Design of HIV-1-PR inhibitors that do not create resistance: Blocking the folding of single monomers

RICARDO A. BROGLIA,^{1,2,3} GUIDO TIANA,^{1,2} LUDOVICO SUTTO,^{1,2}
DAVIDE PROVASI,^{1,2} AND FABIO SIMONA^{1,2}

¹Dipartimento di Fisica, University of Milano, 20133 Milano, Italy

²Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Milano, 20133 Milano, Italy

³Niels Bohr Institute, University of Copenhagen, 2100 Copenhagen, Denmark

(RECEIVED June 27, 2005; FINAL REVISION July 13, 2005; ACCEPTED July 17, 2005)

Abstract

The main problems found in designing drugs are those of optimizing the drug–target interaction and of avoiding the insurgence of resistance. We suggest a scheme for the design of inhibitors that can be used as leads for the development of a drug and that do not face either of these problems, and then apply it to the case of HIV-1-PR. It is based on the knowledge that the folding of single-domain proteins, such as each of the monomers forming the HIV-1-PR homodimer, is controlled by local elementary structures (LES), stabilized by local contacts among hydrophobic, strongly interacting, and highly conserved amino acids that play a central role in the folding process. Because LES have evolved over many generations to recognize and strongly interact with each other so as to make the protein fold fast and avoid aggregation with other proteins, highly specific (and thus little toxic) as well as effective folding-inhibitor molecules suggest themselves: short peptides (or eventually their mimetic molecules) displaying the same amino acid sequence of that of LES (p-LES). Aside from being specific and efficient, these inhibitors are expected not to induce resistance; in fact, mutations in HIV-1-PR that successfully avoid the action of p-LES imply the destabilization of one or more LES and thus should lead to protein denaturation. Making use of Monte Carlo simulations, we first identify the LES of the HIV-1-PR and then show that the corresponding p-LES peptides act as effective inhibitors of the folding of the protease.

Keywords: HIV protease; folding inhibition; simplified model; Monte Carlo simulations

Because human immunodeficiency virus type-1 protease (HIV-1-PR) is an essential enzyme in the viral life cycle, its inhibition can control acquired immune deficiency syndrome (AIDS). The main properties inhibitory drugs must display are efficiency and specificity. Conventionally, this is achieved by either capping the active site of the enzyme (competitive inhibition) or binding to some

other part of the enzyme and provoking structural changes that make the enzyme unfit to bind the substrate (allosteric inhibition). All the inhibitors of HIV-1-PR available on the market (Indinavir, Sanquinavir, etc.) and that have been approved by the FDA follow the former paradigm.

The large production of virions in the cell, coupled with the error-prone replication mechanism of retroviruses, leads to escape mutants, drug resistance, and eventually persistence of the disease. Under the selective pressure of drugs, HIV-1-PR either mutates at the active site or at sites controlling its conformation in such a way that the enzymatic activity is essentially retained, while the drug is no longer able to bind to its target. The first signs of the failure of the drug usually take place 6–8 mo

Reprint requests to: Guido Tiana, Dipartimento di Fisica, University of Milano, 20133 Milano, Italy; e-mail: tiana@mi.infn.it; fax: +39-0250317487.

Abbreviations: HIV-1-PR, human immunodeficiency virus type-1 protease; LES, local elementary structure; p-LES, local-elementary-structure-mimicking peptide; RMSD, root mean square deviation.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.051670905>.

after the starting of the treatment (Tomasselli and Henrikson 2000).

We wish to suggest a novel type of HIV-1-PR inhibitor that interferes with the folding mechanism of the protein, thus destabilizing it and making it prone to proteolysis. These molecules are expected to be, aside from highly specific, perdurably efficient. In fact, as we shall see below, drug-induced mutations will necessarily affect sites important for the folding and stability of the protease, and consequently lead to its denaturation.

HIV-1-PR is a homodimer (Fig. 1), a protein whose native conformation is built out of two (identical) disjointed chains. Sedimentation equilibrium experiments have shown that, in a neutral solution (pH 7, 4°C), the protease folds according to a three-state mechanism ($2U \rightarrow 2N \rightarrow N_2$), populating consistently the monomeric native conformation *N* (Xie et al. 1999). This result (see also Appendix A) is supported by NMR studies of mutants in which the interaction across the interface is weakened but the monomer retain its native conformation (Ishima et al. 2001), by all-atom simulations of the HIV-PR monomer in explicit solvent (Levy and Caflisch 2003), and by Gō model simulations of the dimer (Levy et al. 2004a). The dimer dissociation constant ($2N \rightarrow N_2$) is found to be $k_d = 5.8 \mu\text{M}$ at 4°C (Xie et al. 1999): For instance, in a 30 μM solution, 44% of proteins are in monomeric form. This allows one to conclude that, at neutral pH, each monomer of the protein folds following the same hierarchical folding mechanism of single-domain, monomeric proteins (Tiana and Broglia 2002): After the monomer has reached the native state, it diffuses to find another folded monomer with which to associate.

Experimental and theoretical evidence suggests that globular, single-domain proteins avoid a time-consuming

search in conformational space, folding through a hierarchical mechanism. Ptitsyn and Rashin (1975) observed a hierarchical pathway in the folding of Mb (Ptitsyn and Rashin 1975). Lesk and Rose (1981) identified the units that build the folding hierarchy of Mb and RNase on the basis of geometric arguments, deriving the complete tree of events that leads these proteins to the native state. These studies describe a framework in which small units composed of a few consecutive amino acids build larger units that, in turn, build even larger ones, which eventually involve the whole protein (Baldwin and Rose 1990). The kinetic advantage of this mechanism is that at each level of the hierarchy, only a limited search is needed for the smaller units to coalesce into the larger units belonging to the following level (Panchenko et al. 1995).

Lattice model calculations (Broglia and Tiana 2001a; Tiana and Broglia 2001) have shown that the folding of a small monomeric protein, starting from an unfolded conformation, follows a hierarchical succession of events: (1) formation of local elementary structures (LES, containing 20%–30% of the protein's amino acids) stabilized by a few highly conserved, strongly interacting ("hot"), hydrophobic amino acids ($\approx 10\%$ of the protein's amino acids) lying close along the polypeptide chain; (2) docking of the LES into the (postcritical) folding nucleus (FN) (Abkevich et al. 1994), that is, formation of the minimum set of native contacts that brings the system over the major free energy barrier of the whole folding process; and (3) relaxation of the remaining amino acids in the native structure shortly after the formation of the FN. The "hot" sites, which stabilize the LES, are found to be very sensitive to (nonconservative) point mutations. Since most of the protein stabilization energy is concentrated in these sites, mutating one or two of them has a high probability of denaturing the native state. On the other hand, mutating any other site ("cold" sites, even those "cold" sites belonging to the LES) has in general little effect on the stability of the protein (Broglia et al. 1998; Tiana et al. 1998).

Making use of the same model, it has been shown that it is possible to destabilize the native conformation of a protein making use of peptides whose sequences are identical to that of the protein LES (Broglia et al. 2003). Such peptides (p-LES) interact with the protein (in particular with their complementary fragments in the FN) with the same energy that stabilizes the nucleus, thus competing with its formation.

