Computational Prediction of HIV-1 Resistance to Protease Inhibitors

Ali Hosseini[†], Andreu Alibés[†,‡], Marc Noguera-Julian[§,ǁ,⊥], Victor Gil[†], Roger Paredes[§,ǁ,⊥*], Robert Soliva[†], Modesto Orozco[†,‡,#] and Victor Guallar[†,O*]

[†] Joint BSC-CRG-IRB Research Program in Computational Biology. Barcelona Supercomputing Center. c/ Jordi Girona 29, 08034 Barcelona, Spain.

[‡] Institute for Research in Biomedicine (IRB Barcelona), Baldiri i Reixac 8, Barcelona 08028, Spain

[§] IrsiCaixa AIDS Research Institute, Hospital Universitari Germans Trias i Pujol, Badalona, Catalonia, Spain

[ǁ] Universitat Autònoma de Barcelona, Catalonia, Spain

[⊥] Universitat de Vic – Universitat Central de Catalunya, Vic, Catalonia, Spain

[#] Departament de Bioquímica. Universitat de Barcelona. Avgda Diagona 647. 08029, Spain

[O] Institució Catalana de Recerca i Estudis Avançats, Passeig Lluís Companys 23, 08010 Barcelona, Spain

Abstract

Development of mutations in HIV-1 PR hinders the activity of antiretroviral drugs, forcing changes in drug prescription. Most resistance assessments used to date rely on expert-based rules on predefined sets of stereotypical mutations; such information-driven approach cannot capture new polymorphisms nor be applied for new drugs. Computational modeling could provide a more general assessment of drug resistance, and could be made available to clinicians through the Internet. We have created a protocol involving sequence comparison and all-atom protein-ligand induced fit simulations to predict resistance at the molecular level. We first compared our predictions with experimentally determined IC_{50} of darunavir, amprenavir, ritonavir and indinavir from reference PR mutants displaying different resistance levels. We then performed analyses on a large set of variants harboring more than 10 mutations. Finally, several sequences from real patients were analyzed for amprenavir and darunavir. Our computational approach detected all genotype changes triggering high-level resistance, even those involving a large number of mutations.

INTRODUCTION

Antiretroviral therapy (ART) is one of the most effective interventions in medicine and, in particular, in HIV treatment. In ideal conditions of treatment, ART transforms a deadly disease into a chronic pathology, allowing patients to achieve life expectancy[1] and quality similar to that of non-HIV-infected individuals.[2, 3] Such efficacy, however, can be offset by HIV's ability to develop mutations conferring antiretroviral resistance in the presence of drug selective pressure[4-7] and by patient-to-patient transmission of resistant viruses. In resource-limited settings, where the HIV/AIDS toll is higher, ART is often provided without virological monitoring,[8, 9] which is being associated with an alarming increase in drug resistance.[10-12]

Antiretroviral drug resistance testing is key for clinical management[13] and epidemiologic surveillance,[1, 10-12, 14] but it is not trivial to assess. To this aim, we find multiple predictions techniques using molecular modeling[15, 16] and bioinformatics tools.[17-21] These, however, often fail when addressing real patients sequences containing a large number of mutations and/or new ones. Thus, caregivers often rely on expert-based rules to interpret predefined sets of stereotypical mutations. Such procedure assigns susceptibility scores to different drugs for each mutation encountered after sequencing, from a previously defined list; individual mutation scores are then added into a global score for each drug and combination of drugs. Such scores are usually translated into a susceptible-intermediate-resistant (SIR) interpretation. Interpretation rules based on a predefined list of mutations rely on previous knowledge and are limited in the mutation scope, so they cannot take into account the effect of non-predefined polymorphisms, mutational interactions, or mutation effects on different genetic backgrounds in the virus. Moreover, interpretation rules may initially provide limited resistance assessment to new molecules inhibiting the same target/s until new knowledge is generated.

In an attempt to build a universal computational tool for resistance prediction that would not rely on prior knowledge on drug resistance, we have developed an automatic protocol combining sequence alignment with protein-ligand induced fit sampling techniques. In this first proof-of-concept study we focused on the HIV-1 PR, but the process could be later extended to other enzymes, HIV-1 subtypes or viruses.

