PROTEINS: Structure, Function, and Genetics 43:365–372 (2001)

Molecular Dynamics Studies on HIV-1 Protease Drug Resistance and Folding Pathways

Fabio Cecconi,[1,2] Cristian Micheletti,[1,2,*] Paolo Carloni,[1,2,3] and Amos Maritan[1,2]
[1]*International School for Advanced Studies (SISSA/ISAS), Trieste, Italy*
[2]*INFM and The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy*
[3]*International Centre for Genetic Engineering and Biotechnology (ICGEB), Trieste, Italy*

ABSTRACT **Drug resistance to HIV-1 protease involves the accumulation of multiple mutations in the protein. We investigate the role of these mutations by using molecular dynamics simulations that exploit the influence of the native-state topology in the folding process. Our calculations show that sites contributing to phenotypic resistance of FDA-approved drugs are among the most sensitive positions for the stability of partially folded states and should play a relevant role in the folding process. Furthermore, associations between amino acid sites mutating under drug treatment are shown to be statistically correlated. The striking correlation between clinical data and our calculations suggest a novel approach to the design of drugs tailored to bind regions crucial not only for protein function, but for folding as well. Proteins 2001;43:365–372.**
© 2001 Wiley-Liss, Inc.

Key words: HIV-1 protease; drug resistance; folding pathways

INTRODUCTION

The human immunodeficiency virus encodes a protease (HIV-1 PR) that cleaves the gag and the gag-pol viral polyproteins into enzymes and structural proteins.[1] The discovery that inhibition of this protease (a homodimer of 198 amino acids) leads to the formation of noninfectious virus particle has prompted a tremendous effort to design efficient inhibitors against acquired immunodeficiency syndrome (AIDS) attack.[2] Currently, five antiviral agents are approved by the Food and Drug Administration (FDA)—Saquinavir (SQV), Ritonavir (RTN), Indinavir (IND), Nelfinavir (NLF), Amprenavir (APR)—with several others under clinical investigation.[1–4] Therapeutic benefit is short-lived, unfortunately, as the virus strains—evolving under the selective pressure of the drugs—encode HIV-1 PR multiple mutants with low drug affinity[1,4–8]: mutants resistant to protease inhibitors can emerge in vivo after less than 1 year.[5] The problem of drug resistance also persists when a combination of PR and reverse transcriptase (RT) inhibitor is used.

The occurrence of mutations withstanding antiviral drugs is not a mere consequence of drug action; rather, it results from viral replication itself.[5] Indeed, mutations are found irrespective of drug structural diversity involving virtually every protein domain. This is the case for HIV-1 PR mutants, which are resistant against FDA-approved drugs.[4] Among these mutations, a few involve the active site (residue 25–aspartic acid,[6] while the others belong to protein regions away from it.

By means of molecular dynamics (MD) simulations within the framework introduced previously,[9] we show that the positions where these mutations occur play a key role in the proximity of temperatures at which the specific heat peaks occur. Moreover, we shall argue that residues involved in a frequently observed covariant mutation[5] are statistically correlated. Indeed, while the study of the native state structure can lead to a rational design of drugs binding the active site (or that otherwise disrupt the biological function of the agent by acting on its native structure), the analysis of the folding pathways can provide fundamental information.[10] In particular, it can show, as in the case of HIV-1 PR, the presence of kinetic bottlenecks associated with severe entropy reduction that inhibits progress toward the native state. These bottlenecks represent the most delicate part of the folding process. They are followed by the sudden formation of specific native-like protein subregions; afterward, the folding process proceeds rapidly until another bottleneck is reached. The identification of sites involved in the bottlenecks and their correlation with the active site are crucial pharmaceutically because they are the ideal targets of effective drugs. From this point of view, because of the large amount of data available regarding drug resistance, HIV-1 PR is an excellent candidate to validate our automatic strategy to identify key folding sites. The following discussion presents evidence showing that the crucial sites can be identified with good statistical confidence. The framework introduced is general in nature; applied to other viral proteins, it should be useful for suggesting which sites should be preferentially targeted by effective drugs.

THEORY

The strategy adopted to identify the crucial sites for the folding and assembling of HIV-1 PR is based on a recent theoretical framework[9,11] that permits the capture of the

Grant sponsor: Istituto Nazionale di Fisica della Materia; Grant sponsor: Ministero dell'Università e Ricerca Scientifica e Tecnologica.

*Correspondence to: Cristian Micheletti, SISSA, Via Beirut 2A, I-34014 Trieste, Italy. E-mail: michelet@sissa.it.

Received 8 September 2000; Accepted 22 January 2001

© 2001 WILEY-LISS, INC.

main features of the folding process by a simplified description of both the protein structure and the folding dynamics. At the basis of the method is the observation that the topology of the native state plays a crucial role in steering the folding process.[9,12–17] This statement is supported by an increasing amount of experimental evidence. Perhaps, the most notable examples are: (1) the close similarity of the transition-state conformations of proteins having structurally related native states (despite the very poor sequence similarity)[18,19] and (2) the strong influence that certain simple topological properties, such as contact order, have on protein folding rates.[20] Such observations, and others summarized in the recent review by Baker,[21] complement the findings of Anfinsen,[22] who established that the amino acid sequence of a protein uniquely encodes its native state. Indeed, because the topology of the native state influences the folding process, the amino acid sequence must also encode its possible folding pathways.