There are two important advantages of these folding inhibitors with respect to conventional ones. First, their molecular structure is suggested directly by the target protein. One need not design or optimize anything, just find the LES of the protein to be inhibited, because the design has been performed by evolution through a myriad of generations of the virus (or of the organism that

Figure 1. The native homodimer of HIV-1-PR. In the monomer on the right we have highlighted the local elementary structures, corresponding to fragments 24–34 and 83–93 (see text).

expresses the protein). Moreover, it is unlikely that the protein can develop resistance through mutations. In fact, the present inhibitor binds to a LES, and a protein cannot mutate a LES (Tiana et al. 1998)—in any case not those “hot” amino acids that are essential to stabilize it as well as to bind to the other LES to form the FN, under risk of denaturation. Note that, within this context, neutral mutations (e.g., hydrophobic–hydrophobic) of these “hot” amino acids are possible, as they do not essentially change the stability of the corresponding LES, nor the strength and specificity with which LES dock to form the FN.

Materials and methods

The investigation of the folding of HIV-1-PR has two goals: (1) to validate the model that will be employed to study the effect of the inhibitor peptides, and (2) to help locate the LES of the protein. For this purpose, use is made of a modified G $\bar{o}$ model. In standard G $\bar{o}$ model calculations (G $\bar{o}$ 1983), such as that carried out by Levy and coworkers (2004a) in their study of the HIV-1-PR, the interaction between each pair of amino acids is described by a square well whose bottom lies at the same energy for all native pairs and at zero or positive energy for non-native pairs. Such a treatment insures the native conformation to be the global energy minimum of the system, providing at the same time a realistic description of the entropic features of the chain. It however fails to provide any chemical characterization of the amino acids, treating all of them on equal footing. To account for the diversity existing among amino acids, we assign to each native pair an interaction energy obtained by averaging a Gromacs (Berendsen et al. 1995) force field around the native conformation of the monomer. This procedure has proven useful to account for the folding properties of a number of small, single-domain proteins (L. Sutto, R.A. Broglia, and G. Tiana, unpubl.).

The model pictures each amino acid as a spherical bead, making inextensible links with the following one. The amino acids interact through a contact potential of the kind

$$U(\{r_i\}) = \sum_{i+2 < j} \left[B_{ij} \theta(R - |r_i - r_j|) \theta\left(R - \left|r_i^N - r_j^N\right|\right) + \epsilon_{HC} \theta\left(0.99 \cdot \left|r_i^N - r_j^N\right| - |r_i - r_j|\right) \times \theta\left(R - \left|r_i^N - r_j^N\right|\right) + \epsilon_{HC} \theta(R - |r_i - r_j|) \left[1 - \theta\left(R - \left|r_i^N - r_j^N\right|\right)\right] \right] \quad (1)$$

where r_i is the coordinate of the C $_{\alpha}$ atom of the i th amino acid, r_i^N is the coordinate in the crystallographic native conformation (pdb code 1BVG), $\theta(x)$ is a Heaviside step function, B_{ij} is the interaction energy between

the i th and j th amino acid, R is the interaction range (which in the following calculations is set equal to 7.5 Å), and ϵ_{HC} is the hard core repulsion, set to 100 kT. Accordingly, the first term of the potential function describes the attraction between native pairs; the second term describes the hard core between native pairs, its range being equal to 99% of the native distance; and the last term describes the repulsion between non-native pairs. Moreover, we assume that residue i interacts with residue $i + 2$ only through a hard core repulsion of range $R = 2$ Å. To determine the quantity B_{ij} , all-atom molecular dynamic simulations were carried out making use of the Gromacs package, treating explicitly the solvent. The simulations were done for 1 nsec at room temperature around the native conformation of the dimer. During this time interval the overall root mean square deviation (RMSD) of the system did not exceed 2.5 Å. The values of B_{ij} are the result of the average of the interaction energies between the different pairs of amino acids over the full simulation.

Of course the present model describes in an approximated way both the geometry of the protein and the interaction between the amino acids. In particular, the matrix elements B_{ij} are calculated in the native state and consequently are expected to describe this state suitably, getting worse as the system departs from it. On the other hand, the unfolded state is stabilized entropically, and one expects that the details of the interaction do not play a relevant role in determining its properties. More critical is the use of the same matrix elements B_{ij} to describe the interaction between a LES and a peptide displaying the same sequence as the complementary LES (see the section on inhibition of HIV-1-PR, below). Although this kind of interaction takes place in an environment different from that in which B_{ij} has been determined, the hydrogen bonds, the electrostatic attraction, and the Van der Waals interaction (in the nonpolarizable approximation used in Gromacs) depend mainly on the kind of amino acids, and less on the environment. The hydrophobic interaction, on the other hand, is much more dependent on the environment, and consequently this represents the cruder approximation in the model.

The lack of non-native interactions also is a limitation of the present model, but a large number of studies (for examples, see Levy and Caflisch 2003, Levy et al. 2004a, and references therein) have obtained realistic results both for the folding of single-domain proteins and for dimers, as a consequence of the minimal frustration principle (Levy et al. 2004b). Note, however, that extensive conformational samplings to obtain equilibrium properties of the protease can be performed only with very simplified models, while more realistic all-atom models of the kind of Gromacs are computationally too demanding for our purposes. The key physical feature we want to describe is

the competition between the binding of two complementary LES of the protein and between a LES and a peptide that displays the same sequence as the complementary LES. For this purpose, the exact interaction among residues is not crucial, provided that the FN displays a stabilization energy lower than the rest of the protein.

The simulations were performed, making use of the modified G $\bar{o}$ model, in a 100-Å box with periodic boundary conditions (equivalent to a 1.6 mM concentration), and temperatures ranging from 1 to 5 kJ/mol (temperatures will be expressed in kJ/mol, setting Boltzmann's constant equal to 1; for instance, 300 K corresponds to 2.5 kJ/mol). From these simulations, characterization of the dimer's thermodynamic quantities has been obtained by means of a modified multi-histogram technique (Borg 2001). The resulting dimer's specific heat (per monomer) is displayed in Figure 2A (solid curve). Consistent with the findings of Levy et al. (2004a), it displays two peaks, one at $T_1 = 2.8$ kJ/mol and one at $T_2 = 4.1$ kJ/mol.

To better quantify the properties of the protein we introduce the parameter q_E , defined as the fraction of the native energy (e.g., $q_E = 1$ means that the dimer is in the native conformation). We will use the parameter q_E also with respect to the monomer or to the interface, to indicate the fraction of energy within the monomer or at the interface of a given conformation. Figure 2B displays the value of the parameter q_E associated with the

interaction within each of the monomers forming HIV-1-PR (black dashed curve), as well as that associated with the dimerization, that is, the interaction between the monomers (dashed gray curve). The decrease in the interaction energy at the interface between the two monomers taking place at T_1 indicates that the associated peak in the specific heat of the two chains (continuous curve in Fig. 2A) marks the transition between the dimeric and the monomeric forms of the protein. Just above T_1 , each monomer is still in the native basin of attraction, displaying a $q_E \approx 0.75$. The temperature T_2 corresponds to the (weak) transition to unfolded monomers ($q_E < 0.5$; dashed black curve in Fig. 2B).

