Computational Study of the Resistance Shown by the Subtype B/HIV-1 Protease to Currently Known Inhibitors

Alessandro Genoni, Giulia Morra, Kenneth Merz, Giorgio Colombo

Computational study of the resistance shown by the Subtype B / HIV-1 Protease to currently known inhibitors †

Alessandro Genoni [a,b], Giulia Morra [a], Kenneth M. Merz Jr. [b] *, Giorgio Colombo [a] *

(a) Istituto di Chimica del Riconoscimento Molecolare, CNR, Via Mario Bianco 9, 20131, Milano, Italy.

(b) Quantum Theory Project, University of Florida, P.O. Box 118435, Gainesville, Florida, 32611, USA.

† A.G. and G.C. thank A.I.R.C. for funding. This work was partially supported by a grant (FP6 STREP "BacAbs", grant number LSHB-CT-2006-037325) from the European Community. K.M.M. thanks NIH (GM044974) for financial support of this research.

* Kenneth M. Merz Jr.:

Phone: +1-352-392-6973;

Fax: +1-352-392-8722;

E-mail: merz@qtp.ufl.edu

Giorgio Colombo:

Phone: +39-02-28500031;

Fax: +39-02-28901239;

E-mail: giorgio.colombo@icrm.cnr.it

Abbreviations and Textual Footnotes

HIV: Human Immunodeficiency Virus; HIV-1 PR: Protease of the Type 1 Human Immunodeficiency Virus; HAART: Highly Active Anti-Retroviral Therapy; NRTI: Nucleoside Reverse Transcriptase Inhibitors; NNRTI: Non-Nucleoside Reverse Transcriptase Inhibitors; HIV-1 RT: Reverse Transcriptase of the Type 1 Human Immunodeficiency Virus; WT: Wild Type; MD: Molecular Dynamics; ILCP: Intrinsic Long-Range Communication Propensity; RWSIP: Root Weighted Square Inner Product; EDM: Energy Decomposition Method; MLCE: Matrix of Local Coupling Energies; PCA: Principal Component Analysis; ED: Essential Dynamics.

Abstract

Human Immunodeficiency Virus type 1 Protease (HIV-1 PR) is an essential enzyme in the HIV-1 life cycle. As such, this protein represents a major drug target in AIDS therapy, but emerging resistance to anti-retroviral inhibitor cocktails, due to high viral mutation rates, represents a significant challenge in AIDS treatment. Many mutations are not located within the active site or binding pocket, nor they do significantly modify the 3D structural organization of the enzyme; hence, the mechanism(s) by which they alter inhibitor affinity for the Protease remains uncertain. In this article, we present an all-atom computational analysis of the dynamic residue-residue coordination between the active site residues and the rest of the protein and of the energetic properties of different HIV-1 PR complexes. We analyze both the wild type form and mutated forms that induce drug resistance. In particular, the results show differences between the wild type and the mutants in their mechanism of dynamic coordination, in the signal propagation between the active site residues and the rest of the protein and in the energy-networks responsible for the stabilization of the bound inhibitor conformation. Finally, we propose a dynamic and energetic explanation for HIV-1 Protease drug resistance and, through this model, we identify a possible new site that could be helpful in the design of a new family of HIV-1 PR allosteric inhibitors.

It is well known that a serious problem in overcoming human immunodeficiency virus (HIV) infections is connected to its genetic variability (*1*). HIV exists as a type 1 (HIV-1) or a type 2 (HIV-2) strain, with the former being the most virulent and widespread in the worldwide pandemic. HIV-1 is further subdivided into three groups M, N and O. Furthermore, viruses belonging to group M are classified into subtypes, sub-subtypes and recombinant forms, with each one of them being prevalent in specific geographical regions. For instance, subtype C is dominant in sub-Saharan Africa, while subtype B is the most common subtype in the western world.

Until relatively recently, subtype B has been the target of most drug design efforts, which has led to effective therapies, but it has failed to deal with different virulent subtypes. Nonetheless, the highly active anti-retroviral therapy (HAART) (*2-4*), which consists in a combination of three or more drugs from two different drug classes (nucleoside reverse transcriptase inhibitors (NRTIs) plus either a Protease inhibitor or a non-nucleoside reverse transcriptase inhibitor (NNRTI)), has significantly improved the prognosis of HIV-infected individuals. One of the main targets of HAART is the HIV-1 Protease (HIV-1 PR), which has been considered an optimal target since the early days of HIV research (*5, 6*). This protein plays a fundamental role in virus maturation cleaving the Gag and Gag-Pol poly-proteins into structural and enzymatic proteins. Therefore, its inhibition yields immature virus particles that are unable to spread HIV infection (*7-9*).

HIV-1 PR, whose crystallographic structures are currently available both in the *apo* form (the unbound form) and complexed with inhibitors, is an aspartic Protease existing as a homo-dimer with 99 aminoacids in each subunit. The active site is defined by the triad of Asp25-Thr26-Gly27 (one for each monomer) and enveloped by two β-hairpin flaps (*10,*

11). These flaps are semi-closed when the enzyme is unbound by substrate, while they are completely closed in the presence of substrate or inhibitors. Interestingly, several research groups have carried out molecular dynamics (MD) simulations to investigate the opening and closure of the flaps upon ligand binding (*12-15*). In this context, Galiano *et al.* (*16*) have recently shown by means of electron paramagnetic resonance (EPR) experiments and MD simulations that drug induced mutations alter flap conformations and consequently their motion.

Many efficacious drugs for HIV-1 Protease inhibition have been developed (*7, 8, 11, 17-22*). Nevertheless, due to the error prone activity of HIV-1 Reverse Transcriptase (HIV-1 RT), mutations in HIV-1 PR arise during the course of treatment, inducing resistance to the current generation of inhibitors (*11, 23-27*). The onset of mutation induced resistance poses a major challenge in the struggle against HIV, which has stimulated efforts to discover novel drugs that are able to overcome HIV drug resistance (*28-30*).

Interestingly, it has been observed that most of the mutations are located far from the active site pocket (see Figures 1A and 1B) and, hence, their effect on inhibitor binding cannot be easily rationalized. Clearly, understanding the effect of active site and non active site mutations is extremely important for the development of new drugs that are able to target resistant Proteases and several plausible explanations have been recently proposed (*31-40*).

In particular, several studies have shown that different mutations far from the active site may alter the flexibility of HIV-1 PR, thereby inducing structural adaptations that affect the efficacy of the diversity of small molecules currently used in therapy. Theoretical studies, alone or in combination with experimental approaches, have pointed to a general

increase in the flexibility of the mutated enzymes (either in the flap regions or in the vicinity of the active site) as the possible origin of the weaker affinity for inhibitors and their consequent lower efficacy (*13-16, 41-44*). Schiffer and coworkers (*41*), for instance, showed that drug-resistant mutations may alter the packing of the hydrophobic core affecting the conformational flexibility of HIV-1 PR with an impact on the balance between substrate processing and inhibitor binding. Perryman *et al.* (*42, 43*) have shown that the effect of the mutations could be related to a perturbation of the equilibrium between the semiopen and closed conformations of HIV-1 PR, influencing molecular recognition and thus affecting drug affinity. Differences in the overall dynamic properties of the complexes of the WT and mutated HIV-1 PRs with the substrate and with a gem-diol model intermediate have been related to variations in the enzymatic activity by Piana *et al.* (*44*). The authors have proposed the "flexibility-assisted" mechanism as a common property in the majority of compensatory mutations, which do not change the electrostatic properties of the enzyme. Moreover, MD simulations have been used to characterize transitions between different conformations (*13-16*). Based on these computational models, the authors have suggested new ideas for the development of inhibitors taking conformational flexibility into account.

Despite the important insights obtained by several investigators, there are still open questions regarding the complex interplay between the impact of mutations on the dynamic/energetic properties of HIV-1 PR and ligand binding. Specifically, (minor) differences and/or perturbations at the side-chain level or differences due to mutations of a cluster of residues that have an impact on function may not impact major conformational changes or differential fluctuations at the single residue level.

Consequently, it is important to investigate the molecular mechanisms that determine how the perturbations caused by specific mutations are coupled to the dynamics of the active site and how they can propagate to the active site, affecting ligand recognition, binding and processing.

In this paper, we have carried out 50 ns MD simulations both for the LAI wild type of the Subtype B / HIV-1 PR and for two related multidrug resistant mutants (namely, the pseudo-V6 and the MDR 769 mutants (see the "Materials and Methods" section for more details) in the unbound state, complexed with several inhibitors and with a pseudo-substrate peptide. Thus, we have set out to perform a comprehensive computational analysis of the different HIV-1 PR sequences to shed light on possible correlations between their dynamics, long-range coordination effects that connect mutation loci and binding sites and the networks of the most-relevant stabilizing interactions in relation to the presence or absence of inhibitors or substrate. In particular, we aimed to investigate the processes by which signals originating at certain sites of HIV-1 PR propagate to affect the inhibitor-recognition properties of the remote active site.

It has been shown that protein dynamics and plasticity play a central role in selecting alternative residue-residue interactions and coupled motions that are necessary to perform or modify a certain activity. Analysis of the coupling between distant residues may give hints regarding how a certain perturbation is transmitted (communicated) to a distal site from the interaction site. Indeed, several studies on different systems have identified potential dynamic and energetic coupling pathways that connect different functional sites through the structure of a given protein (*45-51*).

In support of the initial hypothesis, our results provide atomic models for differential mutation-induced modulation not only of the global dynamic properties but also of the internal coordination and of the relevant energetic interactions between different sites of HIV-1 PR. In particular, specific intra-protein dynamic and energetic couplings are likely to be selected and (de-)activated by the presence of mutations distant from the active site, which modify the cross-talk between important protein regions allowing the enzyme to escape inhibition while retaining its activity. Moreover, based on the analysis of energetic couplings, we provide a molecular explanation for the observed lack of mutations inducing resistance in a region of HIV-1 PR localized on residues 11-20 (for both the monomers). We propose that this substructure can be used as a target to develop new inhibitors of HIV-1 PR that allosterically affect its function by perturbing protein dynamics, without being subject to the adverse effects of resistant mutations. Therefore, the results of our study provide further insights into our understanding of the mechanisms of drug resistance development through mutations located far from the active site and it may be of use in the design of new small molecule drugs with improved therapeutic perspectives.

In summary, in this article we provide both a dynamic and an energetic understanding of HIV-1 PR drug resistance and we merge these two facets into an integrated model that, together with those already proposed, will aid in the design of new inhibitors that are able to adapt to mutations onset or that are at least able to effectively bind to resistant Proteases.

Materials and Methods

Systems and structures. In our study we have considered the following genotypes of the Subtype B / HIV-1 Protease: the LAI wild type (from now on BLAI) and the multidrug resistant mutants V6 (with the addition of the I54V and I84V mutations; BV6) (*38*) and MDR 769 (BMDR) (*52*). In particular, BV6 is characterized by 10 mutations per monomer (K20R, V32I, L33F, M36I, I54V, L63P, A71V, V82A, I84V, L90M; see Figure 1A), whereas BMDR by 11 mutations per monomer (L10I, M36V, S37N, M46L, I54V, I62V, L63P, A71V, V82A, I84V, L90M; see Figure 1B). Each genotype has been taken into account both in the *apo* form (i.e., unbound) and with some of the currently known inhibitors (Indinavir (IND) (*18*), Nelfinavir (NFV) (*19*), Ritronavir (RIT) (*20*) and TL-3 (TL3) (*21, 22*)) or with a polypeptide (Arg-Val-Leu-Ala-Glu-Ala-Met) that mimics the real substrate (SUB). If available, we used proper PDB files as starting structures for our MD simulations, otherwise we used Glide (*53, 54*) and modeled the complex of the inhibitor or of the polypeptide with the best-resolved Protease structure (see Table 1 for details). Where necessary, hydrogen atoms were added using the Leap module in the AMBER 9.0 package (*55*) and, assuming that the pH for our simulations is about 4.7, the two histidine residues of the Protease were protonated. Missing parameters for the inhibitors were obtained by means of ANTECHAMBER (*56*) in conjunction with the "General AMBER Force Field" (GAFF) (*57*). Finally, it is worth noting that all of the original PDB structures of the two mutants contain the artificial mutation D25N, which blocks protein self-cleavage. Therefore, before starting each simulation, we performed the back mutation N25D using the Swiss-PDB Viewer 4.0 software (*58*).

