondary structure was determined by DSSP.[49] The α-helix is shown in orange, the H bond turn in blue, and the β-strand in green.

350 K, respectively (Figure 5A). This indicates that most of the monomer flexibility is correlated with the termini and the flap dynamics. The great flexibility of these regions is reflected by the radius of gyration, which becomes stable when these regions are excluded from its calculation (Figure 5C). The fluctuations in the RMSD (RMSF) along these trajectories do not vary significantly with the temperature (Figure 6A) but large fluctuations are seen for the termini residues and residues 45–55, which correspond to the flap. Additionally, residues 24–29 are flexible, though less so than the N and C termini, or the flap region (Figure 6C). The implicit solvent simulations show greater flexibility than the explicit solvent simulations (Figure 5B and D), which is likely a non-equilibrium effect arising from to the lack of friction from water in the implicit solvent model. However, as in the explicit solvent simulations, most of the monomer fluctuations can be attributed to the termini and the flaps. Excluding these regions from the calculations of the RMSD and radius of gyration yield much lower values, indicating that the rest of the monomer is relatively stable. The radii of gyration in both the implicit and explicit solvent simulations (when the highly flexible regions are excluded) have similar values (Figure 5C and D). This observation, together with the similar profile of the RMSF (Figure 6A and B), indicates similar dynamic properties. The dimer simulated at 300 K with the implicit solvent model shows a significantly reduced flexibility of the termini in comparison to the flexibility of the isolated monomer (Figure 5B). While the dimer is, in general, more stable than the monomer it still involves large motion that can be attributed to the flaps.

Figure 6E and F show the time-dependence of the secondary structure content of the monomeric HIV-1 PR in trajectory sampled for the explicit and implicit solvent models at 300 K. It is clear that the secondary structure elements are conserved along these trajectories and that the monomer remains folded. Similar secondary structure contents are obtained from the simulations at 330 and 350 K for both models (data not shown). In our models, while the termini and the flap are very flexible in the monomeric form of the protease, the rest of the protein remains folded and well structured.

Conclusions

The folding and dimerization of HIV-1 PR were studied here by both C^{α} Gō-model and all-atom model simulations. The C^{α} Gō-model was used to explore the dimerization mechanism of HIV-1 PR based on topological information alone and the all-atom simulations were conducted to address the question of the stability of an isolated folded monomer. For 11 homodimers, the C^{α} Gō-model has recently reproduced the gross features of the experimental mechanisms concerning the role of folded monomers in the association.[21] Furthermore, microscopic analysis on the ϕ values for the transition state of dimerization based on the Gō model simulations of Arc-repressor and the P53 tetramerization domain agree well with the experimental values (Y.L. *et al.*, unpublished results).

The simplified Gō simulations indicate that HIV-1 PR folds by association of already folded monomers. Any single monomer folds irrespective of the other monomer and no coupling is observed between monomer folding and dimerization. The stability of a single unbound monomeric HIV-1 PR was additionally treated by all-atom molecular dynamics simulations with both explicit and implicit solvent models. An isolated monomer appears stable at a temperature range of 300–350 K and its secondary structure elements are preserved on both the 5 ns (explicit water) and the 100 ns (implicit water) timescale. Yet, few regions of the monomer are flexible: the N and C termini, the flap, and the region 24–29 that includes the catalytic Asp25. These regions participate in the dimer interface and will exhibit much reduced flexibility upon association.