The same kind of weak transition is found for simulations of an HIV-1-PR monomer alone (dashed curve in Fig. 2A), although at a slightly lower temperature ($T_f^{\text{mon}} = 3.8$ kJ/mol). The weakness of the (monomer) folding transition (taking place at T_2) is associated with a faint degree of cooperativity, as testified by the low value assumed by the two-state parameter $\kappa_2 = 0.18$ (Kaya and Chan 2000). This parameter ranges from one for fully cooperative transitions to zero for noncooperative transitions. Although it is well known that simplified models underestimate the cooperativity of the folding transition (Li et al. 2004), the HIV-1-PR monomer displays a value of κ_2 , which is much lower than that of other proteins simulated with the same model (e.g., src-SH3 displays $\kappa_2 = 0.38$, even though it is shorter than HIV-1-PR).

The physical reason why the folding of the HIV-1-PR monomer is much less cooperative than other monoglobular proteins can be found in the properties of the FN of each of the monomers forming this protein. As discussed below, the FN of the protease is built out of the LES containing the monomers 24–34, 83–93, and 75–78. Due to the distance between the three LES along the chain, the assembly of the FN leaves a conformational freedom to the rest of the chain (specifically, to the fragment 35–75 at least, and likely also to the fragment 35–83) uncommon among other proteins. This is testified to by the large equilibrium RMSD found under folding conditions (up to 10 Å for $T < 3.8$ kJ/mol) in simulations of the monomer alone (dashed curve in Fig. 2C), where the FN is essentially formed (dashed-dotted black curve in Fig. 2B). Within this context, we note that the basin of attraction of the monomeric native state extends to conformations with an RMSD of 10 Å, corresponding to a typical q_E of 0.7 (see Fig. 7A, below), while the unfolded states have a RMSD on the order of ≥ 12 Å. The large fluctuations of the fragment 35–75 (or 35–83) when the nucleus is formed produce a shoulder in the specific heat at low temperatures (dashed curve in Fig. 2A) and blur the folding transition at T_f^{mon} .

Figure 2. (A) The specific heat of the dimer (solid curve; the values obtained from the simulations have been divided by two in order to obtain the specific heat per monomer) and of the monomer (dashed curve); (B) the order parameter q_E associated with the contacts within monomers (dashed black curve), across the dimer (dashed gray curve), and within the nucleus (contacts between fragment 22–32 and 83–93) of the monomer (dot-dashed black curve); (C) the RMSD associated with the whole dimer (solid curve) and with the monomer alone (dashed curve).

With the same model it is possible to run dynamic simulations, starting from random conformations and following the folding of the protein into the native state. Since we are interested in inhibiting the folding of the HIV-1-PR monomer, we will concentrate on the dynamics of the monomer. The overall dynamics can be followed through the plot of $[q_E](t)$, the fractional native energy as a function of time, averaged over 100 independent runs. The result at $T = 2.5$ kJ/mol is displayed as a solid curve in Figure 3, indicating an exponential process of characteristic time $\tau = 2.9 \times 10^{-7}$ sec (see Appendix B), consistent with the two-state picture. The model also provides information concerning the formation of each contact, through the probability $p_{ij}(t)$ that the contact between residues i and j is formed at time t . A number of contacts are stabilized early (sub-nanosecond time scale), following exponential dynamics. This is the case of, for example, contact 87–90. Contacts between residues that are far along the chain are formed later, following non-exponential dynamics, which indicates that their formation is dependent on some other event. The earliest among them involve fragments 22–34 and 77–93 after an average time $\tau \approx 10^{-7}$ sec. As an example, Figure 3 displays the formation probability of the contact 31–89.

Figure 4 summarizes the hierarchy of formation of native contacts of HIV-1-PR (the different gray levels corresponding to different time scales), while Table 1 lists the parameters associated with selected contacts. The picture that emerges is that local contacts within fragments 83–93 and in the β -hairpin 42–58 form

Figure 3. Dynamic evolution of the monomer. The formation probability $[q_E](t)$ of the native conformation (black curve) at $T = 2.5$ kJ/mol, the probability $p_{87-90}(t)$ (dashed gray curve rising steeply from 0–1), and the probability $p_{31-89}(t)$ (dotted gray curve).

Figure 4. The contact map of HIV-1-PR, where different gray levels correspond to different values of the average stabilization time. Darker and lighter symbols correspond to times of the order of 10^{-10} sec and 10^{-7} sec, respectively. The numbers on the abscissa and the ordinates indicate the site number of the 99mer.

first. Note also the very fast formation of contact 25–28 belonging to fragment 22–34. Then the β -turns 14–19 and 64–72 form, again built out of local residues. The next event is the assembly of the FN involving fragments 22–34 and 83–93, which is further stabilized by the contribution of the strongly interacting segment 75–78. Finally, the rest of the residues come into place.

Summing up, the model suggests that LES are built of residues that lie in the regions 83–93, 24–34, and likely also 75–78. This result essentially agrees with the indications provided by the studies of Levy et al. (2004a). These authors have found that the groups of amino acids 27–35 and 79–87 are essential in the folding of the protein.

Localization of the LES: Indirect methods

The direct inspection of the dynamics of HIV-1 PR obtained by means of model simulations of the folding has suggested which are the LES controlling the folding of the monomer. It can be interesting to study other techniques to localize the LES, in order both to substantiate the findings of the model and to develop a more economical method to inhibit other proteins.

Table 1. The dynamics of some native contacts of the protein

Contact	B_{ij}	p_{∞}	τ
25–28	–5.4	0.77	2×10^{-10}
87–90	–9.8	0.99	4×10^{-10}
31–89	–6.1	0.92	2.6×10^{-7}
23–85	–4.9	0.86	3.6×10^{-7}
62–74	–2.4	0.51	1.0×10^{-6}
12–66	–1.5	0.51	1.2×10^{-6}

B_{ij} is the interaction energy expressed in kJ/mol, p_{∞} is the asymptotic stability, and τ is the average formation time of the contact in seconds.

Evolutionary data

Since LES are responsible for guiding a protein to its native conformation, it is likely that evolution pays particular care in conserving their sequence. In fact, model simulations have shown that residues building the LES can undergo only conservative point mutations (e.g., hydrophobic–hydrophobic), at the risk of denaturing the protein (Tiana et al. 1998). Consequently, LES are highly conserved in families of structurally similar proteins (Tiana et al. 2000). Comparative studies of a number of protein families have shown that this is indeed the case (Mirny and Shakhnovich 1999).

A measure of the degree of conservation of residues in a family of proteins is provided by the entropy per site $S(i) = -\sum p_i(\sigma) \log p_i(\sigma)$, where $p_i(\sigma)$ is the frequency of appearance of residues of type σ at site i in the proteins belonging to the family. In order to be statistically meaningful, we have plotted in Figure 5 (with a solid line), the entropy calculated over a family of 28 uncorrelated proteins (i.e., sequence similarity < 25%) structurally similar to HIV-1-PR (Holm and Sander 1996). As seen from the figure, the most conserved regions are those involving residues 22–33 and 81–90. Even disregarding the statistical caution and calculating the entropy over 462 proteins (Sander and Schneider 1991) displaying any sequence similarity to HIV-1-PR (dashed line in Fig. 5) the plot indicates the same regions as the most conserved. Note that the conservation of residues 25–27 is anyway not unexpected, in that they build the active site of the protease (D25, T26, and G27).