HIV-1 protease (PR) is a 22 kDa dimeric aspartic protease that contributes to the maturation progress of the virus, cleaving the polyprotein precursor into functional viral proteins. Being one of the main targets in the treatment against HIV, the development of nine FDA-approved protease inhibitors, along with numerous crystal structures were established in the last two decades.[22, 23] The active form of HIV-1 protease is a homodimer, with ~99 residues per chain. The catalytic residues Asp25 and Asp125 lie at the bottom of the binding cavity shielded by a pair of flaps (residues 44-55 and 144-155).[24] Main features common to all inhibitors are the existence of a hydroxyl group in contact with Asp25/125 residues, and a water molecule that mediates contact between a conserved carbonyl of the inhibitors and the protease amide groups of Ile50/Ile150 located in the flaps (Figure 1). Several resistant mutations have been described along the whole protease chain. Many mutations might occur simultaneously within a single variant highlighting the importance of complex cooperative effects. Typically these changes confer resistance not only by direct contact with the inhibitor, but also by subtle changes in the structure/dynamics incidentally affecting the active site. In addition to reducing or blocking drug inhibition, these mutations may also affect viral fitness by affecting the normal function and lead to the emergence of addition compensatory mutations that restore viral fitness.[4, 25-28]
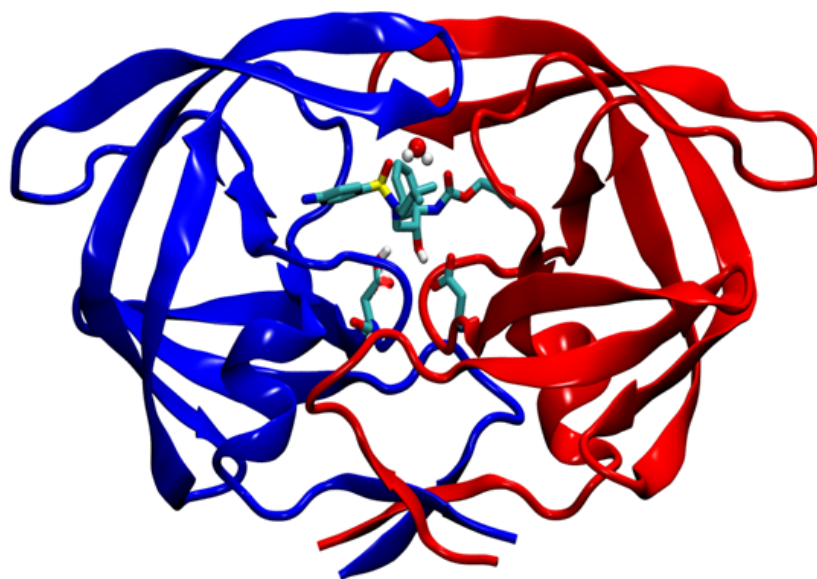

Figure 1. The HIV-1 protease dimer. Cartoon diagram of NL4-3 protease sequence showing monomers in blue and red. Residues Asp25, Asp125 and APV are shown as sticks colored by atom type. A relevant water molecule is displayed as ball and stick.

In this study, we used most of existing HIV-1 PR crystal structures available in the protein data bank and PELE (Protein Energy Landscape Exploration, a technique capable of reproducing complex protein-ligand induced fit[29, 30]), to explore the impact of mutations in

inhibitor binding. Our protocol was first tested by comparing calculated and experimental binding energies of several FDA-approved inhibitors with engineered PR variants containing different amino acid mutations.[15, 31] Then, we used our approach to predict *ab initio*, without any informational-driven bias, the resistance profile of 44 clinical HIV-1 PR variants to amprenavir (APV) and darunavir (DRV). Our technique is fast and, contrary to previous structure-based computational methods that were limited to variants with a reduced number of mutations,[15, 16] it can predict PI resistance to variants containing more than 10 mutations.

METHODS

Systems used

References structures

To build the homology models of the wild type NL4-3 and HXB2 reference sequences (which were considered as non-resistant genotypes unaltered by the selective pressure of any PI), we used the crystal structure of HIV-1 protease bound to BE6 inhibitor (pdb code 1W5Y[32]), with only 2 and 1 mutations per chain, respectively.

Several crystal structures with measured catalytic activities were also used to tune our computational protocol. For the APV ligand, the PDB entries used were: 3NU3[33] a WT reference with analogous relative affinity as NL4-3, 3S43[34] a triple mutant with 15 fold increase in inhibition constant (K_i, involving a decrease in affinity), 3NU5[33] a single mutant with 30 increase in inhibition constant. For the DRV ligand, the PDB entries used were, 2IEN[24] a WT reference with analogous relative affinity as NL4-3, and 3EM6[35] a double mutant with 4 fold higher dissociation constant.

Modeling HIV-1 protease mutants

Due to the strong correlation between sequence similarity and structure, our strategy to model mutants with no solved structure was to use the closest existing crystal structure in terms of sequence. We found ~450 crystal structures of HIV-1 protease which can be used as templates for building a given clinical sequence. Thus, modeling mutants was performed in two steps: (1) a search for the crystal structure with the highest similarity to the mutant sequence using BLAST,[36] and (2) building our model by replacing each mutant using the Maestro[37] software. Mutations were introduced simultaneously in the two protease chains. Using this procedure several sets of mutants were built: *The first set* was selected from the work by Koh et al.,[31] were three NL4-3 variants: (L10I, G48V, I54V, L90M), (L10F, V32I, M46I, I54M, A71V, I84V) and (L10F, D30N, K45I, A71V, T74S) were tested against four inhibitors DRV, APV, indinavir (IDV) and ritonavir (RTV). *The second set* was obtained from the study by Shenderovich et al.[15] For this case, we selected the more difficult set of sequences, with more than 10 mutations (and a maximum number of 17 mutations per

monomer, giving a total of 15 sequences) against IDV inhibitor. *The third set* included 21 and 23 PR FASTA sequences with different resistance profiles to APV and DRV, respectively, obtained from routine genotypic resistance testing (TRUGENE® HIV-1 Genotyping Assay, Siemens Healthcare, Barcelona, Spain) in the HIV Unit and irsiCaixa AIDS Research institute, Hospital Universitari Germans Trias i Pujol, Badalona, Spain. This last set of studies was performed as a blind test, where the modeling team had only access to the sequence but not to the resistance score.