The finding that HIV-1 PR is formed by association of folded monomers (i.e. three-state mechanism) is in disagreement with some equilibrium denaturation experiments,[8,16] which have suggested coupling between the monomer folding and binding. The success of the Gō-model reproducing both the gross and finer aspects of the binding mechanisms of several other protein complexes, as well as the agreement between the Gō-model and all-atom simulations, support our conclusion that monomeric HIV-1 PR is, in fact, stable and its folding is conditional for the dimer formation. We think that the apparent disagreement with experiments stems from the fact that HIV-1 PR has a low dissociation constant ($K_D < 5$ nM).[38] In general, detecting an intermediate requires measurements with concentrations of the order of K_D, which is, unfortunately, impractical with current instrument sensitivity. Support for our conclusion has been obtained recently by detecting folded monomeric mutants of HIV-1 PR.[18] The mutations are at positions 26, 29, and 87, which are involved in the network of contacts at the interface. Accordingly, the mutations (T26A, D29N, and R87K) result in an increased K_D and, thus, detecting a folded monomer is feasible at higher concentrations than those for the wild-type that are dictated by its low K_D. Moreover, a folded monomer was detected when the termini were deleted[18,20] or when the interface was blocked by engineering cysteine residues at the termini that form a disulfide bond thus avoiding their participation in the interfacial contacts.[18] In addition, the high-resolution NMR structure of a monomer of a Mason–Pfizer monkey virus protease, which exhibits conservative structural motif of HIV-1 PR,[39] supports the monomer stability of HIV-1 PR.

The current clinical approach for inhibiting the enzyme activity is based on blocking the active site. It was suggested previously that inhibiting the dimerization might overcome the limitation of this mode of inhibition, which suffers from inevitable drug resistance.[12–15] Our observation that binding of HIV-1 PR occurs by association of two folded monomers opens a new venue for inhibiting the enzyme. Future inhibitors may be designed to incorporate features that complement the surface of the folded monomer and should not be restricted to the N and C termini that are crucial for binding. In particular, the high ϕ values of residues 27–35 and 79–87 at the folding TSE target them as good candidates for designing folding inhibitors. The high ϕ values at the binding TSE found for residues 23–27 and 48–52, which are involved in intermolecular contacts (Figure 4B), directly mark them as good targets for dimerization inhibitors. The fact that the folding nucleus (residues 27–35 and 79–87) is consecutive to an interfacial region (residues 23–27), which includes the catalytic residue, highlights residues 23–35 and 79–87 in the monomeric form as a highly potential candidate for designing a dimerization inhibitor.

Models and Methods

Gō model simulations

The HIV-1 PR (PDB entry 1hhp[40]) was simulated with a C^{α}-based Gō model[41] that takes into account only interactions present in the native structure and therefore does not include energetic frustration (i.e. includes only topological frustration). An interaction between a pair of residues (i, j) exists if the distance between the C^{α} atoms of residues i and j is less than 8 Å or the distance between any side-chain heavy atoms in the two residues is smaller than 4 Å. Native contacts between pairs of residues (i, j) with $|i - j| < 4$ were discarded from the native contact list because contiguous residues already interact through the bond angle and dihedral terms. The total number of native contacts can be divided into monomeric and interfacial contacts. The latter were used to estimate the interface hydrophobicity based on the normalized occurrence of each amino acid in interfacial contacts multiplied by its hydrophobicity factor.[42] The network of native contacts was determined using the CSU software,[43] which gave similar numbers of monomeric and interfacial contacts. A similar analysis was performed for two sets of homodimers that are known to obey either two-state or three-state mechanisms.[21] The structural analysis of the two sets of homodimers together with that of HIV-1 PR is shown in Figure 1B.

The Gō-model has been used to study the folding of many monomers (composed of up to 150 residues) that fold in a two-state or three-state fashion.[44–46] For these proteins, the model, which includes solely topological frustration, is able to predict the existence of intermediates as well as the nature of the transition state ensemble. However, the details of the potential are important to represent adequate stabilities and cooperative folding kinetics.[47] The validity of the model for binding processes has been verified recently by studying the binding mechanism of several protein complexes, including several homodimers, a trimer, and a tetramer.[21,28] For these cases, the model successfully reproduces the known overall experimental binding mechanism with respect to whether stable monomers are needed for binding to take place. Moreover, for two protein complexes, where a ϕ values analysis has been carried out in the laboratory, good agreement was found with the ϕ values based on the Gō model (Y.L. *et al.*, unpublished results).