MD Simulations. All the simulations were performed using the AMBER 9.0 package with the ff99SB force field (*59*), the TIP3P water model (*60*) in a truncated octahedron box with an 8.0 Å buffer (namely, 8.0 Å + protein + 8.0 Å in each direction) to model the explicit solvent, a 10 Å non-bonded cutoff distance and the Particle Mesh Ewald summation method (PME) (*61-63*) to deal with long-range Coulomb interactions. The simulation protocol consisted of the following four steps: 1) 1000-step minimization (steepest descent for the first 500 steps) with a 500.0 $kcal \cdot mol^{-1} \cdot Å^{-2}$ harmonic restraint applied to all solute atoms; 2) 1500-step minimization (steepest descent for the first 750 steps) without constraints; 3) 20 ps MD simulation (2 fs integration step with the SHAKE algorithm (*64*)) at constant volume, heating the system from 0 K to 300 K (Langevin Temperature Equilibration Scheme (*65-67*)) and with a 10.0 $kcal \cdot mol^{-1} \cdot Å^{-2}$ harmonic restraint applied only to the solute atoms; 4) 50 ns MD simulation at constant pressure (1 atm) and at 300 K (Langevin) using a 2 fs integration step with the SHAKE algorithm. It is important to point out that only the production phase of the fourth step (namely, the last 40 ns of MD simulation) was used in the analyses that will be described in the following two sections.

Communications analysis. Bahar and coworkers (*68, 69*) have recently devised a technique for the analysis of signal propagation within proteins. This strategy, which was initially formulated within the framework of the elastic network model, was recently extended to all-atom MD simulation by Morra *et al.* (*70, 71*). This approach associates signal transduction events in proteins with the fluctuation dynamics of atoms. In fact, for

each couple of residues, it is possible to define a communication propensity (λ_{ij}) defined as the variance of the distance between the two residues:

$$\lambda_{ij} = \left\langle \left(d_{ij} - \left\langle d_{ij} \right\rangle \right)^2 \right\rangle \quad (1)$$

where d_{ij} is actually the distance between the C_α atoms of residue *i* and residue *j*.

The communication propensity can be considered as a communication time and, therefore, low λ_{ij} values are often associated with efficiently communicating residues. In other words, two residues characterized by a C_α-C_α distance that fluctuates with a small intensity during the MD trajectory are hypothesized to communicate efficiently and this means that a perturbation at one of the two residues will quickly propagate to the other aminoacid. Conversely, when the distance fluctuations associated with two residues are large, the communication is less efficient and any perturbations at one site will slowly become visible at the other one. Moreover, it is important to note that, to determine whether a communication between two residues is efficient or not, it is possible to introduce a threshold η, one for each simulation, that plays a fundamental role in all of our analyses. It is computed as an average λ_{ij} value that takes into account nearby aminoacids along the sequence. In particular, η is usually calculated by considering for each residue *i* its communication propensities with neighboring residues between *i-4* and *i+4*:

$$\eta = \frac{1}{m} \sum_{i=1}^{N_{res}} \sum_{j=i-4}^{i+4} \lambda_{ij} \quad (\text{with } j \neq i \text{ et } 1 \leq j \leq N_{res}) \quad (2)$$

with *m* as the total number of terms taken into account in the sum.

After the calculation of the λ values between all pairs of residues, it is interesting to determine the intrinsic long-range communication propensity (ILCP) for each aminoacid. In particular, fixing a reference distance δ, the ILCP value for a generic residue *i* is defined as the fraction of residues that efficiently communicate with it ($\lambda_{ij} \leq \eta$) at distances larger than δ. This new quantity greatly simplifies the quite complicated communication networks that are usually obtained from the communication propensities calculations and, using the new simplified picture, it is much easier to observe changes in the communication capabilities of the residues upon mutation or ligand binding.

RWSIP Definition. In order to measure the agreement between two dynamical spaces, we have exploited the Root Weighted Square Inner Product (RWSIP), a quantity introduced by Carnevale *et al. (72*) and defined as

$$RWSIP = \sqrt{\frac{\sum_{l,m} \frac{1}{\lambda_l \mu_m} \left| \mathbf{v}_l \cdot \mathbf{u}_m \right|^2}{\sum_l \frac{1}{\lambda_l \mu_l}}} \qquad (3)$$

wehre $\mathbf{v}_l$ and $\mathbf{u}_m$ are the *l*-th and the *m*-th eigenvectors (with the corresponding eigenvalues λ_l^{-1} and μ_m^{-1}) of the covariance matrices associated with the first and the second MD simulation in the comparison, respectively.

Energy Decomposition Method. The complicated network of energetic interactions between aminoacids represents one of the main drawbacks in the identification of crucial residues for the protein fold and energetic stability. In order to overcome this problem, Colombo and coworkers have recently proposed the energy decomposition method

(EDM) (*73-76*) that, as first step, computes the matrix of non-bonded interaction energies (namely, van der Waals and electrostatic interactions) between pairs of residues. This matrix is afterwards diagonalized and, from the analysis of the eigenvector associated with the lowest eigenvalue, it is possible to determine those residues that behave as strongly interacting and stabilizing centers. It is worthwhile to observe that there are two slightly different versions of the energy decomposition method. In the first one the non-bonded interaction energy matrix is obtained as an average over the MD trajectory. Moreover, in this case, solvation effects are not taken into account directly, although the solvent is present during the MD simulation and, hence, it influences the protein structure. In the second approach, after a cluster analysis is performed on the MD trajectory, only the most representative structure of the most populated cluster is taken into account and the non-bonded interaction energy matrix is computed on that protein structure. Obviously, in the second case, the average solvent effect is not considered and, to resolve this drawback, the solvent is directly taken into account using the PBSA method (*77, 78*) in the calculation of the non-bonded interactions. It is important to observe that the two versions of the EDM provide qualitatively equivalent results, although the second one is less computationally demanding. Of course, due to the great number of simulations to be analyzed in our study, we opted for the second approach, using the GROMOS method (*79*) with 0.1 nm as RMSD cutoff for the cluster analysis.

Now, for the sake of completeness, it is interesting to consider some theoretical details about the EDM. Let us indicate with **M** the non-bonded interaction energy matrix without the diagonal elements, namely without the self-interaction terms. This matrix can be diagonalized and expressed in terms of its eigenvalues and eigenvectors:

$$\mathbf{M}_{ij} = \sum_{k=1}^{N} \lambda_k \, \mathbf{w}_{ik} \, \mathbf{w}_{jk} \tag{4}$$

where:

- N is the number of amino acids in the protein;
- λ_k is the k-th eigenvalue;
- $\mathbf{w}_{ik}$ is the i-th component of the k-th eigenvector

Eigenvalues and eigenvectors are usually labeled following an increasing order. Therefore, λ_1 is the lowest eigenvalue and, from now on, we will refer to the first eigenvector as the eigenvector corresponding to the eigenvalue λ_1.

The total non-bonded energy is defined as:

$$E_{nb} = \frac{1}{2} \sum_{i,j=1}^{N} \mathbf{M}_{ij} = \frac{1}{2} \sum_{i,j=1}^{N} \sum_{k=1}^{N} \lambda_k \, \mathbf{w}_{ik} \, \mathbf{w}_{jk} = \frac{1}{2} \sum_{k=1}^{N} \lambda_k W_k \tag{5}$$

with $W_k = \sum_{i,j=1}^{N} \mathbf{w}_{ik} \, \mathbf{w}_{jk}$. If $\left| \lambda_1 W_1 \right|$ is much larger than $\left| \lambda_k W_k \right|$ for $k \neq 1$, the sum over i and j of $\mathbf{M}_{ij}$ is dominated by the contribution due to the first eigenvalue and eigenvector, such that the total non-boned energy can be approximated by:

$$E_{nb} \approx E_{nb}^{app} = \frac{\lambda_1}{2} \sum_{i,j=1}^{N} \mathbf{w}_{i1} \, \mathbf{w}_{j1} = \frac{\lambda_1 W_1}{2} \tag{6}$$

As mentioned above, the eigenvector associated with the lowest eigenvalue is used to identify the most stabilizing aminoacids. In particular, considering its squared components as the weights of the corresponding residues in the structural stabilization, we can define "hot spots" those residues with a weight higher than a threshold t. This threshold is chosen equal to the squared component of a normalized "flat eigenvector" (namely, a normalized vector whose components provide the same contribution for each

site). This corresponds to a case in which each residue equally contributes to the structural stability and, therefore, the threshold *t* is equal to *1/N*, where *N* is the number of the eigenvector components.

Energy-based detection of putative interaction sites on HIV-1 PR. The search for alternative putative binding regions on HIV-1 PR has been carried out by applying a recently proposed approach for the identification of protein interaction sites (*80*). The method, originally developed to identify antibody-binding sites (epitopes) on protein antigens, is based on the rationale that interaction sites generally display low-intensity intra-protein energy interaction networks. In other words, these sites are characterized by non-optimized intramolecular interaction networks so that they may undergo conformational changes and be recognized by a binding partner with minimal energetic expense. Therefore, low intensity intra-protein energetic couplings would allow these sites to be immediately available for possible interactions with a second partner. Moreover, these interaction sites usually correspond to localized sub-structures of the protein that are exposed or easily accessible on the protein surface.

In the context of HIV-1 PR, sites with minimal energetic couplings with the rest of the protein may be interesting for two reasons. First, they may suggest new sites that can be targeted by inhibitors, either through *de novo* design or through virtual screening efforts where the identified regions are used as docking sites for small molecules. Second, and most importantly, these putative interaction sites would not respond to the effects of drug resistant mutations. In fact, a minimal energetic coupling to the other residues of the

protein would minimize the effects of variations in other (distal) regions on the newly discovered interaction sites, limiting the potential of drug resistant mutations.

Based on these considerations, the identification of interaction sites from the structure of the isolated protein is achieved by combining the analysis of the protein energetics obtained from an MD simulation with the topological information deriving from the contact matrix of the most representative structure of the trajectory. This allows for the detection of spatially close regions in the structure that are minimally coupled to the rest of the protein and that represent possible small molecule binding sites that can be generated at a minimal energetic expense.

The analysis of the protein energetics is based on the Energy Decomposition Method described above and, exploiting the approximation underlying equation (6), the interaction energy matrix **M** can be defined as

$$\mathbf{M}_{ij} \approx \tilde{\mathbf{M}}_{ij} = \lambda_1 \mathbf{w}_{i1} \mathbf{w}_{j1} \tag{7}$$

where the symbols have the same meaning of the corresponding ones in the previous subsection.

As briefly mentioned above, the contact matrix **C** is obtained from the most representative structure of the most populated cluster of the MD trajectory and its elements $\mathbf{C}_{ij}$ are allowed to assume values:

$$\mathbf{C}_{ij} = \begin{cases} 1 & \text{if } r_{ij} < 0.65 \text{ nm} \\ 0 & \text{if } r_{ij} > 0.65 \text{ nm} \end{cases} \tag{8}$$

with r_{ij} as the distance between the C_β atoms of residues *i* and *j*. For the sake of homogeneity with the energy matrix, the contacts between the nearest neighbors *i* and *i+1* are also included.