Another important source of information concerning HIV-1-PR is provided by the study of its sequences in specimens coming from infected individuals. Being a retrovirus, HIV can replicate very fast but very imprecisely, thus displaying a rather fast evolution rate. This evolution is reflected by the appearance of many mutated HIV-1-PR that retain their folding features. Table 2 lists the mutations observed in 28,417 isolates coming from patients infected with HIV-1-PR (Shafer et al. 1999). Since some mutations can be conservative, that

is, substitute an amino acid with another displaying similar chemical properties, we also display on the table the lowest PAM250 (Pearson 1990) score associated with the mutations in each site. A positive PAM250 score indicates a conservative mutation. Although the fragments with no mutations or with only conservative mutations are too many to let one identify the LES from this information alone (probably because of the limited statistics in the database), the fact that only conservative mutations fall in the fragments 24–34, 83–93, and 75–78 is consistent with the description of the FN made above.

Static energy features

In order to become stable at an early stage of the folding process, LES must carry a significant fraction of the total energy of the protein in its native conformation. We have analyzed the native interaction B_{ij} between the amino acids following the scheme described in Tiana et al. (2004), through an eigenvalue decomposition of the matrix. The lowest eigenvalue ($\lambda_1 = -121.2$ kJ/mol) displays a large energy gap (i.e., -12.9 kJ/mol ≈ 5 kT) with respect to the next eigenvalue, indicating a core of strongly interacting amino acids. We show in Figure 6 the eigenvector associated with the lowest energy eigenvalue, which highlights to what extent the different amino acids participate in this core. The largest amplitudes involve residues 25–34, 57–65, 75–77, and 83–90. These regions of residues overlap well with the conserved regions mentioned above in connection with Figure 5.

Figure 5. The entropy per site (see text) of proteins structurally similar to the HIV-1-PR monomer (PDB code 1BVG). The solid line indicates the entropy function calculated over 28 proteins displaying sequence similarity < 25% with HIV-1-PR, while the dashed line is associated with 462 proteins irrespective of sequence similarity.

Table 2. The observed mutations in protease reported in Shafer et al. (1999)

wt	mut	PAM	wt	mut	PAM	wt	mut	PAM	wt	mut	PAM
P1			T26			G51			L76	V	2
Q2			G27			G52			V77	I	4
I3	V	4	A28			F53	L	2	G78		
T4			D29			I54	VMLT	0	P79	A	1
L5			D30	N	2	K55	RH	0	T80		
W6			T31			V56			P81	T	0
Q7			V32	I	4	R57	K	3	V82	TAFIS	-1
R8	KQL	1	L33	FVI	2	Q58	E	2	N83		
P9			E34	DQANG	0	Y59			I84	V	4
L10	FIRV	-3	E35	DG	0	D60	E	3	I85	V	4
V11	IL	2	M36	ILV	2	Q61	ENH	1	G86		
T12	SPAEIKN	0	S37	DSTEKHC	-4	I62	V	4	R87	K	
I13	V	4	L38	F	2	L63	PSTACQH	-6	N88	DS	1
K14	R	3	P39	SQT	0	I64	VLM	2	L89	MVI	2
I15	V	4	G40			E65	D	3	L90	M	4
G16	EA	0	R41	KN	0	I66	FV	1	T91		
G17	E	0	W42			C67	FS	-4	Q92	KR	1
Q18	H	3	K43	RT	0	G68			I93	L	2
L19	ITVQP	-3	P44			H69	KQYRN	0	G94		
K20	MRTIV	-2	K45	IRN	-2	K70	RTE	0	A95	SF	-3
E21			M46	FILV	0	A71	TVI	-1	T96		
A22			I47	V	4	I72	VTLMER	-2	L97	V	2
L23	I	2	G48	V	-1	G73	STCA	1	N98		
L24	IVF	2	G49			T74	SAP	0	F99		
D25			I50	VL	2	V75	I	4			

For each residue of the wild-type sequence (wt) are listed the mutations observed in treated or untreated patients (mut) and the lowest PAM250 score (see text) among the mutations. In bold are sites that undergo nonconservative (i.e., non-positive) mutations.

Another way of representing the interaction matrix B_{ij} is to consider the interaction between fragments S_1 (13–21), S_2 (24–34), S_3 (38–48), S_4 (50–55), S_5 (56–66), S_6 (67–72), S_7 (75–78), and S_8 (83–93). The corresponding 8×8 energy map is essentially codiagonal; the associated energy of the interacting chain segments being -1936 kJ/mol as compared with the molecular dynamics simulation native energy -2722 kJ/mol. In other words, the S_1 – S_8 representation of the folded monomer accounts for $\approx 70\%$ of the calculated native conformation energy. Because this representation contains 70 residues, it would be tempting to conclude that the native energy is uniformly distributed over all the amino acids. Within this context, it is useful to calculate the difference between the native S_1 – S_8 energy map and that associated with the unfolded (U) state ($q_E < 0.3$). For this purpose, $G\bar{o}$ model simulations have been carried out to obtain a statistically representative ensemble of U states, from which we have extracted an average set $\{q_{ij}\}_U$ of similarity parameters. Weighting each contribution to the elements of the 8×8 matrix by the corresponding difference $(q_N - q_U)_{ij}$, one obtains the energy map shown in Table 3. It is seen that the S_1 – S_8 representation divides into three blocks: 1) one composed of segments S_2 , S_7 , and S_8 ; 2) one composed of segments S_1 and S_6 ; and 3) one containing segments S_3 , S_4 , and S_5 . While the average energy

per residue in those blocks is -5.3 kJ/mol, those associated with $S_2 + S_7 + S_8$ and $S_1 + S_6 + S_3 + S_4 + S_5$ are -8.3 kJ/mol and -4.9 kJ/mol, respectively.

The above results indicate that the S_2 , S_7 , and S_8 segments qualify as LES (folding units) of each of the two monomers of the HIV-1-PR dimer, LES which form in their native conformation the FN. It is interesting to note that drug-induced mutations in the amino acids belonging to these LES (L24I, D30N, L33F, V77I, I84V, I85V,

Figure 6. The components of the eigenvector associated with the lowest eigenvalue of the interaction matrix B_{ij} .

Table 3. The all-atom molecular dynamics energy map of the eight amino acid chain segments expressed in terms of the folding nucleus ($S_2 + S_7 + S_8$), the flap region ($S_3 + S_4 + S_5$), and ($S_1 + S_6$)

	$S_1 + S_6$	$S_2 + S_7 + S_8$	$S_3 + S_4 + S_5$
$S_1 + S_6$	-67.2	-72.6	-19.2
$S_2 + S_7 + S_8$	-72.6	-223.6	-117.2
$S_3 + S_4 + S_5$	-19.2	-117.2	-144.1

N88D, and L90M) (Tomasselli and Heinrikson 2000) lead to a FN energy equal to -1500 kJ/mol, as compared with -796 kJ/mol for the wild-type sequence FN, an increase of almost a factor of two in the stability of the system, considering the effect of all mutations simultaneously. In fact, single selected mutations add to the FN stability 5–10 kJ/mol. Within this context and that of Figure 6 one can identify sites 33, 75, 76, 85, and 89 as “hot” sites.