We use here an off-lattice Gō model, where each residue is represented by a single bead centered on its α-carbon atom (C^{α}) position.[44] Adjacent beads are strung together into a polymer chain by means of a potential encoding bond length and angle constraints. The secondary structure is encoded in the dihedral angle potential and the non-bonded (native contact) potential. The interaction energy U at a given protein conformation Γ is given by:

$$
\begin{aligned}
U(\Gamma,\Gamma_0) = & \sum_{\text{bonds}}^{N-1} K_b(b_i - b_{0i})^2 + \sum_{\text{angles}}^{N-2} K_\theta(\theta_i - \theta_{0i})^2 \\
& + \sum_{\text{dihedrals}}^{N-3} \{K_\phi^{(1)}[1 - \cos(1\times(\phi_i - \phi_{0i}))] \\
& + K_\phi^{(3)}[1 - \cos(3\times(\phi_i - \phi_{0i}))]\} \\
& + \sum_{\text{native contacts}|i-j|>3} \left\{\varepsilon\left[5\left(\frac{r_{0ij}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{0ij}}{r_{ij}}\right)^{10}\right]\right\} \\
& + \sum_{\text{non-native contacts},|i-j|>3} \left(\frac{C}{r_{ij}}\right)^{12} \qquad (1)
\end{aligned}
$$

In the equation, b_i, θ_i, and ϕ_i stand for the ith virtual bond length between ith and $(i+1)$th residue, the virtual bond angle between $(i-1)$th and ith bonds, and the virtual dihedral angle around the ith bond, respectively. The parameters b_{0i}, θ_{0i}, and ϕ_{0i} stand for the corresponding variables at the native structure. In the framework of the model, all native contacts are represented by the 10–12 Lennard Jones form without any discrimination between the various chemical types of interaction. Moreover, both the intra- and intermonomeric contacts (interfacial contacts) are treated in the same way without any bias toward separate folding or toward binding. The r_{ij} and r_{0ij}

are the C^{α}–C^{α} distances between the contacting residues i and j in conformation Γ and Γ_0 (the PDB structure), respectively. In the summation over non-native contacts, C ($= 4.0$ Å) parameterizes the excluded volume repulsion between residue pairs that do not belong to the given native contact set. Here, all temperatures and energies are reported in units of ε. For other parameters, we use similar values that have been used in several folding studies;[44,46,47] namely, $K_b = 100.0$, $K_\theta = 20.0$, $K_\phi^{(1)} = 1.0$, $K_\phi^{(3)} = 0.5$, $\varepsilon = 1.0$. To enhance binding events, a harmonic constraint was applied on the distance between residue 17 and 17′ with $K_{\text{constraint}} = 0.04$.

To probe the nature of the transition state ensemble,[23] we computed the ϕ_{ij} values for a native contact pair between i and j from the probability of formation P_{ij}:

$$\phi_{ij} = \frac{\Delta\Delta F^{\text{TS}-\text{U}}}{\Delta\Delta F^{\text{F}-\text{U}}} \approx \frac{P_{ij}^{\text{TS}} - P_{ij}^{\text{U}}}{P_{ij}^{\text{F}} - P_{ij}^{\text{U}}} \tag{2}$$

where $\Delta\Delta F$ is the free energy difference between the wild-type and mutant protein, P_{ij} is the probability of formation of contact between i and j, and the superscripts F, U, and TS correspond for folded, unfolded, and transition state ensembles, respectively. Because in the Gō model all non-bonded contacts have the same energetics, the ϕ_i value of residue i can be calculated from the contact values, ϕ_{ij}, by averaging all the ϕ_{ij} values that are involved with residue I:

$$\phi_i = \frac{1}{n}\sum_{j}^{n} \phi_{ij} \tag{3}$$

The computational ϕ_i value prediction can be compared with the experimental data once they are available.