Performing the Hadamard product (or pointwise product) between the simplified interaction matrix $\tilde{\mathbf{M}}$ and the contact matrix **C**, we obtain the Matrix of Local Coupling Energies (MLCE):

$$\Gamma = \tilde{\mathbf{M}} \bullet \mathbf{C} \tag{9}$$

that allows to determine in a compact way which residue pairs within the contact cutoff are coupled through energetic interactions. The non-zero elements of $\boldsymbol{\Gamma}$ are afterwards ranked in increasing order (namely, from the weakest local interaction to the strongest local interaction) and only the lowest 15% of all the contact-filtered pairs is taken into account. This subset represents the group of local interactions with minimal intensities and the residues belonging to it define possible epitope sequences that identify antigen-antibody or protein interaction sites, namely *soft spots* that can favorably bind a binding partner and that do not respond to variations (e.g., mutations, changes in the binding state) in distal regions of the protein.

Results and Discussion

The modulation of protein-substrate, protein-ligand and protein-protein interactions by perturbation of sites far from the interaction site, either through allosteric ligand binding or point mutation, has been recognized as a general property of many monomeric proteins (*81*). In this paper, we have carried out all-atom MD simulations of the LAI Wild Type Subtype B / HIV-1 Protease and of two related multidrug resistant mutants (pseudo-V6 and MDR-769), both in the unbound state and complexed with inhibitors or with a pseudo-substrate. Our fundamental goal was to shed light on the molecular mechanisms by which perturbations engendered by point mutations at certain sites of HIV-1 PR

propagate to influence the recognition properties of the remote active site, affecting the affinity for inhibitors, while leaving substrate binding and processing unchanged. We have focused on analyzing the correlations and differences in the internal coordination (communication), in the dynamics and in the energetics of the enzyme related to the presence of mutations inducing resistance and to their influence on the selection of enzyme dynamic states with different affinities for inhibitors.

To address these issues, we have carried out mainly two types of analysis of the MD trajectories: 1) the analysis of dynamic communications (residue-residue coordination) and of the signal transduction established between protein residues (*68-71*) and 2) the analysis of the residue-residue energetic couplings. As already described in the previous section, the former method is based on an approach recently proposed by Bahar and coworkers for elastic network models (*68, 69*) that we have extended to the analysis of all-atom MD simulation trajectories (*70, 71*). The analysis of residue-residue communications describes signal transduction events in proteins as directly related to the fluctuation dynamics of atoms, defining the communication propensity of a residue-pair as a function of the fluctuations of the inter-residue distance components. The latter strategy, based on the Energy Decomposition Method (*73-76*), aims to quantitatively map the networks of aminoacid interactions that are the most relevant to the structural and dynamic organization of a protein by connecting distant sites in the tertiary structure. It has been previously shown (*73-76*) that the residues defining the energetic-coupling networks show good correlations with several experimental sets of stability and mechanistic data.

Communication Analysis. For each MD simulation performed, we have computed all the communication propensities (λ_{ij}) between pairs of residues and we have determined the corresponding "efficient communication threshold" η (see Table 2; for the sake of comparison we have also reported, in the Supporting Information, the average, minimum and maximum values of the communication propensities for each simulation). Using 10 Å, 15 Å and 20 Å as reference distances, the λ_{ij} values have afterwards been used to calculate the Intrinsic Long-range Communication Propensities (ILCPs) for each residue in all the simulations and the results obtained have been plotted in Figures 2 and 3. In these Figures, we have depicted the 10 Å (reference distance) ILCP values for the LAI wild type Protease (BLAI) and for the pseudo-V6 mutant (BV6), respectively. The comparison of corresponding data for analogous complexes (for instance, BLAI_APO *vs*. BV6_APO or BLAI_IND *vs*. BV6_IND) shows that, in the presence of any inhibitor, the active site residues of the mutant (namely, residues 25, 26, 27, 124, 125 and 126) show a reduction of the ILCP values compared to the wild type case. Conversely, in the APO form, we observe a slight increase that becomes much larger when we compare the values obtained from the simulations with the substrate.

The same result is consistently observed for all the other comparisons, namely for the ILCP values of BLAI versus those of BV6 relative to the other reference distances and for the ILCP values of BLAI compared to the ones of BMDR (see the ILCP graphs in the Supporting Information).

Based on these observations, we have further investigated the role played by the active site dynamics and correlations in the molecular recognition process. In particular, we have computed the number of aminoacids that efficiently communicate with the active

site residues at distances larger than the reference values (δ = 10 Å, 15 Å and 20 Å) and, for each HIV-1 Protease genotype, we have determined the variation (ΔComm) in the number of efficient communications with respect to the APO simulation (see Tables 3-5). For the sake of completeness, it is worth noting that we have considered a residue as able to establish a long-range communication with the active site when it efficiently communicates with at least one active site residue. Analyzing Tables 3-5 and comparing the BLAI ΔComm values to the corresponding ones associated with the mutants, it is possible to observe that the presence of an inhibitor in a mutated HIV-1 Protease is associated with a reduction in the communication capability of the active site with the rest of the protein.

In the presence of mutations inducing resistance, the active site residues of the enzyme complexed with different inhibitors are characterized by a reduced communication capability compared to the wild type. Mutations distant from the active site modify the dynamic pre-organization that is needed for molecular recognition and effective binding of inhibitors. In fact, a reduced communication propensity is directly related to a decrease of internal coordination with the rest of the protein (*70, 71*). In this condition, active site residues can favorably populate alternative conformational states with reduced affinity for the inhibitor. In contrast, in the WT case, the global dynamic coordination is higher. The resultant lower degree of conformational freedom for the active site residues yields a better pre-organization for inhibitor binding.

In this context, it is important to analyze the ΔComm values corresponding to the substrate simulations. In the BV6 case we observe that the active site communication capability increases. As expected, this means that the mutations do not alter the Protease

ability to bind the substrate, keeping the enzyme active and favoring the virus survival even in the presence of inhibitor drugs. The substrate, due to its intrinsic flexibility, can initially adapt to the increased flexibility of the active site (in the unbound state) and afterwards induce a rearrangement in the Protease that leads to a new well-framed organization that is favorable for molecular recognition and productive binding.

In the BMDR simulation with the polypeptide, we notice an initial reduction in the active site communication capabilities, which seems to be in disagreement with the results obtained for BV6. However this effect was due to poor convergence of this specific simulation in the time scale analyzed. Indeed, higher ΔComm values are obtained considering only the last 10 of the 50 ns of the simulation. To further confirm this observation, we have also extended the simulation for an additional 20 nanoseconds and the analysis of the fully converged part of the trajectory shows much higher ΔComm values (see Tables 3-5) and a trend that is compatible with the results obtained for BV6. The consistency of the observations for the two mutants is also reflected in the results of the Energy Decomposition Method, as we will show in the following section.

Finally, we have completed the analysis of dynamic properties by calculating the pairwise covariance matrix of atomic displacements for each trajectory. The covariance matrix accounts for (anti) correlations in atomic motions and it can be used to highlight protein regions that move coherently in a correlated or anti-correlated ways. Principal Component Analysis (PCA) (*82-83*), a.k.a. Essential Dynamics (ED), reduces the dimensionality of the covariance matrix by diagonalization. This method describes functional and dominant protein motions that are represented by the matrix eigenvectors and eigenvalues, emphasizing their amplitudes and directions. The dominant eigenvectors

of the covariance matrix identify the essential dynamical subspace for a certain complex. In this context, we have calculated the agreement between the essential dynamical subspaces of the HIV-1 protease in different situations (Wild Type and mutants both in the unbound state and complexed with substrate/inhibitors) using a parameter known as the Root Weighted Square Inner Product, RWSIP (see the "Material and Methods" section for its definition). RWSIP was shown to provide an accounting of the degree of agreement between two dynamical spaces (*72*). However, in our case, the RWSIP calculations and analysis does not provide any strong signals differentiating the essential modes of motion of the protein native forms from the ones of the mutants (see Supplementary Information for the RWSIP table).

This observation is further confirmed by the analysis of the residue based root mean square fluctuation (rmsf) profiles obtained by projecting each trajectory on the three main essential eigenvectors associated with the largest eigenvalues. No significant rearrangements are observed, with the exception of non-negligible flap movements, especially in the BMDR multi-drug resistant mutant (see Supporting Information for the rmsf graphs associated with the three main eigenvalues). Many other investigators have already observed or hypothesized similar effects of mutations on flap motions (*38, 42, 43*).

Summarizing, our results suggest that, while global dynamics and main functional motions are preserved among different complexes with local flexibility changes at the single residue level, it is possible to highlight links between local interactions, short timescale fluctuations and residue-residue couplings that result in biologically relevant functions (ILCP and ΔComm analysis).

In other words, our results provide a semiquantitative view of the effect of mutations on the internal coupling dynamics of the protein, which eventually results in hindering the active site conformational selection for efficient inhibitor binding.

The mutations, while not affecting the native and enzymatically active structure, thereby preserving catalytic activity, induces the protein to sample states that are less pre-organized and, thereby, are less favorable for inhibitor binding. One possible mechanism for the virus to achieve this goal would consist in selecting and evolving local networks of interactions, through mutations, that make the Protease sample states and motions that are unfavorable for inhibitor binding. This aspect is reflected by the correlations/fluctuations (Communication Propensities) described above and it is modulated by the sequence differences.

Overall, our results are consistent with previous observations by Schiffer and coworkers, who proposed a structural mechanism for ligand recognition that is conserved both in the presence and in the absence of drug-resistance mutations. Active site and flap motions due to mutations upon binding of the ligand were postulated to negatively impact on inhibitor binding, but not on substrates (*84*). In this context, our results provide an atomic-resolution, semi-quantitative model for the mutations-induced modulation of the dynamics underlying molecular recognition for substrate/ligand binding.

In the free energy landscape perspective of protein dynamics and function, this mechanism is consistent with a hierarchical pre-organization of possible dynamic sub-states that the Protease could select by means of mutations in order to escape inhibition. In fact, the selection of dynamic states through mutations distant from the active site would allow HIV-1 PR to retain activity, while generating mutants with lower inhibitor

affinity that determine drug resistance. The concept of hierarchy of sub-states has been already investigated (*85-88*) together with the idea that preferred relatively small and local fluctuations in enzymes determine the states responsible for molecular recognition and for the catalytically active conformations.

Energy Analysis. The analysis of communication propensities highlighted interesting coordinated motional differences between the *apo*, substrate- and inhibitor-bound HIV-1 PR molecules, showing how mutations mainly affect the dynamics of conformational selection for inhibitor-recognition. Our next step was to investigate possible energetic couplings of residue pairs that define the physically connected networks linking active sites to distant mutations in the tertiary structure.

To accomplish this task, we have carried out an analysis of all the MD trajectories using the EDM (*73-76*). As already mentioned, the method aims at identifying relevant energetic couplings between residues through a principal component decomposition of the non-bonded interaction energy matrix for the whole protein. More specifically, the energy decomposition approach allows one to obtain an effective mean coupling energy between all the residue-pairs and to map the principal energetic interactions in the native state of the protein. This information can then be used to identify the most relevant networks of inter-aminoacid interactions connecting different sites. Furthermore, it was previously shown that, in good agreement with experimental findings, this method efficiently accounts for several properties such as the effects of mutations in terms of stability variation and the modulation of local and global conformational dynamics (*73-76*). These results suggest the possibility of visualizing networks of principal energetic

interactions between pairs of residues and to explain the effects of long-range perturbations in proteins.