We focus our attention on the topological rate-limiting steps along the pathways from unfolded states to the native one. Such bottleneck stages, are usually found in correspondence of non-local amino acid interactions that require the overcoming of a large entropy barrier (due to the flexibility of the peptide chain intervening between them); the formation of such crucial contacts acts as a nucleus for the establishment of further native interactions and leads to rapid progress along the folding reaction coordinate until another barrier is met.

It is striking that the sites involved in the topological bottlenecks are those at which the largest changes in the folding kinetics are observed in site-directed mutagenesis experiments,[10] as first established for CI2 and Barnase.[9] This shows that nature has carefully optimised the protein sequence so to exploit the conformational entropy reduction accompanying the folding process[23] through the careful selection of the amino acids that form the crucial contacts.

With the purpose of identifying the key sites we investigate the topological obstacles encountered during the formation of the native HIV-1 PR structure. Such sites are, intuitively, the ideal candidate targets of effective drugs, as they take part to the most delicate steps of the folding process. This fact was first recognized by Anfinsen[22] in connection with the staphylococcal nuclease. The most effective strategy to prevent the protein formation is acting on residues involved in the key contacts and undermining the formation/overcoming of bottleneck stages. One of the distinctive features of the HIV virus is the extremely high rate of mutations. The capability of encoding several mutants provides a possibility for HIV-1 PR to elude the disruptive action of the drug by intervening on the key sites. This seems an unavoidable countermeasure, since the viable mutants (i.e., those with native-like enzymatic activity) retain the original native structure and hence, arguably, encounter the same bottlenecks as the wild-type. Within this framework, the lapse of time during which the drug therapy is temporarily effective corresponds to the time taken by the virus to encode, through random mutations, a mutant form of HIV-1 PR for which the crucial sites have been fine-tuned to overcome not only the kinetic bottlenecks (as for the wild-type) but also the additional drug attack. The key sites identified through the method explained in the next sections, have been compared with the known key mutating positions of HIV-1 PR, finding a highly significant correlation between the two of them. In addition, previously unexplained covariant mutations seen in HIV-1 PR are explained as arising due to the correlation between distinct topological bottlenecks.

METHODS

The model that we adopted encompasses an energy-scoring function of the Go type.[24] This is one of the simplest energy functionals and provides a natural topological bias to the native state by rewarding the formation of native pairwise interactions. In the version used in the present study, which is a generalization of a previous report,[9] applicable to molecular dynamics studies, the cooperativity of the folding process is enhanced by the introduction of repulsive non-native interactions. In our Hamiltonian, each pair of nonconsecutive amino acids interacts with the following strength:

$$5V_0\varepsilon_{ij}^N\left[\left(\frac{r_{ij}^N}{r_{ij}}\right)^{12}-\frac{6}{5}\left(\frac{r_{ij}^N}{r_{ij}}\right)^{10}\right]+V_1(1-\varepsilon_{ij}^N)\left(\frac{r_0}{r_{ij}}\right)^{12} \qquad (1)$$

where $r_0 = 6.8\,\text{Å}$, $r_{i,j}^N$ denotes the distance of C_α atoms of amino acids i and j in the native structure and ε_{ij}^N is the native contact map, whose entries are 1 (0) if i and j are (not) in contact in the native conformation (i.e., below or above 6.5 Å). V_0 and V_1 are constants controlling the strength of interactions ($V_0 = 20$, $V_1 = V_0/400$ in our simulations). In addition, the peptide bond between two consecutive amino acids, i, $i+1$ at distance $r_{i,i+1}$ is described by the unharmonic potential:

$$\frac{a}{2}(r_{i,i+1}-r_d)^2+\frac{b}{4}(r_{i,i+1}-r_d)^4 \qquad (2)$$

with parameters $a = V_0$, $b = 10V_0$, and $r_d = 3.8$ Å is the rest distance between consecutive C_α atoms.

It is important to note that in eq. 1 the formation of any native contact is rewarded in the same way, since V_0 does not depend on i and j. This choice is done deliberately, so that the only information entering eq. 1 is the native contact map and not the types of interacting amino acids (i.e., no sequence information). Although by construction, the minimum of the energy scoring function is achieved in correspondence of the native state, there is no a priori guarantee that the folding process, under the influence the pair interaction in eq. 1, occurs, on average, through the same stages encountered in nature, or even in a more sophisticated atomic MD simulation with ab initio force fields. Certainly, there are situations where the influence of the native-state topology on the folding process may be overridden by strong chemical propensities to form definite pairs of amino acids (e.g., disulfide bridges). In addition, given the explicit bias toward the native state, one should not expect that it would be possible to observe intermediate states with low concentration of native con-

tacts. Aside from similar circumstances, it is appropriate to ask whether one can reproduce the key steps of the folding process by exploiting only the structural information of the native state. The basis and justification for the present study are the growing evidence that the above question has a positive answer. In fact, starting from our previous work,[9] and of refs. 12–17, it has become clear that the characterization of the transition states can be confidently done within a Go-model scheme for a variety of proteins.

In the present study, the starting structural model (target) is the free enzyme,[5] which is a homodimer with each subunit composed by 99 residues (Fig. 1). Following Clementi et al.,[25] the crystallographic C2 symmetry was enforced during MD simulations to reduce the computational effort. The progress toward the fully folded (native) state was estimated in terms of the fraction of native contacts formed at any given time in the partially folded structure, Γ.[26] This quantity, also termed overlap, is defined as

$$Q = \frac{\sum_{i,j} \varepsilon_{ij}^{N} \varepsilon_{ij}^{\Gamma}}{\sum_{i,j} \varepsilon_{ij}^{N}} \qquad (3)$$

where ε^{Γ} is the contact matrix of Γ.