All-atom simulations

The structural stability of monomeric HIV-1 PR was studied by all-atom molecular dynamic simulations, where the solvent molecules are represented both by explicit and by implicit solvent models. For each model, three simulations were performed, at 300, 330, and 350 K. The initial conformation in the simulations was a single subunit of the crystal structure of the unliganded protease. In addition, a dimeric HIV-1 PR was simulated with the implicit solvent model to enable a comparison between the monomer and dimer flexibility. The details of the dimer simulation are as described.[6] In the explicit solvent simulations, the protease monomer is solvated by 3105 TIP3P water molecules. Each simulation was performed for 5 ns and a conformation was sampled every 0.5 ps resulting in a total of 10,000 conformations. The implicit solvent simulations were performed with the EEF1 model.[36] Each EEF1 simulation was performed for 100 ns and a conformation was sampled every 4 ps, resulting in a total of 25,000 conformations. In a previous work, shorter molecular dynamics simulations (each of 20 ns) were done with the EEF1 model for both monomeric and dimeric form of the HIV-1 PR.[6] All simulations were performed with the molecular dynamics program CHARMM,[48] and the param19 polar hydrogen force-field[48] using 2 fs timesteps and a 10 Å cutoff for the non-bonding interactions.

Acknowledgements

This work has been funded by the NSF sponsored Center for Theoretical Biological Physics (grants PHY-0216576 and 0225630) with additional support from MCB-0084797. Y.L. acknowledges the support from Rothschild and Fulbright foundations. Computations were carried out at the UCSD KeckII computing facility (supported partially by NSF-MCB). A.C. was supported partially by the Swiss National Competence Center in Research (NCCR) in Structural Biology and the Swiss National Science Foundation.

References

1. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T. *et al.* (1989). Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science*, **245**, 616–621.
2. Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S. *et al.* (1989). X-ray analysis of HIV-1 protease at 2.7 Å resolution confirms structural homology among retroviral enzymes. *Nature*, **342**, 299–302.
3. Ishima, R., Freedberg, D. I., Wang, Y.-X., Louis, J. M. & Torchia, D. A. (1999). Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function. *Structure*, **7**, 1047–1055.
4. Scott, W. R. P. & Schiffer, C. A. (2000). Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. *Structure*, **8**, 1259–1265.
5. Freedberg, D. I., Ishima, R., Jacob, J., Wang, Y.-X., Kustanovich, I., Louis, J. M. & Torchia, D. A. (2002). Rapid structural fluctuations of the free HIV protease flaps in solution: relationship to crystal structures and comparison with predictions of dynamics calculations. *Protein Sci.* **11**, 221–232.
6. Levy, Y. & Caflisch, A. (2003). The flexibility of monomeric and dimeric HIV-1 PR. *J. Phys. Chem. ser. B*, **107**, 3068–3079.
7. Perryman, A. L., Lin, J.-H. & McCammon, J. A. (2004). HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. *Protein Sci.* **13**, 1108–1123.
8. Todd, M. J., Semo, N. & Freire, E. (1998). The structural stability of the HIV-1 protease. *J. Mol. Biol.* **283**, 475–488.
9. Fitzgerald, P. M. D. & Spirnger, J. P. (1991). Structure