Along this line of thought, the application of the method to the different complexes studied herein has allowed for the determination of the residues responsible for the most relevant energetic interactions in the different situations. These crucial residues have afterwards been projected onto the HIV-1 Protease 3D structure in order to make our comparisons easier (see Figures 4-6). In particular, for the sake of simplicity and without losing generality, in all the figures that we will analyze in this section we have depicted only one of the two HIV-1 PR monomers and, furthermore, we have provided two different monomer orientations in order to better display the differences.

At first, we have focused on the APO forms and we have compared the wild type energy couplings with those associated with the mutants. The results are shown in Figures 4A and 4B, where we can observe that a large number of "hot spots" turn off in BV6_APO and BMDR_APO, respectively, with respect to the wild type case. For the sake of clarity, it is worthwhile to point out that in Figures 4-6 the residues that turn off (with respect to a reference situation) are represented with red van der Waals spheres, those that turn on with blue van der Waals spheres and common "hot aminoacids" with purple van der Waals spheres.

Afterwards, we have also considered the substrate and inhibitors effects on the modulation of the energetic coupling networks. In particular, in Figure 5 we have examined the differences between corresponding "unbound" and "substrate-bound" interaction networks. Strikingly, while the BLAI_APO "hot residues" generally correspond to the ones of BLAI_SUB (see Figure 5A), in the mutants most of the wild

type strongly coupled centers, which were not present in BV6_APO and BMDR_APO, turn on again in presence of the substrate (see Figures 5B and 5C). The combined effects of substrate and mutations cooperate in recovering the same network of principal couplings observed in the active wild type Protease.

Finally, we have compared the energetically crucial residues for substrate-complexed Proteases with the ones characterizing the different Proteases in the presence of inhibitors. In Figure 6 we have reported the Ritronavir case. For the wild type the organization of the principal interaction networks is almost equivalent to that of the substrate bound case (see Figure 6A). In the other two cases, featuring mutations inducing resistance, many residues that are classified as "hot spots" in the presence of the substrate turn off when HIV-1 PR is bound to inhibitors (see Figures 6B and 6C).

These results suggest a role of mutations linked to a modulation of the energetic properties of the protein: many "hot spot residues" responsible for the principal interaction networks in the unbound wild type situation, disappear in BV6_APO and BMDR_APO. This observation supports the idea that specific sites play a key role in stabilizing and selecting the protein conformation necessary for the molecular recognition of inhibitors. The disappearance of a number of stabilizing hot spots in the apo state of mutated HIV-1 PRs indicates that these proteins may shift to alternative conformations available on the free energy landscape, characterized by lower drug-binding affinities.

Strikingly, these hot spot residues gain back their role in the mutated Proteases only in the presence of the substrate, showing that these specific interaction networks are necessary to achieve a catalytically active conformation and turn on again when a cleavable substrate is bound (see Figures 5A, 5B, 5C and 6A). Interestingly, the

differences between the wild type and the resistant Proteases re-appear in the presence of inhibitors (see Figures 6B and 6C).

Hence, this means that mutations render the important HIV-1 PR interaction networks non-optimal for the tight binding of inhibitors. However, these unfavorable situations for efficient inhibitor binding can be obviated by the presence of a flexible natural substrate. This may not be true for the rigid inhibitors examined herein because their low flexibility hinders the establishment of favorable interactions with the Protease, which consequently does not allow a proper reorganization of the protein energetic coupling networks.

Dynamic and Energetic Model of HIV-1 PR Drug Resistance. Long-range interactions and long-range coordination in proteins are crucial for their function and evolution. Therefore, it is not surprising that pathogenic proteins can evolve mechanisms based on these properties to escape the challenges posed by drug inhibition to their survival.

In this paper, we have shown that mutations inducing resistance in HIV-1 PR impact on the recognition of inhibitors by affecting the dynamics of long-range coordination and the energetic interaction networks responsible for the conformational organization of the distant active site that characterizes the WT and that can be efficiently targeted by known drugs.

In fact, from previous results, it is possible to see that mutations determine the organization of the energetic interactions that is reflected in a dynamic state characterized by reduced communication propensities of the active site with the rest of the protein. More specifically, the active site of the mutants in the presence of inhibitors appears to be

less correlated with the rest of the protein and consequently less pre-organized for efficacious binding when compared to the wild type situation.

It is important to observe that the substrate can react to this unfavorable situation. Thanks to its high flexibility relative to the inhibitors, it can adapt to the new environment and induce local protein rearrangements that eventually lead to a recovery of the interaction networks and of the dynamic communication properties necessary for an efficient binding. This is evident both from the communication and from the energy decomposition data that show an increased rigidity of the active site for the substrate-bound Proteases and a recovery of the energy coupling networks characterizing the active wild type APO form.

The situation in presence of inhibitors is opposite. In fact, due to the inhibitor rigidity, establishing a set of interactions typical of the wild type cases fails, as it can be clearly seen from the reduced communication propensity of the Proteases active site compared to the APO case and from the variation in energetic couplings (i.e., crucial stabilizing residues for binding do not turn on in the presence of inhibitors).

A central result of this work is that a subset of residues is physically coupled to determine the global dynamics of the protein necessary for proper substrate recognition and activity. Mutations affecting these networks through long-range and allosteric coupling effects can impact inhibitor recognition properties without affecting activity.

Detecting a possible new site for the design of allosteric inhibitors of drug-resistant HIV-1 Proteases. Our investigations, highlighting specific clusters of energetically coupled residues that respond to mutations inducing resistance, have provided a model for the

complex interplay between dynamics and energetics that is associated with the variation in inhibitor affinity for the HIV-1 PR mutants. The possibility to detect these sites also allows identifying protease positions that are not subject to mutations in response to the presence of inhibitors. In fact, in the framework of the present model, they coincide with residues characterized by minimal energetic couplings with the rest of the protein. It is possible to see that regions with a minimal energetic coupling are not responsive to variations in the binding state of the protein and that their mutation does not represent an advantage for the virus. Moreover, if these residues define organized 3D substructures of the protein, they can be used as targets for the design of new allosteric inhibitors aimed at interfering with the functional dynamics and with the intermolecular protein-protein interactions present in HIV-1 PR.

To detect these sites we have calculated the *Matrix of Local Coupling Energies (MLCE)* (*80*) as described in the Materials and Methods section. This has allowed us to identify a β-hairpin structure (one for each monomer; see Figure 7), and in particular residues 11-20, as the protein sub-structure whose sequence shows the lowest energy coupling to the rest of the protein and to the active site residues. According to our hypotheses, due to the minimal coupling with the rest of the protein, mutations in this sequence should not be reflected in changes in binding affinity, and thus they would not represent an evolutionary advantage for the virus. Consequently, the occurrence of drug-resistant mutations in this region should be much lower than in other protein sites. Interestingly, analysis of the Stanford HIV database (http://hivdb.stanford.edu) confirms our assumptions and calculations. For patients treated with inhibitor cocktails, mutation frequency in this part of the sequence does not change significantly compared to

untreated patients (with the exception of mutation K20R that has been considered in the BV6 multi-drug resistant mutant BV6). Conversely, larger differences are observed for drug resistance mutations. For instance, it is possible to observe that mutation L19I occurs in the 14.0 % of untreated patients, which is comparable with the presence of the same mutation (10.3 %) in individuals treated with a cocktail of inhibitors. On the otherhand, the drug resistance mutation L90M is present in 0.2 % of untreated patients, whereas its occurrence in individuals treated with a cocktail dramatically increases (40.2 %).

Moreover, in the *MLCE* approach, the *soft spots* define the non-optimized interaction-networks in the structure that are typically found clustered on the protein surface. Previous validation of the method has shown that these sites coincide with protein interaction sites, where antibodies or small molecule ligands can bind. Strikingly, analysis of the epitopes list in the HIV immunology database of the Los Alamos National Laboratory (http://www.hiv.lanl.gov) reveal that the 11-20 β-hairpin sequence is actually recognized by antibodies against HIV-1 PR (*89-92*).

We propose that this β-hairpin may also be a target for the development of allosteric inhibitors of HIV-1 PR. Small molecules binding to this region would have the potential to perturb the normal Protease functional dynamics and to interfere with possible protein-protein interactions in which the Protease is involved. To this end several theoretical methods ranging from large scale docking and screening of multiple compounds to receptor based pharmacophore design can be applied. Such an approach will clearly expand the chemical space and diversity of known HIV-1 PR inhibitors, providing an

entry point for small molecules that target a region where mutations and the onset of drug resistance have low probability.

Conclusions

The mutation of wild type HIV-1 PR represents one of the key problems that hinder the development of a definitive treatment against the HIV virus. In this context, the HIV-1 PR, which is one of the most important targets to block the progression of HIV, is a perfect example because HIV-1 PR mutants appear after some weeks of therapy and show strong resistance to the current generation of inhibitors. Therefore, the development of new drugs is of significant importance, but it is also obvious that in order to optimize the design of new plausible drug molecules, it is desirable to fully understand the role that is played by HIV-1 PR mutations in drug resistance.

In order to address this problem we have analyzed MD simulations performed on the LAI wild type of the Subtype B / HIV-1 PR and on two related multidrug resistant mutants, both unbound and "complexed" with inhibitors or with a polypeptide that mimics the "true" substrate. In particular, we studied intra-protein communication pathways within wild type and mutated HIV-1 proteases by means of a communication analysis and via the energy decomposition method.

The results obtained from these analyses have revealed that HIV-1 PR mutations alter the energetic organization and dynamic coupling, which influences the inhibitor binding properties. In fact, the non-optimal energetic scaffold of mutated HIV-1 PR is associated with altered dynamics of the active site that affects inhibitor and substrate binding relative to the wild type case. However, as one might expect, we have found that the

relatively flexible substrate is able to successfully restore favorable protein-ligand contacts enabling it to efficiently bind to the active site. Moreover, we found that the relatively rigid inhibitors are unable to reverse the unfavorable situation and their binding to HIV-1 PR is less effective than in the wild type. Based on these models and approaches, we were also able to rationally identify and validate a region of the protein that is characterized by a low frequency for drug resistant mutations and leading us to propose it as a possible target for structure-based design of new allosteric inhibitors.

Hence, our study suggests that next generation of HIV-1 PR inhibitors should be able to adapt, at least temporarily, to the unfavorable environment of mutated Proteases in order to restore the required wild type pre-organization for the molecular recognition process.

We believe that in order to discover a second-generation of HIV-1 PR inhibitors, the present insights, together with other recent efforts examining HIV-1 PR mutants, indicate the need for novel strategies that fully model both ligand and protein dynamics into the drug design and discovery process.

Acknowledgment

We would like to thank Massimiliano Meli, Guido Scarabelli and Rubben Torella for helpful discussions.

Supporting Information Available

Tables S1-S3: average, minimum and maximum values of the communication propensities for each performed simulation; Table S4: RWSIP values; Figures S1-S6: Root Mean square deviations of the $C\alpha$ atoms with respect to the starting structure for

each MD simulation; Figures S7-S13: ILCP values for BLAI (reference distances: 15 and 20 Å), BV6 (reference distances: 15, and 20 Å) and BMDR (reference distances: 10, 15 and 20 Å); Figures S14-S31: RMSf (*Root Mean Square Fluctuation*) profiles corresponding to the first three eigenvectors of the covariance matrix for each performed simulation.