Constant temperature MD simulations were carried out for several decreasing temperatures from the unfolded to folded state of the protein. The equations of motion for the C_α atoms were integrated by a velocity–Verlet algorithm with time step $\Delta t = 0.01$ combined with the standard gaussian isokinetic scheme.[27] We performed unfolding simulations within the same framework by starting from the native structure and taking it through a sequence of increasing temperatures (heat denaturation). The temperature was measured in reduced units V_0/k_B (with k_B the Boltzmann constant, and V_0 the energy of the native contacts in eq. 1). At each temperature, we let the system equilibrate from the last structure reached at the previous run at a lower temperature. Each equilibration involved 5×10^5 MD steps, a time much longer than the largest correlation times observed for the system. After equilibration, we sampled 4,000 structures again at time intervals twice the estimated correlation time. At each temperature, we collected the energy histogram of such uncorrelated structures. Using the multiple histogram techniques,[28] the energy measurements for all temperatures have been reweighted to provide optimal estimates of thermodynamic quantities such as the average energy and the specific heat (by differentiation of the former) for a continuous range of temperatures. The statistical significance of the data collected in our runs was checked by verifying that the reweighted thermodynamic quantities did not change by more than a few percentage points upon addition of energy histograms obtained from folding/unfolding simulations with different temperature schedules or initial conditions.

Within the approximation in which the reaction coordinate is the internal energy, from the high temperature side, the slowest dynamics occurs at temperatures near the specific heat peak. Thus, the contacts contributing more to the specific heat peak are identified as the key contacts belonging to the folding bottleneck and sites sharing them as the most probable to be sensitive to mutations. Furthermore, by following several individual folding trajectories (by suddenly quenching unfolded conformations below the folding temperature, T_{fold}) we ascertained that all such individual dynamic pathways underwent the same kinetic bottlenecks described in the next section.

A reliable and convenient way to identify and characterize the kinetic bottlenecks is through the location of peaks and shoulders in the specific heat (which denote the overcoming of free energy barriers). Moreover, because the specific heat results from the contribution of each pair of native contacts, it is also useful to monitor the formation of each native interaction throughout the folding process. Indeed, the probability of formation of a native contact is a decreasing function of T and has a sigmoidal shape fitted by suitably shifted hyperbolic tangent (Fig. 2). The smooth interpolation allows to identify a crossover temperature, T_0, where the slope reaches its maximum, C_0 (that is the inflection point of the curves in Fig. 2). T_0 defines a local "transition" temperature at which each contact is locked, whereas C_0, which represents the "rapidity" of its formation, can also be regarded as a measure of the local contribution to the specific heat.

RESULTS

At very low temperatures, the observed structures have nearly 100% native-state similarity, measured as the fraction Q of established native contacts (eq. 3). Further increase in temperature causes structural rearrangements into a configuration that cannot be assembled into a dimer anymore (Fig. 3): the number of subunit–subunit contacts vanished and the two subunits behaved independently. The dissociation mechanism is well described in Figure 4, which shows the fractional occupation of native contacts for the individual subunits and at the monomer–monomer interface, for several temperatures. The dissociation is also signaled by an abrupt increase of the specific heat of the dimer (see inset in Figure 5; which defines the the dissociation temperature T_{diss}). A typical structure at this temperature is shown in Figure 3a. At even higher temperatures ($T = 1.4T_{\text{diss}}$), a large increase of specific heat is observed, indicating the presence of a strong transition of the single subunits,[29] see Figure 5. This temperature is identified with the folding temperature, T_{fold}. Consistently with other studies on different proteins, the native overlap at T_{fold} was about 50%. A typical structure at this temperature is shown in Figure 3b.

A further set of bottlenecks is encountered at $T \approx 1.4T_{\text{fold}}$, where the formation of the three β-strands of HIV-1 PR is involved. Upon increasing the temperature, β-sheet β_2 is encountered, followed by β_1, and then β_3. It is found that the kinetic bottleneck for the general β-sheet formation is not the establishment of the contact(s) closest to the β turn (that involves amino acids near in sequence)

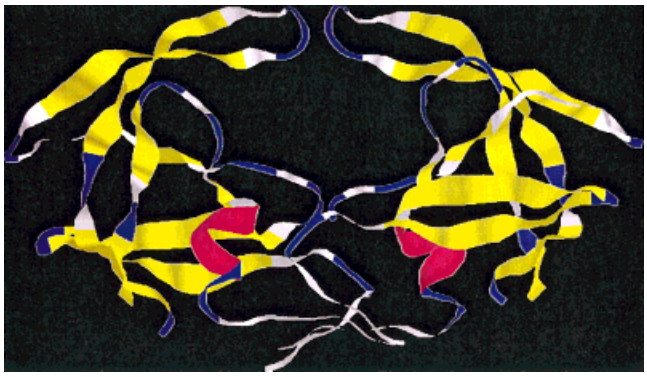

Fig. 1. Structure of HIV-1 PR.[5]