Further evidence

Wallqvist and coworkers (1998) investigated HIV-1-PR for the occurrence of cooperative folding units that exhibit a relatively stronger protection against unfolding than other parts of the molecule. Unfolding penalties are calculated forming all possible combinations of interactions between segments of the native conformation and making use of a knowledge-based potential. This procedure identifies a folding core in HIV-1-PR comprising residues 22–32, 74–78, and 84–91, residues that form a spatially close unit of a helix (84–91) with a sheet (74–78) above another β -strand ([22–25], containing the active site residues D25, T26, and G27) perpendicular to these elements.

Making use of a Gaussian network model, Bahar and coworkers (1998) studied the normal modes about the native conformation of HIV-1-PR. “Hot” residues, playing a key role in the stability of the protein, are defined as those displaying the fastest modes. In this way regions 22–32, 74–78, and 84–91 are identified as the folding core of the protein. These regions match with those displaying low experimental Debye-Waller factors, that is, low fluctuations in the crystallographic structure.

Calculation of ϕ -values by means of $G\bar{o}$ model simulations performed by Levy and coworkers (2004a) has located a major transition state in which only regions 27–35 and 79–87 are structured. The protein then reaches the native state, overcoming another minor transition state, and subsequently dimerizes into the biologically active structure.

Cecconi and coworkers (2001) have calculated the stability temperatures associated with each contact of

the protease, again making use of a $G\bar{o}$ model. They find that key sites to the stability of partially folded states, that is, those displaying the lowest stability temperatures, are 22, 29, 32, 76, 84, and 86.

Results and Discussion

The central issue of the present work is to show that it is possible to destabilize the native state of the HIV-1-PR monomer, shifting the equilibrium to the unfolded state, by means of short peptides displaying the same sequence as one of the LES (which we shall call p-LES). As emerged from the calculations and evidence presented above, segments 24–34 (S_2), 83–93 (S_8), and likely 75–78 (S_7) qualify as LES of the HIV-1-PR monomer and thus as leads of inhibitors of the enzyme, with the following provisos. Segment S_7 is a so-called open LES (Broglia and Tiana 2001b), too short to be specific. Concerning the S_2 LES, it contains the active site (residues 25–27). In model studies of the design of good folders, neither we nor any other group has ever considered the role the conserved amino acids belonging to the active site play in the resulting sequence, nor in the folding properties of the protein. Nonetheless one knows that this role is likely to be important. In particular, while considerations of hydrophobicity and/or capability to establish the largest number of native contacts suggest that the most strongly interacting amino acids providing stabilization of the LES and thus of the FN should be, in the native conformation, well protected and buried inside the protein, those associated with the active site should be reachable by the substrate and thus lay on the surface of the enzyme. The corresponding frustration is well exemplified by the anticorrelation observed between the entropy values associated with sites 26 and 27 (low) and the corresponding eigenvector components (also low) shown in Figures 5 and 6, respectively. This anticorrelation becomes even stronger if one compares it with the perfect correlation existing between the values of the entropy (low) and of eigenvector (high) associated with sites 33 and 85 essential in the folding process (“hot” sites) but not connected with the active site. Summing up, we do not know what the consequences are of this frustration in the design of nonconventional inhibitors. Consequently, in what follows we shall exclusively concentrate on LES S_8 and on the inhibitory properties of the peptide p- S_8 .

In the model calculations, the residues of a p-LES interact with other residues of the same p-LES, with residues of the other p-LESSs, and with residues of the protein through the same matrix elements that control the interaction between the corresponding LES (i.e., the LES displaying the same sequence as the p-LES under consideration) and the rest of the protein. For example,

since residues 87 and 90, belonging to LES S_8 , build a contact in the native conformation of the HIV-1-PR monomer, they interact with energy B_{87-90} (cf. Equation 1). Also, residues 87' and 90' belonging to the peptide p- S_8 , residues 87' and 90' belonging to two different p- S_8 peptides, and residues 87' and 90 belonging to p- S_8 and to the protein, respectively, are considered to interact through the same matrix element B_{87-90} . This extension of the $G\bar{o}$ model seems quite realistic, in spite of the crudity of the $G\bar{o}$ model itself, in that a p-LES is chemically identical to the corresponding LES and consequently the interactions made with the rest of the system are expected to be the same. Even if the matrix element B_{87-90} was affected by a consistent error, the model contains the key ingredient that a p-LES interacts with the complementary LES with the same energy that stabilizes the two LESs, and thus will compete with it, destabilizing the native state. On the other hand, interactions among p-LES are not so well described, and are, in general, underestimated by the model. Consequently, the p-LES could have a larger propensity to aggregate than what is predicted by the present calculations.

We have performed equilibrium simulations of the system composed of the HIV-1-PR monomer and a number of p-LES (p- S_8) corresponding to the fragment 83–93 of the protein (S_8). The joint probability distribu-

tion $p(q_E, \text{RMSD})$ of the native relative energy fraction q_E and of the normalized (i.e., divided by the number of residues) RMSD for the case of the monomer plus three p-LES at $T = 2.5$ kJ/mol is displayed in Figure 7B, as compared with that of the monomer alone in Figure 7A. Consistent with the folding transition observed in Figure 2 and with the discussion in the Materials and Methods section, we define as the native state the region of Figure 7A characterized by $q_E > 0.7$ and $\text{RMSD} < 10$ Å (this region is delimited with a dashed curve in the figure). The effect of p- S_8 is to decrease drastically the population of the native state and increase at the same time that of the unfolded state.

The increase of the peak associated with the unfolded state is caused by the appearance of conformations where the p- S_8 peptides are bound to fragment 24–34 (S_2) of the protein, preventing the actual S_2 LES from finding its native conformation. An example of such a conformation is shown in Figure 8, corresponding to the values $q_E = 0.6$ and $\text{RMSD} = 11$ Å. This conformation is particularly stable because the interaction between the p-LES and the monomer is on the order of -165 kJ/mol, the same amount of energy that stabilizes the nucleus of the protein. Note also that more than one p- S_8 is able to bind at the same time to the monomer, increasing the degree of denaturation.

Figure 7. The equilibrium probability of the HIV-1-PR monomer as a function of the energetic parameter q_E and of the RMSD for the monomer alone (A), of the system composed of the monomer and three peptides p- S_8 (B), and of the monomer and three peptides corresponding to the sequence 5–15 control peptide (C). The dashed curve indicates the native state.

Figure 8. A snapshot of an unfolded conformation taken from the simulation of Figure 7, in which the peptides p-S₈ prevent the FN of the monomer to form.

From the above Monte Carlo (MC) simulations it is possible to estimate the inhibition constant K_i of p-S₈ at $T = 2.5$ kJ/mol, assuming that the different MC steps represent different replicas of the system. From the number $N_F = 31.7 \times 10^9$ of MC steps in which the protein is folded and the number $N_I = 3 \times 10^8$ in which it is inhibited, one finds $K_i = n_p N_F^2 / (V N_T N_I)$, where $n_p = 3$ is the number of p-LES, $V = 100 \text{ \AA}^3$ is the volume of the box, and $N_T = N_I + N_F$ is the total number of MC steps of the simulation. One obtains $K_i = 0.5 \text{ \mu M}$, which is comparable to that of FDA-approved drugs.