References

1. Katzenstein, D. (2006) Diversity, Drug Resistance, and the Epidemic of Subtype C HIV-1 in Africa, *J. Infect. Dis. 194*, S45-S50.
2. Hammer, S. M., Squires, K. E., Hughes, M. D., Grimes, J. M., Demeter, L. M., Currier, J. S., Eron, J. J. Jr., Feinberg, J. E., Balfour, H. H., Jr., Deyton, L. R., Chodakewitz, J. A., and Fischl, M. A. (1997) A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team, *N. Engl. J. Med. 337*, 725-733.
3. Deeks, S. G., Smith, M., Holodniy, M., and Kahn, J. O. (1997) HIV-1 protease inhibitors. A review for clinicians, *J. Am. Med. Assoc. 277*, 145-153.
4. Carpeenter, C. C. J., Fischl, M. A., Hammer, S. M., Hirsch, M. S., Jacobsen, D. M., Katzenstein, D. A., Montaner, J. S. G., Richman, D. D., Saag, M. S., Schooley, R. T., Thompson, M. A., Vella, S., Yeni, P. G., and Volberding, P. A. (1997) Antiretroviral therapy for HIV infection in 1997. Updated recommendations of the International AIDS Society-USA panel, *J. Am. Med. Assoc. 277*, 1962-1969.
5. Wlodawer, A., and Erickson, J. W. (1993) Structure-based inhibitors of HIV-1 protease, *Annu. Rev. Biochem*. *62*, 543-585.
6. Debouck C. (1992) The HIV-1 protease as a therapeutic target for AIDS, *AIDS Res. Hum. Retroviruses 8*, 153-164.
7. Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A., Scolnick, E. M., and Sigal, I. S. (1988) Active human immunodeficiency virus

protease is required for viral infectivity, *Proc. Natl. Acad. Sci. U.S.A. 85*, 4686-4690.

8. Peng, C., Ho, B. K., Chang, T. W., and Chang, N. T. (1989) Role of human immunodeficiency virus type 1-specific protease in core protein maturation and viral infectivity, *J. Virol. 63*, 2550-2556.
9. Frankel, A. D., and Young, J. A. T. (1998) HIV-1: fifteen proteins and an RNA, *Annu. Rev. Biochem. 67*, 1-25.
10. Wlodawer, A., and Gustchina, A. (2000) Structural and biochemical studies of retroviral proteases, *Biochim. Biophys. Acta 1477*, 16-34.
11. Tomasselli, A. G., and Heinrikson, R. L. (2000) Targeting the HIV-protease in AIDS therapy: a current clinical perspective, *Biochim. Biophys. Acta 1477*, 189-214.
12. Hamelberg, D., and McCammon, J. A. (2005) Fast Peptidyl cis−trans Isomerization within the Flexible Gly-Rich Flaps of HIV-1 Protease, *J. Am. Chem. Soc. 127*, 13778-13779.
13. Hornak, V., Okur, A., Rizzo, R. C., and Simmerling, C. (2006) HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations, *Proc. Natl. Acad. Sci. U.S.A. 103*, 915-920.
14. Hornak, V., and Simmerling, C. (2007) Targeting structural flexibility in HIV-1 protease inhibitor binding, *Drug Discov. Today 12*, 132-138.
15. Ding, F., Layten, M., and Simmerling, C. (2008) Solution Structure of HIV-1 Protease Flaps Probed by Comparison of Molecular Dynamics Simulation Ensembles and EPR Experiments, *J. Am. Chem. Soc. 130*, 7184-7185.

16. Galiano, L., Ding, F., Veloro, A. M., Blackburn, M., Simmerling, C. and Fanucci, G. E. (2009) Drug Pressure Selected Mutations in HIV-1 Protease Alter Flap Conformations, *J. Am. Chem. Soc. 131*, 430-431.

17. Roberts, N. A., Martin, J. A., Kinchington, D., Broadhurst, A. V., Craig, J. C., Duncan, I. B., Galpin, S. A., Handa, B. K., Kay, J., Krohn, A., Lambert, R. W., Merrett, J. H., Mills, J. S., Parkes, K. E. B., Redshaw, S., Ritchie, A. J., Taylor, D. L., Thomas, G. J., and Machin, P. J. (1990) Rational design of peptide-based HIV proteinase inhibitors *Science 248*, 358-361.

18. Dorsey, B. D., Levin, R. B., McDaniel, S. L., Vacca, J. P., Guare, J. P., Darke, P. L., Zugay, J. A., Emini, E. A., Schleif, W. A., Quintero, J. C., Lin, J. H., Chen, I.-W., Holloway, M. K., Fitzgerald, P. M. D., Axel, M. G., Ostovic, D., Anderson, P. S., and Huff, J. R. (1994) L-735,524: the design of a potent and orally bioavailable HIV protease inhibitor, *J. Med. Chem. 37*, 3443-3451.

19. Kaldor, S. W., Kalish, V. J., Davies, J. F. II, Shetty, B. V., Fritz, J. E., Appelt, K., Burgess, J. A., Campanale, K. M., Chirgadze, N. Y., Clawson, D. K., Dressman, B. A., Hatch, S. D., Khalil, D. A., Kosa, M. B., Lubbehusen, P. P., Muesing, M. A., Patick, A. K., Reich, S. H., Su, K. S., and Tatlock, J. H. (1997) Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease, *J. Med. Chem. 40*, 3979-3985.

20. Kempf, D. J., Marsh, K.C., Denissen, J. F., McDonald, E., Vasavanonda, S., Flentge, C. A., Green, B. E., Fino, L., Park, C. H., Kong, X. P., Wideburg, N. E., Saldivar, A., Ruiz, L., Kati, W. M., Sham, H. L., Robins, T., Stewart, K. D., Hsu,

A., Plattner, J. J., Leonard, J. M., and Norbeck, D. W. (1995) ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans, *Proc. Natl. Acad. Sci. U.S.A. 92*, 2484-2488.

21. Lee, T. Y., Le, V.-D., Lim, D. Y., Lin, Y.-C., Morris, G. M., Wong, A. L., Olson, A. J., Elder, J. H., and Wong, C. H. (1999) Development of a New Type of Protease Inhibitors, Efficacious against FIV and HIV Variants, *J. Am. Chem. Soc. 121*, 1145-1155.
22. Li, M., Morris, G. M., Lee, T., Laco, G. S., Wong, C. H., Olson, A. J., Elder, J. H., Wlodawer, A., and Gustchina, A. (2000) Structural studies of FIV and HIV-1 proteases complexed with an efficient inhibitor of FIV protease, *Proteins 38*, 29-40.
23. Condra, J. H., Schleif, W. A., Blahy, O. M., Gabryelski, L. J., Graham, D. J., Quintero, J. C., Rhodes, A., Robbins, H. L., Roth, E., Shivaprakash, M., Titus, D., Yang, T., Teppler, H., Squires, K. E., Deutsch, P. J., and Emini, E. A. (1995) In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors, *Nature 374*, 569-571.
24. Molla, A., Korneyeva, M., Gao, Q., Vasavanonda, S., Shipper, P. J., Mo, H.-M., Markowitz, M., Chernyavskiy, T., Niu, P., Lyons, N., Hsu, A., Granneman, G. R., Ho, D. D., Boucher, C. A. B., Leonard, J. M., Norbeck, D. W., and Kempf, D. J. (1996) Ordered accumulation of mutations in HIV protease confers resistance to ritonavir, *Nat. Med. 2*, 760-766.
25. Erickson, J. W., and Burt, S. K. (1996) Structural mechanisms of HIV drug resistance, *Annu. Rev. Pharmacol. Toxicol. 36*, 545-571.

26. Wlodawer, A., and Vonrasek, J. (1998) Inhibitors of HIV-1 protease: a major success of structure-assisted drug design, *Annu. Rev. Biophys. Biomol. Struct. 27*, 249-284.

27. Dunn, B. M. (2002) Anatomy and pathology of HIV-1 peptidase, *Essays Biochem*. *38*, 113-127.

28. Clemente, J. C., Robbins, A., Gaña, P., Paleo, M. R., Correa, J. F., Villaverde, M. C., Sardina, F. J., Govindasamy, L., Agbandje-McKenna, M., McKenna, R., Dunn, B. M., and Sussman, F. (2008) Design, synthesis, evaluation, and crystallographic-based structural studies of HIV-1 protease inhibitors with reduced response to the V82A mutation, *J. Med. Chem. 51*, 852-860.

29. Bannwarth, L., Rose, T., Dufau, L., Vaderesse, R., Dumond, J., Jamart-Grégoire, B., Pannecouque, C., De Clercq, E., and Reboud-Ravaux, M. (2009) Dimer disruption and monomer sequestration by alkyl tripeptides are successful strategies for inhibiting wild-type and multidrug-resistant mutated HIV-1 proteases, *Biochemistry 48*, 379-387.

30. Ala, P. J., Huston, E. E., Klabe, R. M., Jadhav, P. K., Lam, P. Y. S., and Chang, C.-H. (1998) Counteracting HIV-1 protease drug resistance: structural analysis of mutant proteases complexed with XV638 and SD146, cyclic urea amides with broad specificities, *Biochemistry 37*, 15042-15049.

31. Luque, I., Todd, M. J., Gómez, J., Semo, N., and Freire, E. (1998) Molecular basis of resistance to HIV-1 protease inhibition: a plausible hypothesis, *Biochemistry 37*, 5791-5797.

32. Rose, R. B., Craik, C. S., and Stroud, R. M. (1998) Domain flexibility in retroviral proteases: structural implications for drug resistant mutations, *Biochemistry 37*, 2607-2621.

33. Velázquez-Campoy, A., Todd, M. J., and Freire, E. (2000) HIV-1 protease inhibitors: enthalpic versus entropic optimization of the binding affinity, *Biochemistry 39*, 2201-2207.

34. Todd, M. J., Luque, I., Velázquez-Campoy, A., and Freire, E. (2000) Thermodynamic basis of resistance to HIV-1 protease inhibition: calorimetric analysis of the V82F/I84V active site resistant mutant, *Biochemistry 39*, 11876-11883.

35. Velázquez-Campoy, A., Luque, I., Todd, M. J., Milutinovich, M., Kiso, Y., and Freire, E. (2000) Thermodynamic dissection of the binding energetics of KNI-272, a potent HIV-1 protease inhibitor, *Protein Sci. 9*, 1801-1809.

36. Wang, W., and Kollman, P. A. (2001) Computational study of protein specificity: The molecular basis of HIV-1 protease drug resistance, *Proc. Natl. Acad. U.S.A. 98*, 14937-14942.

37. Clemente, J. C., Hemrajani, R., Blum, L. E., Goodenow, M. M., and Dunn, B. M. (2003) Secondary mutations M36I and A71V in the human immunodeficiency virus type 1 protease can provide an advantage for the emergence of the primary mutation D30N, *Biochemistry 42*, 15029-15035.

38. Clemente, J. C., Moose, R. E., Hemrajani, R., Whitfors, L. R. S., Govinasamy, L., Reutzel, R., McKenna, R., Agbandje-McKenna, M., Goodenow, M. M., and

Dunn, B. M. (2004) Comparing the accumulation of active- and nonactive-site mutations in the HIV-1 protease, *Biochemistry 43*, 12141-12151.