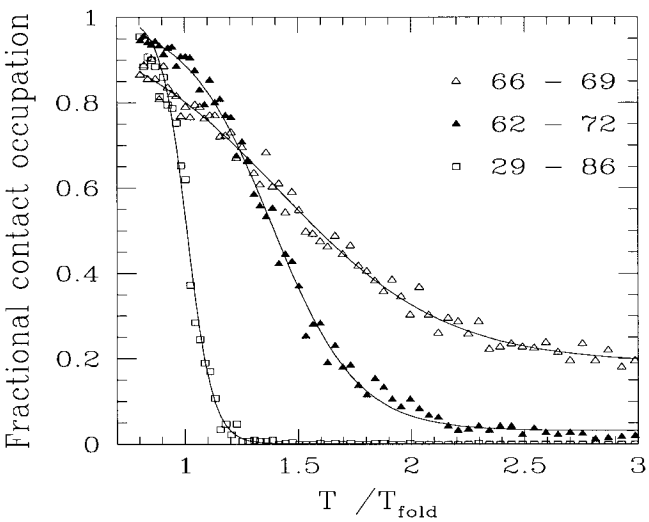

Fig. 2. Fractional occupation for three different native contacts. In all cases the fractional occupation approaches 1 as $T \to 0$ and vanishes for very large T. However, the rapidity of formation (slope at the inflexion point) is very different. The contact binding residues 66 and 69, located at the turn of β sheet 1, forms very gradually. The highest rapidity of formation of contacts in β_1 is observed for the pair 62, 72 (solid triangles). At the folding temperature, one of the highest formation rapidities is found in correspondence of the contact bonding residues 29 and 86 (open squares). The continuous curve is obtained from a smooth interpolation of the points.

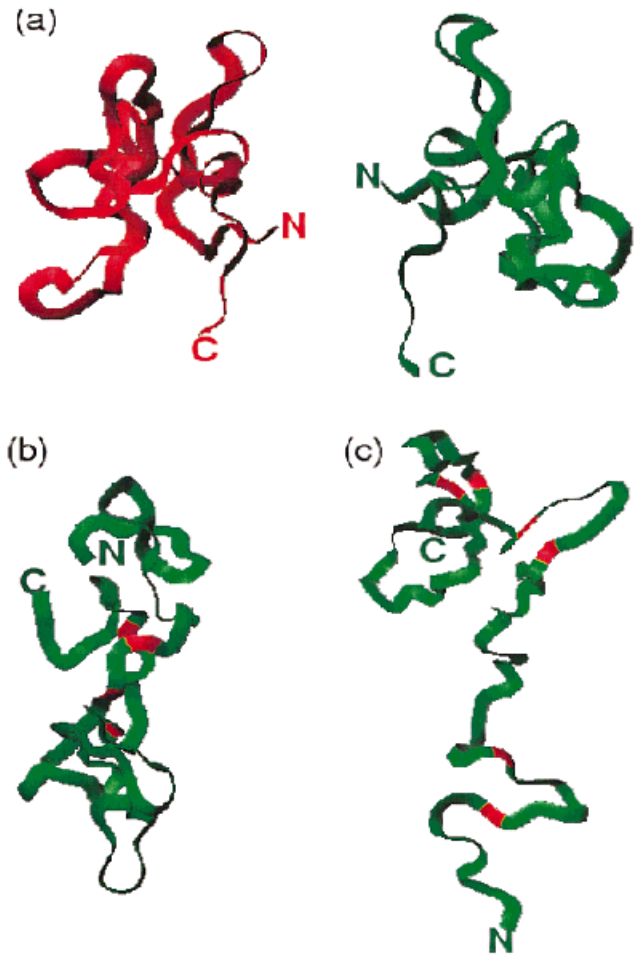

Fig. 3. Typical dimer conformations near the dissociation temperature T_{diss} (a). Typical monomer structures at the folding transition, T_{fold} (b) (key residues 29, 32, 76, 86 are highlighted in red) and at $1.3\,T_{fold}$ (c) (sites 11, 21, 46, 55, 61, 74 responsible for the initiation of the β sheets are shown in red.)

but it is located farther away. A quantitative analysis of the amino acids most involved in the folding bottleneck is again obtained by monitoring during the folding/unfolding process each pair of amino acids, which are in contact in the native state. Examples of the probability with which individual contacts are formed is shown in Figure 2.

At each temperature at which the dynamical evolution of the HIV-1 PR is followed, the formation probability of each native contact (fractional occupation) is calculated. Such quantities are 1 at very low temperatures (all native contacts always present) and decrease to zero at temperatures larger than the folding temperature. It may be anticipated that the rate of decrease as a function of temperature will not be the same for all contacting pairs. In particular, trivial local contacts between residues with a small sequence separation will have a high probability of being formed even at high temperatures. Our interest focuses on those contacts that show a dramatic increase in fractional occupation near the folding transition. Those will be the key contacts responsible for the appearance of the specific heat peak. Examples of the fractional occupation for three native contacts are presented in Figure 3. Given the monotonic behavior of the fractional occupation, the formation of each contacting pair can be synthetically characterized by the temperature at which the point of inflection of the curve is seen and also by the slope at that very same point. Both data can be conveniently summarised in two scatter plots where the slope, C_0, and the temperature of formation, T_0, are reported for each residue taking part in native contacts. Such graphs are reported in the scatter plot of Figure 6. Note that, for each site, there are as many entries as the number of contacts involving it (a number that typically differs from site to site). Figure 6 clearly displays the clusters of contacts that are turned on at similar temperatures.