and function of retroviral proteases. *Annu. Rev. Biophys. Chem.* **20**, 299–320.
10. Wlodawer, A. & Vondrasek, J. (1998). Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. *Annu. Rev. Biophys. Struct.* **27**, 249–284.
11. Freire, E. (2002). Designing drugs against heterogeneous targets. *Nature Biotechnol.* **20**, 15–16.
12. Zutshi, R., Franciskovich, J., Shultz, M., Schweitzer, B., Bishop, P., Wilson, M. & Chmielewski, J. (1997). Targeting the dimerization interface of HIV-1 protease: inhibition with cross-linked interfacial peptides. *J. Am. Chem. Soc.* **119**, 4841–4845.
13. Bowman, M. J. & Chemielewski, J. (2002). Novel strategies for targeting the dimerization interface of HIV protease with cross-linked interfacial peptides. *Biopolymers*, **66**, 126–133.
14. Boggetto, N. & Reboud-Ravaux, M. (2002). Dimerization inhibitors of HIV-1 protease. *Biol. Chem.* **383**, 1321–1324.
15. Caflisch, A., Schramm, H. J. & Karplus, M. (2000). Design of dimerization inhibitors of HIV-1 aspartic proteinase: a computer-based combinatorial approach. *J. Comput. Aided Mol. Des.* **14**, 161–179.
16. Grant, S. K., Deckman, I. C., Culp, J. S., Minnich, M. D., Brooks, I. S., Hensley, P. *et al.* (1992). Use of protein folding studies to determine the conformational and dimeric stabilities of HIV-1 and SIV proteases. *Biochemistry*, **31**, 9491–9501.
17. Xie, D., Gulnik, S., Gusrchina, E., Yu, B., Shao, W., Qoronfleh, W. *et al.* (1999). Drug resistance mutations can affect dimer stability of HIV-1 protease at neutral pH. *Protein Sci.* **8**, 1702–1707.
18. Ishima, R., Ghirlando, R., Tozser, J., Gronenborn, A. M., Torchia, D. A. & Louis, J. M. (2001). Folded monomer of HIV-1 protease. *J. Biol. Chem.* **276**, 49110–49116.
19. Louis, J. M., Ishima, R., Nesheiwat, I., Pannell, L. K., Lynch, S. M., Torchia, D. A. & Gronenborn, A. M. (2003). Revisiting monomeric HIV-1 protease. *J. Biol. Chem.* **278**, 6085–6092.
20. Ishima, R., Torchia, D. A., Lynch, S. M., Gronrnborn, A. M. & Louis, J. M. (2003). Solution structure of the mature HIV-1 protease monomer. *J. Biol. Chem.* **278**, 43311–43319.
21. Levy, Y., Wolynes, P. G. & Onuchic, J. N. (2004). Protein topology determines binding mechanisms. *Proc. Natl Acad. Sci. USA*, **101**, 511–516.
22. Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. (1997). Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* **48**, 539–594.
23. Onuchic, J. N., Socci, N. D., Luthey-Schulten, Z. & Wolynes, P. G. (1996). Protein folding funnels: the nature of the transition state ensemble. *Fold. Des.* **1**, 441–450.
24. Xu, D., Tsai, C.-J. & Nussinov, R. (1998). Mechanism and evolution of protein dimerization. *Protein Sci.* **7**, 533–544.
25. Papoian, G. A. & Wolynes, P. G. (2003). The physics and bioinformatics of binding and folding—an energy landscape perspective. *Biopolymers*, **68**, 333–349.
26. Caflisch, A. (2003). Folding for binding or binding for folding? *Trends Biotechnol.* **21**, 423–425.
27. Ferrenberg, A. M. & Swendsen, R. H. (1989). Optimized Monte Carlo data analysis. *Phys. Rev. Letters*, **63**, 1195–1198.
28. Levy, Y., Papoian, G. A., Onuchic, J. & Wolynes, P. G. (2004). The energy landscape of protein dimers. *Isr. J. Chem.* **44**, In the press..
29. Fersht, A. R. (1994). Characterizing transition states in protein folding: an essential step in the puzzle. *Curr. Opin. Struct. Biol.* **5**, 79–84.
30. Milla, M. E., Brown, B. M., Waldburger, C. D. & Sauer, R. T. (1995). P22 Arc repressor: transition state properties inferred from mutational effects on the rates of protein unfolding and refolding. *Biochemistry*, **34**, 13914–13919.