39. Clemente, J. C., Coman, R. M., Thiaville, M. M., Janka, L. K., Jeung, J. A., Nukoolkarn, S., Govindasamy, L., Agbandje-McKenna, M., McKenna, R., Leelamanit, W., Goodenow, M. M., and Dunn, B. M. (2006) Analysis of HIV-1 CRF_01 A/E protease inhibitor resistance: structural determinants for maintaining sensitivity and developing resistance to atazanavir, *Biochemistry 45*, 5468-5477.
40. Sanches, M., Krauchenco, S., Martins, N. H., Gustchina, A., Wlodawer, A., and Polikarpov, I. (2007) Structural characterization of B and non-B subtypes of HIV-protease: insights into the natural susceptibility to drug resistance development, *J. Mol. Biol. 369*, 1029-1040.
41. Foulkes-Murzycki, J. E., Scott, W. R. P., and Schiffer, C. A. (2007) Hydrophobic sliding: a possible mechanism for drug resistance in human immunodeficiency virus type 1 protease, *Structure 15*, 225-233.
42. Perryman, A. L., Lin, J. H., and McCammon, J. A. (2004) HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: Possible contributions to drug resistance and a potential new target site for drugs, *Prot. Sci. 13*, 1108-1123.
43. Perryman, A. L., Lin, J. H., and McCammon, J. A. (2004) HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: Possible contributions to drug resistance and a potential new target site for drugs (Correction), *Prot. Sci. 13*, 1434-1434.

44. Piana, S., Carloni, P., and Rothlisberger, U. (2002) Drug resistance in HIV-1 protease: Flexibility-assisted mechanism of compensatory mutations, *Protein Sci. 11*, 2393-2402.

45. Smock, R. G., and Gierasch, L. M. (2009) Sending signals dinamically, *Science 324*, 198-203.

46. Lockless, S. W., and Ranganathan, R. (1999) Evolutionarily conserved pathways of energetic connectivity in protein families, *Science 286*, 295-299.

47. Fuentes, E. J., Der, C. J., and Lee, A. L. (2004) Ligand-dependent dynamics and intramolecular signaling in a PDZ domain, *J. Mol. Biol. 335*, 1105-1115.

48. Dhulesia, A., Gsponer, J., and Vendruscolo, M. (2008) Mapping of Two Networks of Residues That Exhibit Structural and Dynamical Changes upon Binding in a PDZ Domain Protein, *J. Am. Chem. Soc. 130*, 8931-8939.

49. Ota, N., and Agard, D. A. (2005) Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion, *J. Mol. Biol. 351*, 345-354.

50. De Los Rios., P., Cecconi, F., Pretre, A., Dietler, G., Michielin, O., Piazza, F., and Juanico, B. (2005) Functional Dynamics of PDZ Binding Domains: A Normal-Modes Analysis, *Biophys. Journ. 89*, 14-21.

51. Kong, Y., and Karplus, M. (2009) Signaling pathways of PDZ2 domain: a molecular dynamics interaction correlation analysis, *Proteins 74*, 145-154.

52. Martin, P., Vickrey, J. F., Proteasa, G., Jimenez, Y. L., Wawrzak, Z., Winters, M. A., Merigan, T. C., and Kovari, L. C. (2005) "Wide-open" 1.3 A structure of a multidrug-resistant HIV-1 protease as a drug target, *Structure 13*, 1887-1895.

53. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *J. Med. Chem. 47*, 1739-1749.

54. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., and Banks, J. L. (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening, *J. Med. Chem. 47*, 1750-1759.

55. Case, D. A., Darden, T. A., Cheatman, T. E. III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Merz, K. M. Jr., Pearlman, D. A., Crowley, M., Walker, R. C., Zhang, W., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Wong, K. F., Paesani, F., Wu, X., Brozell, S., Tsui, V., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Beroza, P., Mathews, D. H., Schafmeister, C., Ross, W. S., and Kollman, P. A. (2009) AMBER 9, University of California - San Fancisco.

56. Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006) Automatic atom type and bond type perception in molecular mechanical calculations, *J. Mol. Graphics Model. 25*, 247-260.

57. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field, *J. Comput Chem. 25*, 1157-1174.

58. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, *Electrophoresis 18*, 2714-2723.

59. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins 65*, 712-725.

60. Jorgensen, W. L., Chandrasekar, J., Madura, J., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys. 79*, 926-935.

61. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems, *J. Chem. Phys. 98*, 10089-10092.

62. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method, *J. Chem. Phys. 103*, 8577-8593.

63. Crowley, M. F., Darden, T. A., Cheatman, T. E. III, and Deerfield, D. W. II (1997) Adventures in improving the scaling and accuracy of a parallel molecular dynamics program, *J. Supercomput. 11*, 255.

64. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes, *J. Comput. Phys. 23*, 327-341.

65. Pastor, R. W., Brooks, B. R., and Szabo, A. (1988) An analysis of the accuracy of Langevin and molecular dynamics algorithms, *Mol. Phys. 65*, 1409-1419.

66. Loncharich, R. J., Brooks, B. R., and Pastor, R. W. (1992) Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide, *Biopolymers 32*, 523-535.

67. Izaguirre, J. A., Catarello, D. P., Wozniak, J. M., and Skeel, R. D. (2001) Langevin stabilization of molecular dynamics, *J. Chem. Phys. 114*, 2090-2098.

68. Chennubhotla, C., and Bahar, I. (2007) Signal Propagation in Proteins and Relation to Equilibrium Fluctuations, *PLoS Comput. Biol. 3*, e172.

69. Chennubhotla, C., Yang, Z., and Bahar, I. (2008) Coupling between global dynamics and signal transduction pathways: a mechanism of allostery for chaperonin GroEL, *Mol. Biosyst. 4*, 287-292.

70. Morra, G., Verkhivker, G., and Colombo, G. (2009) Modeling Signal Propagation Mechanisms and Ligand-Based Conformational Dynamics of the Hsp90 Molecular Chaperone Full-Length Dimer, *PLoS Comput. Biol. 5*, e1000323.

71. Colombo, G., Meli, M., Morra, G., Gabizon, R., and Gasset, M. (2009) Methionine Sulfoxides on Prion Protein Helix-3 Switch on the α-Fold Destabilization Required for Conversion, *PLos ONE 4*, e4296.

72. Carnevale, V., Pontiggia, F., and Micheletti, M. (2007) Structural and dynamical alignment of enzymes with partial structural similarity, *J. Phys. Condens. Matter. 19*, 285206.

73. 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*, 113-124.

74. Colacino, S., Tiana, G. , and Colombo, G. (2006) Similar folds with different stabilization mechanisms: the cases of Prion and Doppel proteins, *BMC Struct. Biol. 6*, 17.

75. Ragona, L., Colombo, G., Catalano, M., and Molinari, H. (2005) Determinants of protein stability and folding: comparative analysis of beta-lactoglobulins and liver basic fatty acid binding protein, *Proteins 61*, 366-376.

76. Morra, G., and Colombo, G. (2008) Relationship between energy distribution and fold stability: Insights from molecular dynamics simulations of native and mutant proteins, *Proteins 72*, 660-672.

77. Sharp, K. A., Honig, B. (1990) Electrostatic interactions in macromolecules: theory and applications, *Annu. Rev. Biophys. Biophys. Chem. 19*, 301-332.

78. Davis, M. E., McCammon, J. A. (1990) Electrostatics in biomolecular structure and dynamics, *Chem. Rev. 90*, 509-521.

79. Daura, X., Gademann, K., Jaun, B., Seebacj, D., van Gunsteren, W. F., and Mark, A. E. (1999) Peptide Folding: When Simulation Meets Experiment, *Angew. Chem. Int. Ed. 38*, 236-240.

80. Scarabelli, G., Morra, G., and Colombo, G. (2010) Predicting Interaction Sites from the Energetics of Isolated Proteins: A New Approach to Epitope Mapping, *Biophys. Journ. 98*, 1-10 (in press).

81. Swain, J. F., and Gierasch, L. M. (2006) The changing landscape of protein allostery, *Curr. Op. Struct. Biol. 16*, 102-108.

82. Amadei, A., Linnsen, A. B. M., and Berendsen, H. J. C. (1993) Essential dynamics of proteins, *Proteins 17*, 412-425.

83. Van Alten, D. M. F., Amadei, A., Linnsen, A. B. M., Eijsink, V. G. H., Vriend, G., and Berendsen, H. J. C. (1995) The essential dynamics of thermolysin: confirmation of the hinge-bending motion and comparison of simulations in vacuum and water, *Proteins 22*, 45-54.

84. Prabu-Jeyabalan, M., Nalivaika, E. A., Romano, K., and Schiffer, C. A. (2006) Mechanism of substrate recognition by drug-resistant human immunodeficiency virus type 1 protease variants revealed by a novel structural intermediate, *J. Virol. 80*, 3607-3616.

85. Austin, R. H., Beeson, K. W., Eisenstein, L., Frauenfelder, H., and Gunsalus, I. C. (1975) Dynamics of ligand binding to myoglobin, *Biochemistry 14*, 5355-5373.

86. Frauenfelder, H., Sligar, S. G., and Wolynes, P. G. (1991) The energy landscapes and motions of proteins, *Science 254*, 1598-1603.

87. Frauenfelder, H., McMahon, B. H., and Fenimore, P. W. (2003) Myoglobin: the hydrogen atom of biology and a paradigm of complexity, *Proc. Natl. Acad. Sci. U.S.A. 100*, 8615-8617.

88. Henzler-Wildman, K. A., Thai, V., Lei, M., Ott, M., Wolf-Watz, M., Fenn, T., Pozharski, E., Wilson, M. A., Petsko, G. A., Karplus, M., Hübner, C. G., and Kern, D. (2007) Intrinsic motions along an enzymatic reaction trajectory, *Nature 450*, 838-844.

89. Wilson, C. C., McKinney, D., Anders, M., MaWhinney, S., Forster, J., Crimi, C., Southwood, S., Sette, A., Chesnut, R., Newman, M. J., and Livingston, B. D. (2003) Development of a DNA vaccine designed to induce cytotoxic T

lymphocyte responses to multiple conserved epitopes in HIV-1, *J. Immunol. 171*, 5611-5623.

90. McKinney, D. M., Skvoretz, R., Livingston, B. D., Wilson, C. C., Anders, M., Chesnut, R. W., Sette, A., Essex, M., Novitsky, V., and Newman, M. J. (2004) Recognition of variant HIV-1 epitopes from diverse viral subtypes by vaccine-induced CTL, *J. Immunol. 173*, 1941-1950.
91. Wilson, C. C., Newman, M. J., Livingston, B. D., MaWhinney, S., Forster, J. E., Scott, J., Schooley, T., and Benson, C. A. (2008) Clinical phase 1 testing of the safety and immunogenicity of an epitope-based DNA vaccine in human immunodeficiency virus type 1-infected subjects receiving highly active antiretroviral therapy, *Clin. Vaccine Immunol. 15*, 986-994.
92. Gorse, G. J., Baden, L. R., Wecker, M., Newman, M. J., Ferrari, G., Weinhold, K. J., Livingston, B. D., Villafana, T. L., Li, H., Noonan, E., and Russel, N. D. (2008) Safety and immunogenicity of cytotoxic T-lymphocyte poly-epitope, DNA plasmid (EP HIV-1090) vaccine in healthy, human immunodeficiency virus type 1 (HIV-1)-uninfected adults, *Vaccine 26*, 215-223.

Table 1 - PDB files used as starting structures for the MD simulations. The letter R preceding the PDB code is used to indicate that the inhibitor has been removed from the original structure, while the letter D means that the original inhibitor has been replaced with a new docked and modeled substrate (inhibitor or polypeptide) in the active site.

	BLAI	BV6	BMDR
APO	R-1HSG	R-1SGU	1TW7
INDINAVIR	1HSG	1SGU	D-1RQ9
NELFINAVIR	D-1HSG	D-1SGU	D-1RQ9
RITRONAVIR	D-1HSG	1SH9	D-1RQ9
TL-3	D-1HSG	D-1SGU	D-1RQ9
SUBSTRATE	D-1HSG	D-1SGU	D-1RQ9

Table 2 - "Efficient communication threshold" (η) for the systems taken into account.