The bottlenecks for the folding process are identified by isolating the contacts having both a formation temperature, T_0 matching the location of the peaks and shoulders

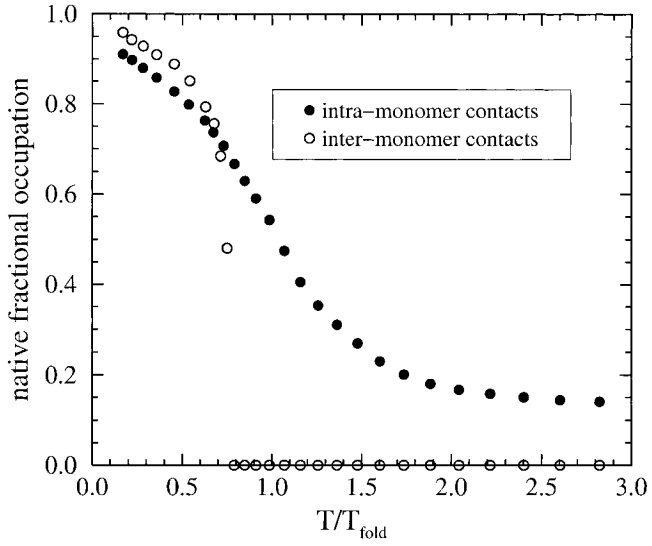

Fig. 4. Behavior, as a function of temperature, of the average fractional occupation of native contacts within each HIV-1 PR monomer (solid circles) and at the interface between the two monomeric units (open circles). The dissociation of the two subunits clearly is seen to occur at $T/T_{fold} = 0.6$.

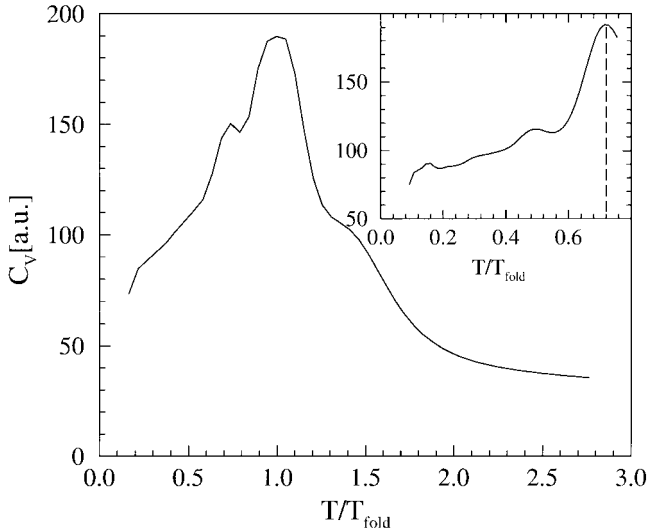

Fig. 5. Specific heat of the HIV-PR single monomer. The dimer specific heat at temperatures below the T_{diss} is shown in the inset.

of the specific heat, and a high rapidity of formation, C_0. Figure 6a shows the presence of four distinct clusters of contacts. The first three, labeled β_1, β_2, and β_3, are associated with the formation of the three antiparallel β-sheets in HIV-1 PR. Their temperature of formation is about $1.4T_{fold}$ and corresponds to the shoulder visible in the larger plot of the specific heat of Figure 5. The sites sharing the most important contacts involved in such three bottlenecks are listed in Table I and highlighted in Figure 3c, where a typical structure at $T = 1.4T_{fold}$ is shown. It is interesting to see that T_0 is maximum for sites close to the β-turn, in accord with the intuitive expectation that the β formation is initiated at the turn. Figure 7a–c presents the values of C_0 only for the pairs of contacting

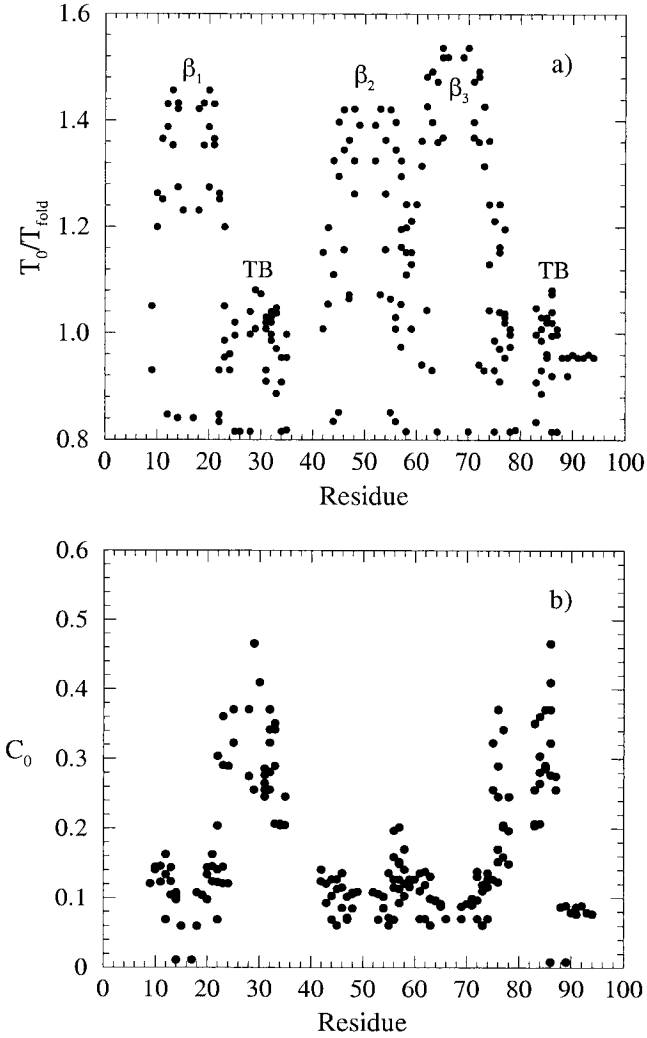

Fig. 6. (a) Characteristic temperatures T_o (a) and maximum "rapidity" C_o associated to each native contact versus the amino acid position sharing the contact. (b) Distribution of the values of the maximum rapidity C_o of contacts involving each residue.