Figure 9 displays the equilibrium population p_N of the native state as a function of the number n_p of p-S₈ peptides. The inhibitory effect of the peptides is already present at $n_p = 1$ (i.e., a concentration of p-S₈ equal to the concentration of protein, in terms of number of chains), in which case the stability of the native state is reduced by $\approx 30\%$ with respect to the situation with no peptides, while for $n_p = 4$, the value of p_N becomes ≈ 0.25 . We have repeated the same calculation with control peptides whose sequences are equal to those of fragments 5–15 and 61–70 of the monomer. In all cases a slight decrease in stability has been observed. However, the control peptides are not able to disrupt to any extent the FN and thus prevent the monomer from reaching the native state.

We have also calculated how inhibition depends on temperature, carrying out simulations of folding of the monomer protein in the presence of three p-S₈ at two temperatures different from that used in connection with the results displayed in Figure 9. At $T = 3.5$ kJ/mol, the value of p_N drops to 0.01, while at the (very) low tem-

perature $T = 2$ kJ/mol, the simulation leads to $p_N = 0.98$, indicating that p-S₈ becomes ineffective. Within this context, the following considerations are in place. At the “biological” temperature $T = 2.5$ kJ/mol, our simulations are well equilibrated, in the sense that the system has gone in and out the native conformation a large number of times and the system has lost memory of the initial condition (e.g., simulations starting from the native conformation give the same results as simulations starting from a random, unfolded conformation). At $T = 2$ kJ/mol this is no longer true. In fact, since in this case the dynamics are much slower, it is not feasible to perform simulations for a long enough time for the system to become equilibrated.

In any case, the low-temperature simulations performed give a strong signal about a diminished inhibitory effect of p-S₈. This is not unexpected, because the preference the protein has to bind a p-LES instead of its own LES is mainly entropic. More precisely, the protein can either make its S₂-LES and S₈-LES interact to build the FN (in which case the protein folds) or make its S₂-LES interact with the p-S₈ peptide (in which case the protein does not fold). The difference between the two situations can thus be understood as follows. In the case where two LES bind to each other in the native conformation, the whole system gains potential energy from the folding of the rest of the protein, paying at the same time the entropic cost associated with this phenomenon. In the case in which a LES binds a p-LES, the system essentially gains the same potential energy as in the previous case, due to the fact that most of the stabilization energy is concentrated in the interaction among the LES. On the other hand, the entropy cost is only that associated with freezing out the degrees of freedom of the peptide. This entropic cost decreases with the num-

Figure 9. The equilibrium population $p_N \equiv p(q_E > 0.7; \text{RMSD} < 10 \text{ \AA})$ (see text) of the native state of the monomer as a function of the number of peptides p-S₈ ($T = 2.5$ kJ/mol).

ber n_p of p-LES in the system as $TS^{\text{rot-transl}} - T \log n_p$, the roto-translational entropy of the p-LES depending weakly on n_p at low concentrations, as in the case under discussion. Consequently, regardless of the stabilization energy of the protein monomer, there always exists a number n_p of p-LES such that the free energy of the unfolded state is lower than that of the native state. At the “biological” temperature $T = 2.5$ kJ/mol we observe (see Fig. 9) that $n_p = 1$ is enough to destabilize the native state to a measurable extent. When the temperature is lowered, entropy plays a less important role (i.e., $F = E - TS$), and the equilibrium state becomes the lowest potential energy state; that is, the native state.

The fact that p-S₈ destabilizes the monomeric protease is a sufficient condition to prevent the replication of the virus. This is in keeping with the fact that the monomer is at equilibrium with the dimer. Consequently, the destabilization of the monomer shifts the equilibrium of the system towards the unfolded state. Moreover, the fact that the monomeric state is consistently populated under physiological conditions suggests that this shift would be fast and effective.

Nonetheless, it is interesting to follow the interaction between p-LES and HIV-1-PR, starting from the protein in its dimeric native conformation. For this purpose, we have performed 10^{10} Monte Carlo Step simulations of a system composed of the dimer and three p-S₂ peptides at $T = 2.5$ kJ/mol. Since the two monomers that build out the dimer are identical, the model gives the same interaction energy to a pair of residues X–Y building a native contact in the monomer, and to a pair X–Y', where Y' belongs to the other monomer (consequently, the dimer–dimer interaction departs from a Gō model). Due to its large size, the equilibration of the system is computationally quite demanding (3 wk on a Xeon work station). The population p_N of the native dimeric state (using the same definition as above) is 0.05, compared with 0.30 for the case in which the isolated dimer is evolved starting again from the native conformation. The detailed values for the populations of the various states are given in Table 4. The large errors ascribed to these numbers reflect the computational difficulties found in equilibrating the system.

Table 4. The equilibrium population p_N ($q_E > 0.7$; $RMSD < 10$ Å) of the different states of the HIV-1-PR dimer in the presence of n_p p-S₈ peptides at $T = 2.5$ kJ/mol

n_p	$N_2(\%)$	$2N(\%)$	$U(\%)$	$U_2(\%)$
0	30.4 ± 29.1	68.4 ± 29.0	1.1 ± 0.1	0.1 ± 0.1
3	0.05 ± 0.05	31.3 ± 16.2	68.6 ± 21.1	0.05 ± 0.05

N_2 , Native dimeric state; $2N$, monomers in the native state but not dimerized; U , monomers unfolded and separated; U_2 , monomers unfolded but native interactions at the dimer interface formed.

Figure 10. A snapshot of the simulation of the dimer with three p-LES, starting from the native conformation. The LES of the protein are indicated with a thicker gray tube, while the p-LES are highlighted in black. Also given are the number of the initial and final sites of each of the local elementary structures belonging to the left (L)- and right (R)-drawn monomers.

A snapshot of the result of the three p-S₈ plus dimer simulation is shown in Figure 10. It is seen that the p-S₈ peptides are able to bind to the protein, blocking the way of native S₈ LES to dock S₂ LES, thus disrupting the FN. This situation is similar to that shown in Figure 8 (monomer plus three p-S₈ peptides). In the present case, when p-S₈ enters the protein, the RMSD of the associated monomer is increased to a value ≈ 14 Å, which implies that the native conformation is lost.

Effects of mutations in the HIV-1-PR sequence

In order to assess the effect that mutations of the protease have on the effectiveness of p-S₈ as an inhibitor, we have performed a number of simulations of mutated protease with and without p-S₈. Within the framework of the present model a mutation on a given site is made operative by switching off all the native interactions made by that site in the wild-type sequence, treating them as if they were non-native (see Equation 1). We have applied this procedure to a number of sites that are known to be mutated by the virus to escape drugs (e.g., 19, 37, 63, 67, 72, and 95), and to sites that belong to a LES (31, 33, and 85) or that interact with a LES (68).