	BLAI	BV6	BMDR
APO	0.0090	0.0085	0.0115
INDINAVIR	0.0097	0.0086	0.0099
NELFINAVIR	0.0080	0.0097	0.0142
RITRONAVIR	0.0080	0.0079	0.0115
TL-3	0.0056	0.0073	0.0130
SUBSTRATE	0.0077	0.0086	0.0128

Table 3 – Variations in the number of efficient communications established with the active site with respect to *apo* simulations (reference distance: 10 Å). The number of new and lost communications is reported in parenthesis.

Ligand	ΔComm (BLAI)	ΔComm (BV6)	ΔComm (BMDR)
INDINAVIR	+25 (38 / 13)	-16 (20 /36)	-26 (3 / 29)
NELFINAVIR	+24 (32 / 8)	-62 (0 / 62)	0 (13 / 13)
RITRONAVIR	+43 (47 / 4)	-22 (7 / 29)	-22 (3 / 25)
TL-3	+37 (41 / 4)	+3 (15 /12)	-6 (9 / 15)
SUBSTRATE	+17 (28 / 11)	+21 (26 / 5)	-15 (3 / 18)
SUBSTRATE (*)	//	//	-8 (7 / 15)
SUBSTRATE (**)	//	//	+1 (13 / 12)

(*) The BMDR value has been calculated on the last 10 ns of the original MD simulation.

(**) The BMDR value has been calculated on the 20 ns extension of the MD simulation.

Table 4 – Variations in the number of efficient communications established with the active site with respect to *apo* simulations (reference distance: 15 Å). The number of new and lost communications is reported in parenthesis.

Ligand	ΔComm (BLAI)	ΔComm (BV6)	ΔComm (BMDR)
INDINAVIR	+30 (38 / 8)	+8 (19 / 11)	-24 (6 / 30)
NELFINAVIR	+28 (36 / 8)	-19 (0 / 19)	+ 10 (26 / 16)
RITRONAVIR	+37 (43 / 6)	-8 (2 / 10)	-16 (12 / 28)
TL-3	+48 (51 / 3)	+9 (15 / 6)	+2 (21 / 19)
SUBSTRATE	+17 (27 / 10)	+26 (32 / 6)	-9 (8 / 17)
SUBSTRATE (*)	//	//	0 (13 / 13)
SUBSTRATE (**)	//	//	+ 8 (18 / 10)

(*) The BMDR value has been calculated on the last 10 ns of the original MD simulation.

(**) The BMDR value has been calculated on the 20 ns extension of the MD simulation.

Table 5 – Variations in the number of efficient communications established with the active site with respect to *apo* simulations (reference distance: 20 Å). The number of new and lost communications is reported in parenthesis.

Ligand	ΔComm (BLAI)	ΔComm (BV6)	ΔComm (BMDR)
INDINAVIR	+14 (16 / 2)	+4 (7 /3)	-14 (3 / 17)
NELFINAVIR	+13 (15 / 2)	-4 (0 / 4)	+11 (17 / 6)
RITRONAVIR	+22 (26 / 4)	-3 (1 / 4)	-8 (10 / 18)
TL-3	+21 (21 / 0)	+3 (5 / 2)	+5 (13 / 8)
SUBSTRATE	+6 (10 / 4)	+10 (12 / 2)	-8 (2 /10)
SUBSTRATE (*)	//	//	-1 (8 / 9)
SUBSTRATE (**)	//	//	+10 (14 /4)

(*) The BMDR value has been calculated on the last 10 ns of the original MD simulation.

(**) The BMDR value has been calculated on the 20 ns extension of the MD simulation.

Figures Legends

Figure 1. Ribbon diagrams of HIV-1 Protease. The red and blue colored spheres represent the Cα position of mutations in the BV6 (A) and BMDR (B) mutants, respectively. The active site is highlighted in orange.

Figure 2. Intrinsic Long-Range Communication Propensities (ILCPs) for each residue of the BLAI HIV-1 Protease (namely, the wild type LAI) in all the simulations performed (δ = 10 Å). The red arrows indicate the active site residues in the Protease sequence.

Figure 3. Intrinsic Long-Range Communication Propensities (ILCPs) for each residue of the BV6 mutant (namely, the pseudo-V6 mutant) in all the simulations performed (δ = 10 Å). The red arrows indicate the active site residues in the Protease sequence.

Figure 4. "Hot spots" comparisons. The stabilizing centers are highlighted on the ribbon diagram of the 1hsg PDB structure by means of van der Waals spheres. (A) BLAI_APO *vs*. BV6_APO: red spheres indicate "hot residues" only for BLAI_APO, blue spheres "hot residues" only for BV6_APO and purple spheres common stabilizing centers; (B) BLAI_APO *vs*. BMDR_APO: red spheres indicate "hot residues" only for BLAI_APO, blue spheres "hot residues" only for BMDR_APO and purple spheres common stabilizing centers.

Figure 5. “Hot spots” comparisons. The stabilizing centers are highlighted on the ribbon diagram of the 1hsg PDB structure by means of van der Waals spheres. (A) BLAI_APO *vs*. BLAI_SUB: red spheres indicate “hot residues” only for BLAI_APO, blue spheres “hot residues” only for BLAI_SUB and purple spheres common stabilizing centers; (B) BV6_APO *vs*. BV6_SUB: red spheres indicate “hot residues” only for BV6_APO, blue spheres “hot residues” only for BV6_SUB and purple spheres common stabilizing centers; (C) BMDR_APO *vs*. BMDR_SUB: red spheres indicate “hot residues” only for BMDR_APO, blue spheres “hot residues” only for BMDR_SUB and purple spheres common stabilizing centers;

Figure 6. “Hot spots” comparisons. The stabilizing centers are highlighted on the ribbon diagram of the 1hsg PDB structure by means of van der Waals spheres. (A) BLAI_SUB *vs*. BLAI_RIT: red spheres indicate “hot residues” only for BLAI_SUB, blue spheres “hot residues” only for BLAI_RIT and purple spheres common stabilizing centers; (B) BV6_SUB *vs*. BV6_RIT: red spheres indicate “hot residues” only for BV6_SUB, blue spheres “hot residues” only for BV6_RIT and purple spheres common stabilizing centers; (C) BMDR_SUB *vs*. BMDR_RIT: red spheres indicate “hot residues” only for BMDR_SUB, blue spheres “hot residues” only for BMDR_RIT and purple spheres common stabilizing centers;

Figure 7. Ribbon diagram of HIV-1 Protease with the 11-20 β-hairpin structures (one for each monomer) highlighted in blue. These protease sub-structures represent a new target for the development of alternative allosteric inhibitors.