31. Taylor, M. G., Rajpal, A. & Kirsch, J. F. (1998). Kinetic epitope mapping of the chicken lysozyme HyHEL-10 Fab complex: delineation of docking trajectories. *Protein Sci.* **7**, 1857–1867.
32. Mateu, M. G., Pino, M. M. S. D. & Fersht, A. R. (1999). Mechanism of folding and assembly of a small tetrameric protein domain from tumor suppressor p53. *Nature Struct. Biol.* **6**, 191–198.
33. Ingr, M., Uhlikova, T., Strisovsky, K., Majerova, E. & Konvalinka, J. (2003). Kinetics of the dimerization of retroviral proteases: the fireman's grip and dimerization. *Protein Sci.* **12**, 2173–2182.
34. Cecconi, F., Micheletti, C., Carloni, P. & Maritan, A. (2001). Molecular dynamics studies on HIV-1 protease drug resistance and folding pathways. *Proteins: Struct. Funct. Genet.* **43**, 365–372.
35. Micheletti, C., Cecconi, F., Flammini, A. & Maritan, A. (2002). Crucial stages of protein folding through a solvable model: predicting target sites for enzyme-inhibiting drugs. *Protein Sci.* **11**, 1878–1887.
36. Lazaridis, T. & Karplus, M. (1999). Effective energy function for protein in solution. *Proteins: Struct. Funct. Genet.* **35**, 133–152.
37. Fan, H. & Mark, A. E. (2003). Relative stability of protein structures determined by X-ray crystallography or NMR spectroscopy: a molecular dynamics simulation study. *Proteins: Struct. Funct. Genet.* **52**, 111–120.
38. Louis, J. M., Clore, M. G. & Gronenborn, A. M. (1999). Autoprocessing of HIV-1 protease is tightly coupled to protein folding. *Nature Struct. Biol.* **6**, 868–875.
39. Veverka, V., Bauerova, H., Zabransky, A., Lang, J., Ruml, T., Pichova, I. & Hrabal, R. (2003). Three-dimensional structure of a monomeric form of a retroviral protease. *J. Mol. Biol.* **333**, 771–780.
40. Spinelli, S., Liu, Q. Z., Alzari, P. M., Hirel, P. H. & Poljak, R. J. (1991). The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. *Biochimie*, **73**, 1391–1396.
41. Go, N. (1983). Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.* **12**, 183–210.
42. Pacios, L. (2001). Distinct molecular surfaces and hydrophobicity of amino acid residues in proteins. *J. Chem. Inf. Comput. Sci.* **41**, 1427–1435.
43. Sobolev, V., Sorokine, A., Prilusky, L., Abola, E. E. & Edelman, M. (1999). Automated analysis of interatomic contacts in proteins. *Bioinformatics*, **15**, 327–332.
44. Clementi, C., Nymeyer, H. & Onuchic, J. N. (2000). Topological and energetical factors: what determines the structural details of the transition state ensemble and En-route intermediate for protein folding? An investigation of small globular proteins. *J. Mol. Biol.* **298**, 937–953.
45. Clementi, C., Jennings, P. A. & Onuchic, J. N. (2000). How native-state topology affects the folding of dihydrofolate reductase and interleukin-1 β. *Proc. Natl Acad. Sci. USA*, **97**, 5871–5876.

46. Koga, N. & Takada, S. (2001). Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. *J. Mol. Biol.* **313**, 171–180.
47. Kaya, H. & Chan, H. S. (2003). Solvation effects and driving forces for protein thermodynamic and kinetic cooperativity: how adequate is native-centric topological modeling? *J. Mol. Biol.* **326**, 911–931.
48. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization and dynamic calculation. *J. Comput. Chem.* **4**, 187–217.
49. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Edited by M. Levitt

(Received 5 December 2003; received in revised form 27 March 2004; accepted 3 April 2004)