In Figure 11 we display the population of the native state p_N for the mutated monomeric protease (continuous curve) and for the system composed of the mutated protease and three p-S₈ (dotted curve). All mutations except those on sites 85 and 33 have little effect on the stability of the protein. On the other hand, the denaturing effect of the p-S₈ is fully retained, as expected from the fact that the interaction between p-S₈ and the monomer is unchanged. In fact, the interaction between p-S₈ and the monomer is the same as those that stabilize the

Figure 11. The effect of mutations on a number of sites of the monomer (X-axis) on the stability p_N of the native state (Y-axis). The solid curve indicates the stability of the monomer alone ($T = 2.5$ kJ/mol), while the dotted curve refers to the case of the monomer plus three p-S₈ peptides. The point drawn at abscissa zero indicates the wild-type sequence.

monomer itself. Consequently, the maintained stability of the mutated protein is proof of the maintained affinity between p-S₈ and the mutated protein.

The mutation on site 33 causes a consistent destabilization of the protease because it is a “hot” site belonging to a LES (S₂), where a large fraction of the stabilization energy of the protein is concentrated. In this case the affinity of p-S₈ to the protease is also diminished, and consequently its destabilizing effect is greatly reduced. In any case the net effect of this mutation is that the protein becomes quite unstable.

Conclusions

We have identified fragments 24–34, 83–93, and 75–78 as the local elementary structures that guide the folding and are responsible for the stability of HIV-1-PR. Peptides with the same sequence as fragment 83–93 have shown good inhibitory properties, causing a consistent unfolding of its native state. It is not likely that HIV-1-PR can develop resistance against these peptides. To do so, the protease should mutate amino acids that play an important role in its folding. The strategy presented in this paper seems to solve the two major problems encountered in drug design: optimization and resistance. Due to the universality of the approach, it is suggested that this kind of folding-inhibitor molecule can be designed and used in connection with other target proteins.

Appendix A: Nature of the HIV-1-PR dimer

At low pH, calorimetry experiments (Todd et al. 1998) have shown that there is a single transition at $T = 59^\circ\text{C}$ (pH 3.4, 25 μM protein, 100 mM NaCl) between the

dimeric native state and a monomeric unfolded state. This means that, for example, at $T = 25^\circ\text{C}$ there is essentially no monomeric protease in solution. At this temperature, the stabilization energy of the dimer is about 10 kcal/mol, but each pair of isolated monomers is unstable (the stabilization free energy being ≈ -13 kcal/mol), the stabilization energy coming from the interface ($\approx +20$ kcal/mol).

The analysis of the distribution of stabilization energy among the residues at the interface indicates that this is concentrated in few “hot” spots, namely 1, 3, 5, and 95–99. This behavior is somewhat odd for a two-state dimer, in which typically the stabilization energy is spread out uniformly on the interface (Tiana and Broglia 2002). Note, however, that most likely the evolution of the HIV protease has mostly taken place in the cytoplasm, which is neutral, and consequently one should compare its evolutionary properties with its stability features at higher pH.

When increasing the pH value, acidic residues acquire a negative charge. In particular, the pair of D25 that lie close on the interface repel each other through Coulomb force. The overall effect is to increase the dissociation constant (measured by sedimentation equilibrium experiments), which assumes the value $k_D = 5.8$ μM at pH 7 (and $T = 4^\circ\text{C}$; Xie et al. 1999), further increasing at higher temperatures. The dissociation of the dimer is accompanied by a decrease in the internal structure of the monomers, as testified by the fact that CD experiments highlight a decrease of $\sim 26\%$ of the β -structure of the protein (Xie et al. 1999). Consequently, one expects a detectable ratio of folded monomers in solution. For instance, according to a scenario in which a loss of 26% structure means that 26% of monomers have no structure at all, the ratio of folded monomers is 18 %. (Of course, this is an idealized situation. In a more realistic scenario, a larger ratio of monomers is partially destabilized; e.g., 52% of monomers lose half of the structure).

One should note that sedimentation equilibrium results are apparently contradicted by fluorescence assays. The intensity of fluorescence at equilibrium at different pH values displays maximum values in the region between pH 4.5 and pH 8, decreasing both at acidic and basic values of pH (Grant et al. 1992; Szeltner and Polgar 1996). Moreover, equilibrium experiments in which the fluorescence is recorded upon addition of urea at constant pH and the concentration of urea corresponding to the midpoint in the change in fluorescence (Todd et al. 1998) show that the midpoint concentration of urea increases if pH is increased from 4 to 5.5. These results seem to contradict the results discussed above, suggesting that the protease is more stable at neutral pH. They also seem to contradict kinetic fluorescence measurements, in which the kinetics of fluorescence emission upon addition of urea is recorded at different pH. The

Table 5. The values of dissociation constant k_D of the dimer, calculated at different conditions

pH	T(C)	NaCl	k_D	Method	Reference
3.4	25	100 mM	17 nM	fluorescence	Todd et al. 1998
4.5	4	0	<100 nM	sedim. eq.	Xie et al. 1999
4.5	9	0	28 μ M	sedim. eq. (HIV-2)	Holzman et al. 1991
5	4	0.2 M	<10 nM	sedim. eq.	Grant et al. 1992
5	25	100 mM	23 pM	fluorescence	Todd et al. 1998
5	37	1 M	3.6 nM	activity	Zhang 1997
7	4	0	5.8 μ M	sedim. eq.	Xie et al. 1999
7	25	1 M	50 nM	activity	Cheng et al. 1990
7.5	9	0	87 μ M	sedim. eq. (HIV-2)	Holzman et al. 1991

“unfolding rates” thus obtained behave in a specular way as compared to the equilibrium measurement, displaying low rates (i.e., “more stable”) between pH 4 and 7, and increasing (i.e., “less stable”) at the extremes (Szeltner and Polgar 1996).

A summary of the dissociation constants obtained with different techniques and under different conditions is presented in Table 5. In considering these results, one should remember that each monomer of the protease has two Trp residues, at position six at the interface and at position 42 on the surface, at the rear part of the flap. Consequently, not only is fluorescence intensity not able to distinguish between unfolding of the monomer and dissociation of the dimer, but more subtle processes such as opening and closing of the flaps can affect the recorded signal. As a consequence, we consider more reliable the results obtained in sedimentation equilibrium experiments, which is the only direct inspection of the monomer/dimer character of the solution.

Appendix B: Dynamic simulations

Metropolis MC simulations are meant to sample the conformational space of a system in order to investigate thermodynamic quantities at equilibrium. Nonetheless, it has been shown that, if the elementary movement of the chain is small enough, the algorithm describes in an approximated way the dynamics of the system (Rey and Skolnick 1991), solving effectively the Langevin equations associated to the system (Kikuchi et al. 1991).

In order to transform MC steps into seconds, we have calculated the diffusion coefficient of the center of mass of the protein monomer and compared it with the one obtained by Stokes' approximation, describing diffusion of a spherical object of radius 30 Å in water at room temperature. The resulting relation is one Monte Carlo step equal to 10^{-13} sec.

References

Abkevich, V.I., Gutin, A.M., and Shakhnovich, E.I. 1994. Specific nucleus as the transition state for protein folding. *Biochemistry* **33**: 10026–10032.