Figure 1-

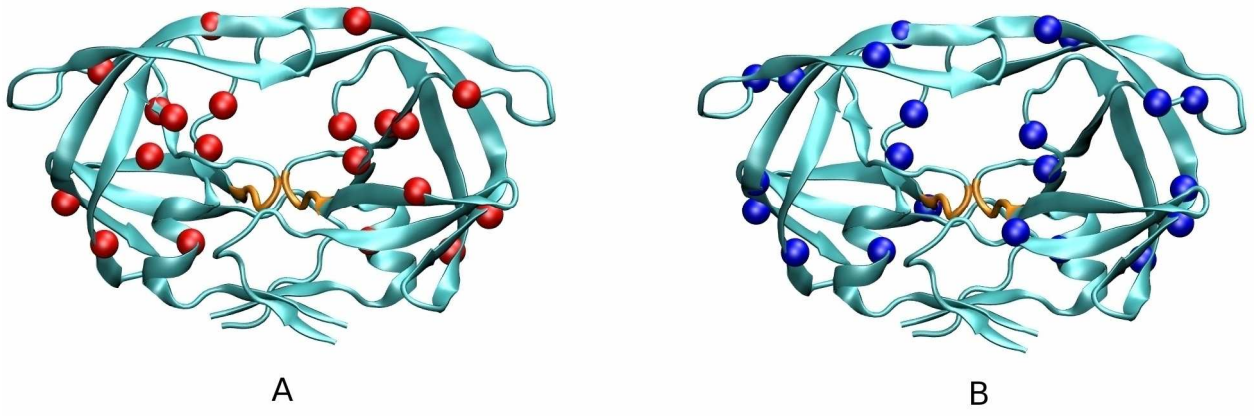


Figure 2 –

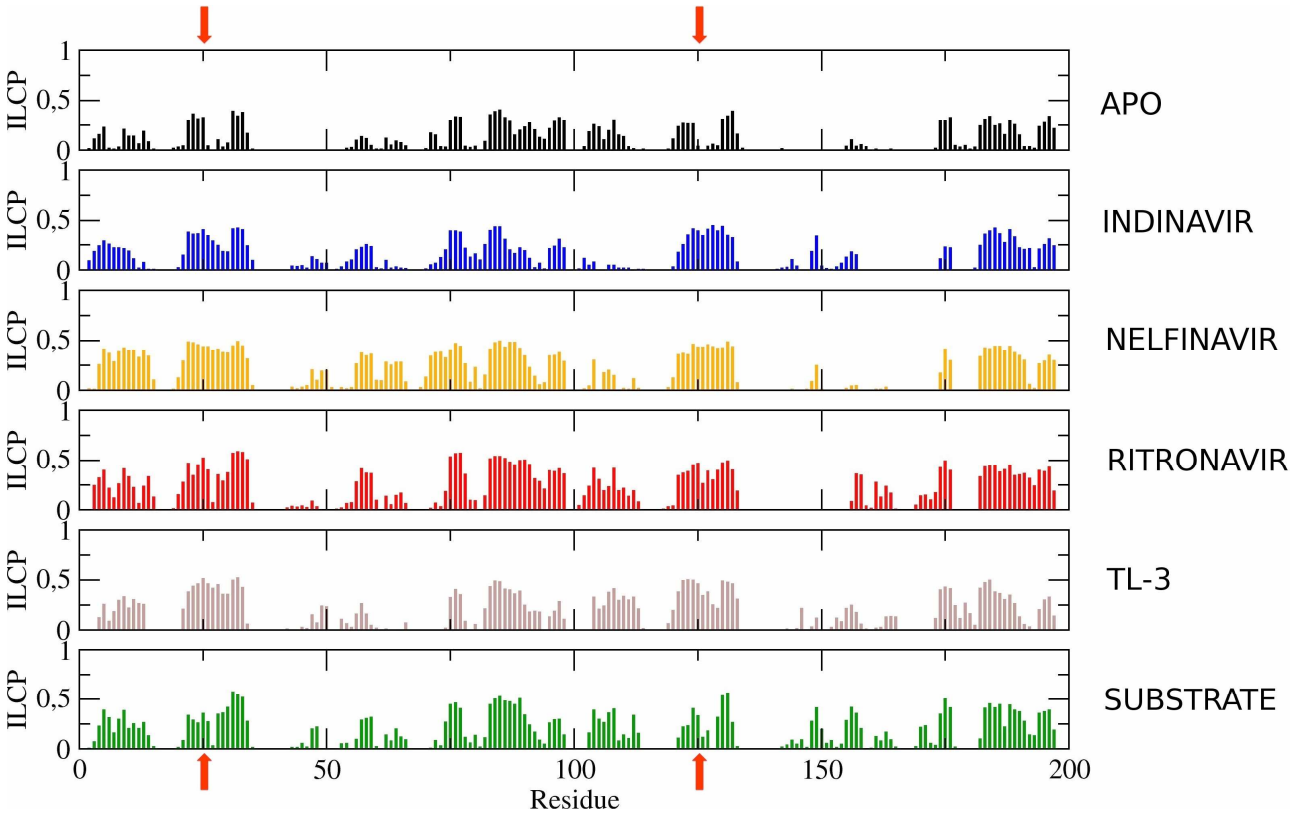

ILCP
1
0,5
0
APO
INDINAVIR
NELFINAVIR
RITRONAVIR
TL-3
SUBSTRATE
0
50
100
150
200
Residue

Figure 3 –

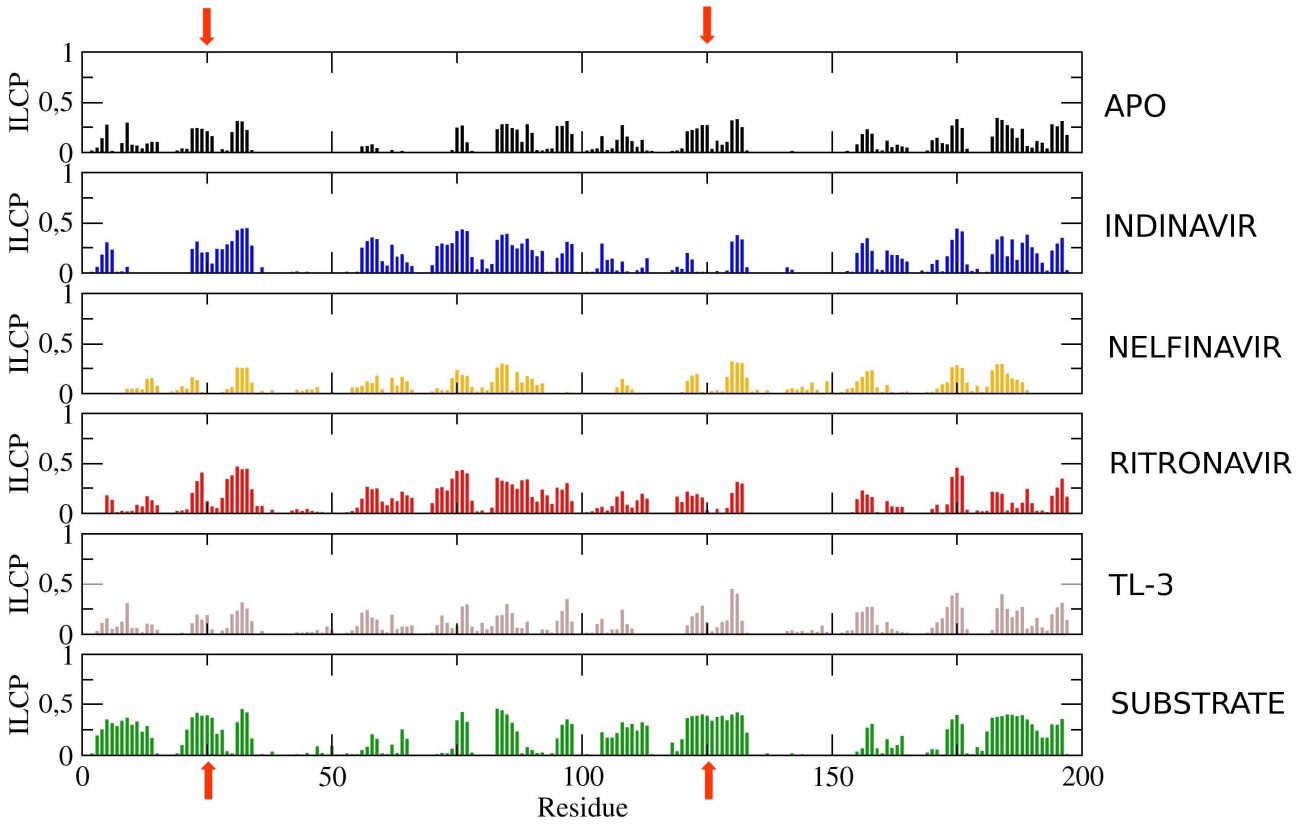

ILCP
1
0,5
0
APO
INDINAVIR
NELFINAVIR
RITRONAVIR
TL-3
SUBSTRATE
0
50
100
150
200
Residue

Figure 4 –

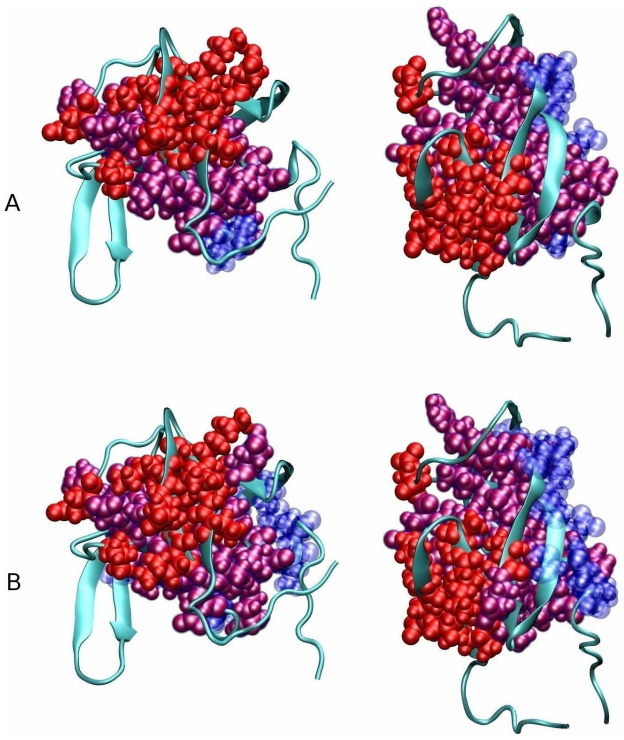


Figure 5 –

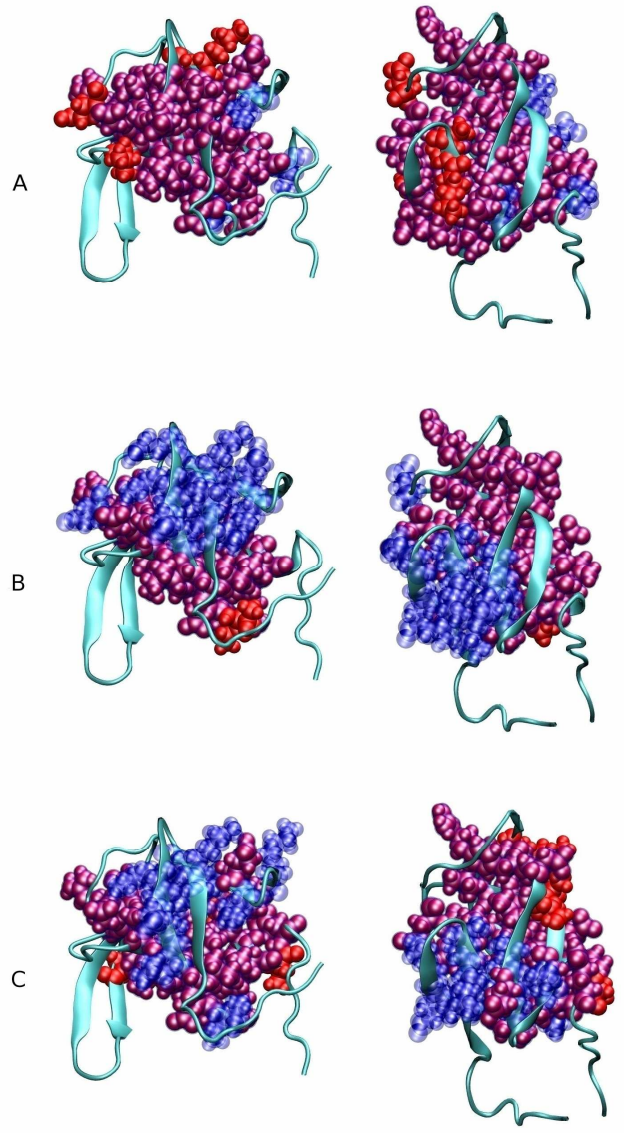


Figure 6 –

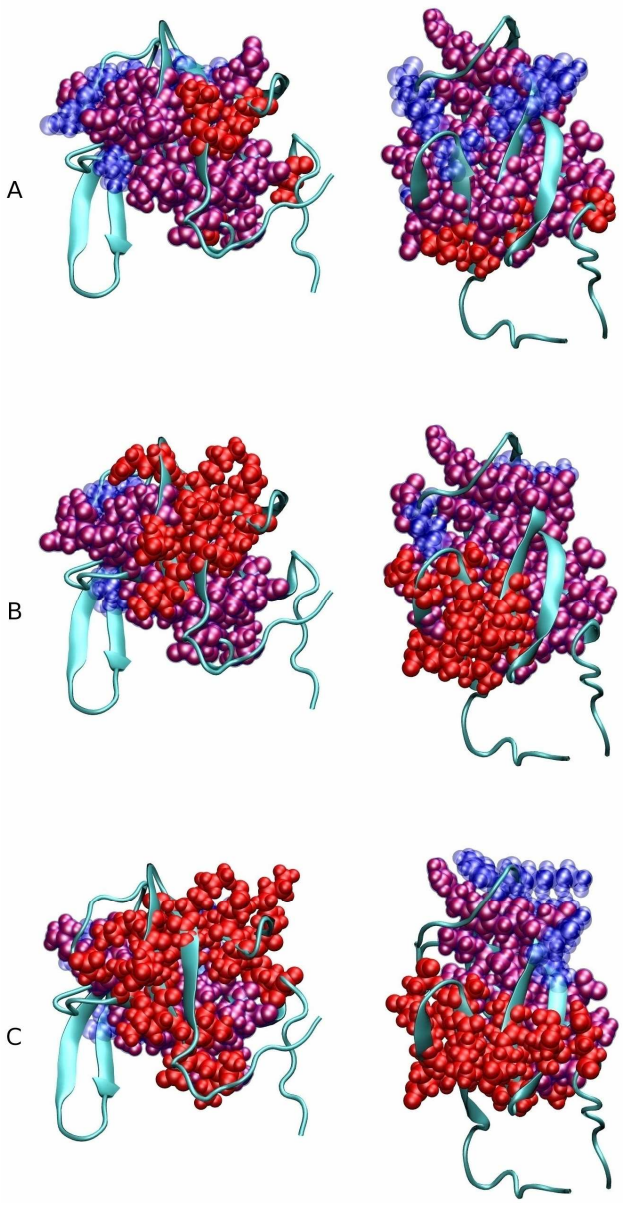
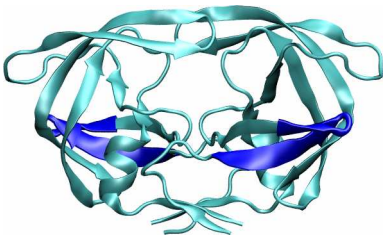
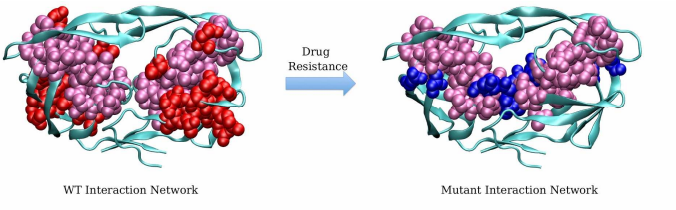


Figure 7 –



For Table of Contents Use Only



Manuscript Title: *Computational study of the resistance shown by the Subtype B / HIV-1 Protease to currently known inhibitors.*

Authors: *Alessandro Genoni, Giulia Morra, Kenneth M. Merz Jr., Giorgio Colombo.*