TABLE I. Key Sites for the Four Bottlenecks*

Bottleneck	Key Sites
TB	22,29,32,76,84,86
β_1	10,11,13,20,21,23
β_2	44,45,46,55,56,57
β_3	61,62,63,72, 74

*For each bottleneck, only the sites in the top three pairs of contacts have been reported.

sites in the β-sheet. It is seen that the sites closest to the turn have a small formation rapidity. This can be understood since, being very close along the sequence, they can be easily formed/broken. The highest rapidity, C_0, i.e., the highest difficulty of formation, is encountered typically three to four sites away from the turn. The corresponding contacts are then identified as the bottleneck for this particular folding stage. For the β-sheets, the bottlenecks

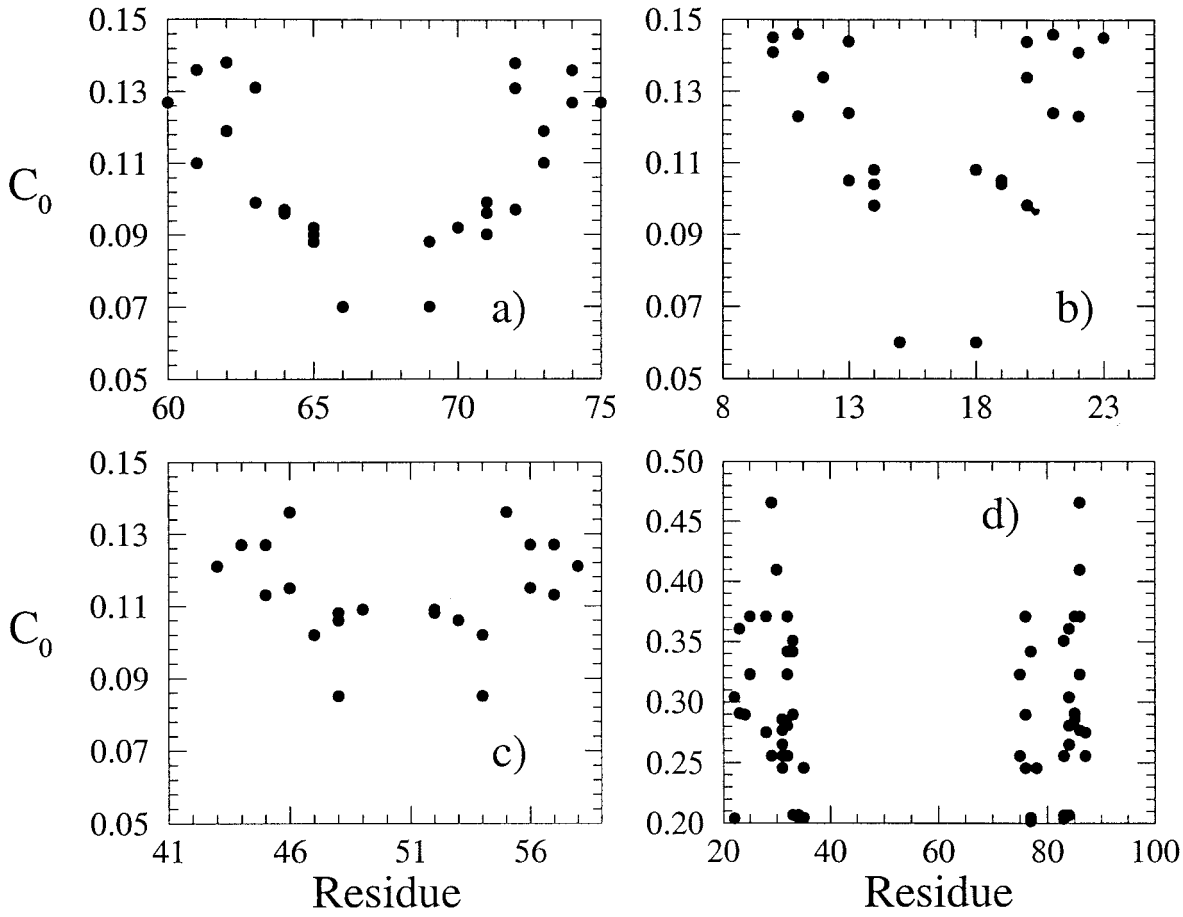

Fig. 7. Formation "rapidity" C_0, for contacts in the four subregions of Fig. 6a. In order of decreasing formation, temperature data are shown for a) β_3, b) β_1, c) β_2 and d) TB. Notice that the scale of C_o depends on temperatures. The highest C_o's for each of the regions increase as the temperature T_o decreases.

involve amino acids that are typically three to four residues away from the turns.