- Bahar, I., Atilgan, A.R., Demirel, M.C., and 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.
- Baldwin, R.L. and Rose, G.D. 1990. Is protein folding hierarchic? *TIBS* **24**: 26–83.
- Berendsen, H.J.C., van der Spoel, D., and van Drunen, R. 1995. GRO-MACS: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.* **91**: 43–56.
- Borg, J. 2001. Optimized Monte Carlo analysis for generalized ensembles. *Europ. Phys. J. B* **29**: 481–484.
- Broglia, R.A. and Tiana, G. 2001a. Hierarchy of events in the folding of model proteins. *J. Chem. Phys.* **114**: 7267–7273.
- . 2001b. Reading the three-dimensional structure of a protein from its amino acid sequence. *Proteins* **45**: 421–427.
- Broglia, R.A., Tiana, G., Pasquali, S., Roman, H.E., and Vigezzi, E. 1998. Folding and aggregation of designed protein chains. *Proc. Natl. Acad. Sci.* **95**: 12930–12934.
- Broglia, R.A., Tiana, G., and Berera, R. 2003. Resistance proof, folding-inhibitor drugs. *J. Chem. Phys.* **118**: 4754–4758.
- Cecconi, F., Micheletti, C., Carloni, P., and Maritan, A. 2001. Molecular dynamics studies on HIV-1 Protease drug resistance and folding pathways. *Proteins* **43**: 365–372.
- Cheng, Y.S., Yin, F.H., Foundling, S., Blomstrom, K., and Kettner, C.A. 1990. Stability and activity of human virus protease. *Proc. Natl. Acad. Sci.* **87**: 9660–9664.
- Gō, N. 1983. Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioengin.* **12**: 183–210.
- Grant, S.K., Deckman, I.C., Culp, J.S., Minnich, M.D., Brooks, I.S., Hensley, P., Debouck, C., and Meek, T.D. 1992. Use of protein unfolding studies to determine the conformational and dimer stability of HIV-1 and SIV proteases. *Biochemistry* **31**: 9491–9501.
- Holm, L. and Sander, C. 1996. Mapping the protein universe. *Science* **273**: 595–602.
- Holzman, T.F., Kohlbrenner, W.E., Weigl, D., Rittenhouse, J., Kempf, D., and Erickson, J. 1991. Inhibitor stabilization of human immunodeficiency virus type-2 proteinase dimer formation. *J. Biol. Chem.* **266**: 19217–19220.
- Ishima, R., Ghirlando, R., Todtzer, J., Gronenborn, A.M., Torchia, D.A., and Louis, J.M. 2001. Folded monomer of HIV-1 protease. *J. Biol. Chem.* **276**: 49110–49116.
- Kaya, H. and Chan, H.S. 2000. Polymer principles of protein calorimetric two-state cooperativity. *Proteins* **40**: 637–661.
- Kikuchi, K., Yoshida, M., Maekawa, T., and Watanabe, H. 1991. Metropolis Monte Carlo method as a numerical technique to solve Fokker-Planck equation. *Chem. Phys. Lett.* **185**: 335–338.
- Lesk, A.M. and Rose, G.D. 1981. Folding units in globular proteins. *Proc. Natl. Acad. Sci.* **78**: 4304–4308.
- Levy, Y. and Caflisch, A. 2003. The flexibility of monomeric and dimeric HIV-1 PR. *J. Phys. Chem. B* **107**: 3068–3079.
- Levy, Y., Caflisch, A., Onuchic, J.N., and Wolynes, P.G. 2004a. The folding and dimerization of HIV-1 Protease: Evidence for a stable monomer from simulations. *J. Mol. Biol.* **340**: 67–79.
- Levy, Y., Wolynes, P.G., and Onuchic, J.N. 2004b. Protein topology determines binding mechanism. *Proc. Natl. Acad. Sci.* **101**: 511–516.
- Li, M.S., Klimov, D.K., and Thirumalai, D. 2004. Finite size effects on calorimetric cooperativity of two-state proteins. *Physica A* **350**: 38–44.
- Mirny, L. and Shakhnovich, E.I. 1999. Universally conserved positions in protein folds: Reading evolutionary signals about stability, folding kinetics and function. *J. Mol. Biol.* **291**: 177–184.

- Panchenko, A.R., Luthey-Schulten, Z., and Wolynes, P.G. 1995. Foldons, protein structural molecules, and exons. *Proc. Natl. Acad. Sci.* **93**: 2008–2013.
- Pearson, W.A. 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.* **183**: 63–98.
- Ptitsyn, O.B. and Rashin, A.A. 1975. A model of myoglobin self-organization. *Biophys. Chem.* **3**: 1–20.
- Rey, A. and Skolnick, J. 1991. Comparison of lattice Monte Carlo dynamics and Brownian dynamics folding pathways of α -helical hairpins. *Chem. Phys.* **158**: 199–205.
- Sander, C. and Schneider, R. 1991. Database of homology derived protein structures and the structural meaning of sequence alignment. *Proteins* **9**: 56–68.
- Shafer, R.W., Hsu, P., Patick, A.K., Craig, C., and Brendel, V. 1999. Identification of biased amino acid substitutions patterns in HIV type 1 isolated from patient treated with protease inhibitors. *J. Virol.* **73**: 6197–6202.
- Szeltner, Z. and Polgar, L. 1996. Conformational stability and catalytic activity of HIV-1 protease are both enhanced at high salt concentration. *J. Biol. Chem.* **271**: 5458–5463.
- Tiana, G. and Broglia, R.A. 2001. Statistical analysis of native contact formation in the folding of designed model proteins. *J. Chem. Phys.* **114**: 2503–2507.
- . 2002. Folding and design of dimeric proteins. *Proteins* **49**: 82–92.
- Tiana, G., Broglia, R.A., Roman, H.E., Vigezzi, E., and Shakhnovich, E.I. 1998. Folding and misfolding of designed protein-like chains with mutations. *J. Chem. Phys.* **108**: 757–761.
- Tiana, G., Broglia, R.A., and Shakhnovich, E.I. 2000. Hiking in the energy landscape in sequence space: A bumpy road to good folders. *Proteins* **39**: 244–251.
- Tiana, G., Simona, F., De Mori, G.M.S., Broglia, R.A., and Colombo, G. 2004. Understanding the determinants of stability and folding of small globular proteins from their energetics. *Protein Sci.* **13**: 113–117.
- Todd, M.J., Semo, N., and Freire, E. 1998. The structural stability of the HIV-1 protease. *J. Mol. Biol.* **283**: 475–488.
- Tomasselli, A.G. and Heinrikson, R.L. 2000. Targeting the HIV-Protease in AIDS therapy: A current clinical perspective. *Biochem. Biophys. Acta.* **1477**: 189–214.
- Wallqvist, A., Smythers, G.W., and Covell, D.G. 1998. A cooperative folding unit in HIV-1 protease. Implications for protein stability and occurrence of drug-induced mutations. *Prot. Engin.* **11**: 999–1005.
- Xie, D., Gulnik, S., Gustchina, E., Yu, B., Shao, W., Qoroneh, W., Nathan, A., and Erickson, J.W. 1999. Drug resistance mutations can affect dimer stability of HIV-1 protease at neutral pH. *Protein Sci.* **8**: 1702–1713.
- Zhang, Z.Y. 1997. Targeting the dimerization interface of HIV-1 protease: Inhibition with crosslinked interfacial peptides. *J. Am. Chem. Soc.* **119**: 4841–4845.