Going back to Figure 5, one observes that there is a fourth group of contacts around residues 30 and 86, labelled TB after "transition bottleneck," that are formed cooperatively at the folding transition. The sites involved in the TB contacts are listed in Table I. Among those contacts we have recorded the largest values of C_0, as shown more clearly in Figure 6d. Again, we considered the sites with the highest values of C_0 as responsible for the main bottleneck of the folding process. The highest "rapidity" is measured in correspondence of contacts 29–86 and 30–84 (see also Fig. 7), which are, consequently, identified as the most crucial for the folding/unfolding process.

DISCUSSION

The sites involved in the main folding bottleneck (TB) are located at the active site of HIV-1 PR, which is targeted by anti-AIDS drugs.[6] Hence, within the limitations of our simplified approach, we predict that changes in the detailed chemistry at the active site also affect a key step of the folding process. To counteract the drug action, the virus has to perform some very delicate mutations corresponding to the key sites. Within a random mutation scheme, this requires many trials (occurring over several months). The time required for the biosynthesis of a mutant with native-like activity is even longer if the drug attack correlates with several bottlenecks simultaneously.

This is certainly the case for several anti-AIDS drugs. Indeed, Table II summarizes the mutations emerged for the FDA-approved drugs.[4] Remarkably, among the first 23 most crucial sites predicted by our method and listed in Table I), there are 7 sites in common with the 16 distinct mutating sites of Table II. The probability that two sets of 16 and 23 sites randomly taken from a total population of 99 (the length of the HIV-1 PR monomer) share at least 7 sites is only 3%. Also note that, all the mutation sites of Table I except 82, 35, 36 and 90 fall within a mismatch of at most one position from the sites of Table II. These results highlight the highly statistical correlation between our prediction and the evidence accumulated from clinical trials.

All mutations causing resistance involve crucial residues for the main folding bottleneck, (particularly residue 84) in combination with key sites for one or more of the β-sheets. Mutation in this sites are expected to modify the energetics and structure of partially folded states. By contrast, the folded state appears to be weakly affected by

TABLE II. Mutations in the Protease Associated With FDA-Approved Drug Resistance*

Name	Point Mutations	Bottlenecks
RTN[30,31]	20,[a] 33, 35, 36, 46,[a] 54, 63,[a] 71, 82, 84,[a] 90	TB, β_1, β_2, β_3
NLF[32]	30, 46,[a] 63,[a] 71, 77, 84[a]	TB, β_2, β_3
IND[5,33]	10,[a] 32,[a] 46,[a] 63,[a] 71, 82, 84[a]	TB, β_1, β_2, β_3
SQV[5,33,34]	10,[a] 46,[a] 48, 63,[a] 71, 82, 84,[a] 90	TB, β_1, β_2, β_3
APR[3]	46,[a] 63,[a] 82, 84[a]	TB, β_2, β_3

[a]Sites involved in the folding bottlenecks as predicted in our approach. β_i refers to the bottleneck associated to the formation of the ith β-sheet, whereas TB refers to the bottleneck that occurs at folding transition temperature T_{fold}.

specific mutations, such as M46I, L63P, V82T, and I84V, which lead to a C_α RMS distance of 0.5 Å from the wild-type.[35,36] In the light of these results, it is possible to interpret the experimental evidence for the existence of correlations between mutations at residue 82 and residues 10, 54, and 71 as correlations between the main kinetic barrier, TB, and the others β_1, β_2, and β_3. The large separation of these associated sites, both along the sequence and in space, suggests that their correlations arise by virtue of the folding process itself. This kinetic effects is particularly clear in one of these cases, namely the comutation of sites in the TB and at residue 10, which occurs under IND therapy. The mediator of the correlation is residue 23 that takes part to two bottlenecks: TB and β_1 through direct contact with residues 84 and 10, respectively. Covarying mutations between the two sites are observed because changes in TB will affect the environment of the other key site 10, which must mutate accordingly.

CONCLUSIONS

The strategy presented allows both to identify the bottleneck of the folding process and to explain their highly significant match with known mutating residues. This approach should be readily applicable to identify the kinetic bottlenecks of other viral enzymes of pharmaceutical interest, aiding the development of novel inhibitor targeting the kinetic bottlenecks. This is expected to enhance dramatically the difficulty with which the virus can express mutated proteins, which continue to fold efficiently into the same native state with unaltered functionality.

REFERENCES

1. Gulnik S, Erickson JW, Xie D. HIV protease: enzyme function and drug resistance. Vitam Horm 2000;58:213–256.
2. Wlodawer A, Erickson JW. Structure-based inhibitors of HIV-1 protease. Annu Rev Biochem 1993;62:543–585.
3. Reddy P, Ross J. A protease inhibitor for the treatment of patients with HIV-1 infection. Formulary 1999;34:567–675.
4. Ala PJ, Huston EE, Klabe RM, Jadhav PK, Lam PY, Chang CH. Counteracting HIV-1 protease drug resistance: structural analysis of mutant proteases complexed with XV638 and SD146, cyclic urea amides with broad specificities. Biochemistry 1998;37:15042–15049.
5. Condra JH, Schleif WA, Blahy OM, Gabryelski LJ, Graham DJ, Quintero JC, Rhodes A, Robbins HL, Roth E, Shivaprakash M, et al. In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. Nature 1995;374:569–571.
6. Brown AJ, Korber BT, Condra JH. Associations between amino acids in the evolution of HIV type 1 protease sequences under indinavir therapy. AIDS Res Hum Retroviruses 1999;15:247–253.
7. Durant J, Clevenbergh P, Halfon P, Delgiudice P, Porsin S, Simonet P, Montagne N, Boucher CA, Schapiro JM, Dellamonica P. Drug-resistance genotyping in HIV-1 therapy: the VIRADAPT randomised controlled trial. Lancet 1999;353:2195–2199.
8. Boucher C. Rational approaches to resistance: using saquinavir. AIDS 1996(suppl 1):S15–S19.
9. Micheletti C, Banavar JR, Maritan A, Seno F. Protein structures and optimal folding from a geometrical variational principle. Phys Rev Lett 1999;82:3372–3375.
10. Fersht AR. Optimization of rates of protein folding: the nucleation–condensation mechanism and its implications. Proc Natl Acad Sci USA 1995;92:10869–10873.
11. Maritan A, Micheletti C, Banavar JR. Role of secondary motifs in fast folding polymers: a dynamical variational principle. Phys Rev Lett 2000;84:3009–3012.
12. Galzitskaya OV, Finkelstein AV. A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. Proc Natl Acad Sci USA 1999;96:11299–11304.
13. Munoz V, Henry ER, Hofrichter J, Eaton WA. A statistical mechanical model for beta-hairpin kinetics. Proc Natl Acad Sci USA 199826;95:5872–5879.
14. Alm E, Baker D. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. Proc Natl Acad Sci USA 1999;96:11305–11310.
15. Chiti F, Taddei N, White PM, Bucciantini M, Magherini F, Stefani M, Dobson CM. Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. Nature Struct Biol 1999;6:1005–1009.
16. Martinez JC, Pisabarro MT, Serrano L. Obligatory steps in protein folding and the conformational diversity of the transition state. Nature Struct Biol 1998;5:721–729.
17. Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. J Mol Biol 2000;298:937–953.
18. Martinez JC, Serrano L. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. Nature Struct Biol 1999;6:1010–1016.
19. Riddle DS, Grantcharova VP, Santiago JV, Alm E, Ruczinski I, Baker D. Experiment and theory highlight role of native state topology in SH3 folding. Nature Struct Biol 1999;6:1016–1024.
20. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 1998;277:985–994.
21. Baker D. A surprising simplicity to protein folding. Nature 2000;405:39–42.
22. Anfinsen CB. Principles that govern the folding of protein chains. Science 1973;181:223–230.
23. Wolynes PG, Onuchic JN, Thirumalai D. Navigating the folding routes. Science 1995;267:1619–1620.
24. Go N, Scheraga HA. On the use of classical statistical mechanics in the treatment of polymer chain conformations. Macromolecules 1976;9:535–542.
25. Clementi C, Carloni P, Maritan A. Protein design is a key factor for subunit–subunit association. Proc Natl Acad Sci USA 1999;96:9616–9621.

26. Chan HS, Dill KA. Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. Proteins 1998;30:2–33.
27. Evans DJ, Hoover WG, Failor BH, Moran B, Ladd AJC. Phys Rev A 1983;28:1016–1021.
28. Ferrenberg AM, Swendsen RH. Optimized Monte Carlo data analysis. Phys Rev Lett 1989;63:1195–1198.
29. Hao MH, Scheraga HA. On foldable protein-like models; a statistical-mechanical study with Montecarlo simulations. Physica A 1997;244:124–145.
30. Molla A, Korneyeva M, Gao Q, Vasavanonda S, Schipper PJ, Mo HM, Markowitz M, Chernyavskiy T, Niu P, Lyons N, Hsu A, Granneman GR, Ho DD, Boucher CA, Leonard JM, Norbeck DW, Kempf DJ. Ordered accumulation of mutations in HIV protease confers resistance to ritonavir. Nature Med 1996;2:760–766.
31. Markowitz M, Mo H, Kempf DJ, Norbeck DW, Bhat TN, Erickson JW, Ho DD. Selection and analysis of human immunodeficiency virus type 1 variants with increased resistance to ABT-538, a novel protease inhibitor. J Virol 1995;69:701–706.
32. Patick AK, Mo H, Markowitz M, Appelt K, Wu B, Musick L, Kalish V, Kaldor S, Reich S, Ho D, Webber S. Antiviral and resistance studies of AG1343, an orally bioavailable inhibitor of human immunodeficiency virus protease. Antimicrob Agents Chemother 1996;40:292–297.
33. Tisdale M, Myers RE, Maschera B, Parry NR, Oliver NM, Blair ED. Cross-resistance analysis of human immunodeficiency virus type 1 variants individually selected for resistance to five different protease inhibitors. Antimicrob Agents Chemother 1995;39:1704–1710.
34. Jacobsen H, Hanggi M, Ott M, Duncan IB, Owen S, Andreoni M, Vella S, Mous J. In vivo resistance to a human immunodeficiency virus type 1 proteinase inhibitor: mutations, kinetics, and frequencies. J Infect Dis 1996;173:1379–1387.
35. Chen Z, Li Y, Schock HB, Hall D, Chen E, Kuo LC. Three-dimensional structure of a mutant HIV-1 protease displaying cross-resistance to all protease inhibitors in clinical trials. J Biol Chem 1995;270:21433–21436.
36. Nair AC, Miertus S, Tossi A, Romeo D. A computational study of the resistance of HIV-1 aspartic protease to the inhibitors ABT-538 and VX-478 and design of new analogues. Biochem Biophys Res Commun 1998;242:545–551.