Systems preparations

All systems were prepared with Schrodinger's Protein Wizard.[37] This algorithm builds hydrogen-bonded clusters and performs 100000 Monte Carlo moves by reorienting hydroxyl and thiol groups, water molecules, amide groups of Asn and Gln, and the imidazole ring of His, to correct for typical crystal structure refinement errors. The algorithm also predicts protonation states of His, Asp, Glu, Lys and Arg. Each possibility is scored based on the total number of hydrogen bonds and their quality (relative to an idealized hydrogen bond). In this work, Asp25 was protonated in all structures, whereas Asp125 was considered ionized. Histidines 69 and 169 were either epsilon or double protonated depending on the structural environment; all other histidines kept delta protonation. A water molecule, generally found in all protease-inhibitor crystal structures was kept. The water preserved mediates a contact between the P2/P1′ carbonyl oxygen atoms from the inhibitors and the amide groups of Ile50/Ile50′.

Once the atom model was built, the ligand was initially docked in the active site using Glide.[38] The top score model (XP scoring) obtained by glide was selected, followed by a molecular mechanics minimization using Schrodinger's Protein Wizard and the OPLS2005 force field to remove possible geometric clashes. In order to keep the system close to its initial geometry, a restrain was applied to all heavy atoms (allowing a maximum displacement of 0.3 Å). All ligands were quantum mechanically minimized separately in the gas phase using the DFT/B3LYP and 6-31G** basis set level of theory. Atomic charges were then derived from the electrostatic potential fitting at the same level of theory. Additional ligand parameters were taken from the OPLS2 force field.[39]

Assessing the induce fit and binding energy

To map the protein-ligand conformational space we used PELE[29], which implements a Monte Carlo algorithm where new trial configurations are produced with sequential ligand (and protein) perturbation, side chain prediction and minimization steps. Ligand perturbation includes a ligand specific rotamer library.[30] Trial configurations are then filtered with a Metropolis acceptance test, where the energy is described with an all-atom OPLS force field and a surface generalized born solvent model.[40] PELE has recently shown to provide competitive advantages with respect to *state of the art* induced fit commercial

software and to reproduce the conformational sampling obtained in microsecond molecular dynamics (MD) trajectories.[41, 42]

A total of 12 independent MC trajectories were produced for each inhibitor and mutant sequence. Trajectories were interrupted after 12h of CPU, providing approximately a total of 6000 Monte Carlo steps and ~2000 accepted minima. PELE's binding energy was then obtained by averaging the interaction energies of all accepted minima (approximately 2000 snapshots). The overall protocol has been added to our in-house server (pele.bsc.es) for public usage.

In addition to PELE, HIV-1 resistance to APV and DRV was modeled with MD, using the Amber 12 software.[43] For each system we performed 10 independent 10 ns runs with explicit waters, a 10 Å layer of pre-equilibrated water molecules in a truncated octahedral box, and physiological salt (NaCl) conditions. A time step of 1 fs was used throughout the simulations in combination with the SHAKE algorithm to constrain bond lengths involving hydrogen atoms.[44, 45] Non-bonded interactions were explicitly evaluated for distances below 9 Å. The particle mesh Ewald method was employed to treat long-range electrostatic interactions.[46] Constant pressure and temperature (NPT ensemble) were maintained by weakly coupling the system to an external bath at 1 bar and 298 K, using the Berendsen barostat and thermostat.[47] Finally, MMPBSA.py[48] was used to extract the average binding energy, where we tried several time ranges: 1, 5 and 10 ns; best results were obtained using 20 snapshots between 3 and 5 ns.

RESULTS

In the first part of the study, we validated the method by comparing in vitro IC_{50} results for specific mutations with our theoretical predictions. In the second part, attention was centered on predicting how mutations in HIV-1 protease taken from real patients affect drug resistance against APV and DRV.

Validation for the set of data with known experimental binding affinity (IC_{50})

We first calculated the interaction energies of APV, DRV, IDV and RTV for three different sequences (each ligand) among the set described in Koh et al.[31] Besides NL4-3, for each drug we selected the mutation with the lowest and highest IC_{50} value relative to the value for NL4-3 (see Table 3 in Koh et al.[31]). Figure 2, summarizes PELE's relative increase in binding energy for each ligand and target, compared to the reference NL4-3 (the higher the increase in binding energy, the greater the resistance conferred by the mutation). The experimental relative (with respect to NL4-3) increase in IC_{50} is also shown with numbers in Figure 2. Clearly, PELE's simulations can distinguish between the high (dark gray) and low (light gray) resistance mutant in each specific inhibitor. Moreover, the relative increase

in PELE's binding energy correlates nicely with the experimental increase in resistance in this set of data. This is more clearly seen in the right panel where the natural logarithm of the IC50 values is used to estimate changes in binding affinities.
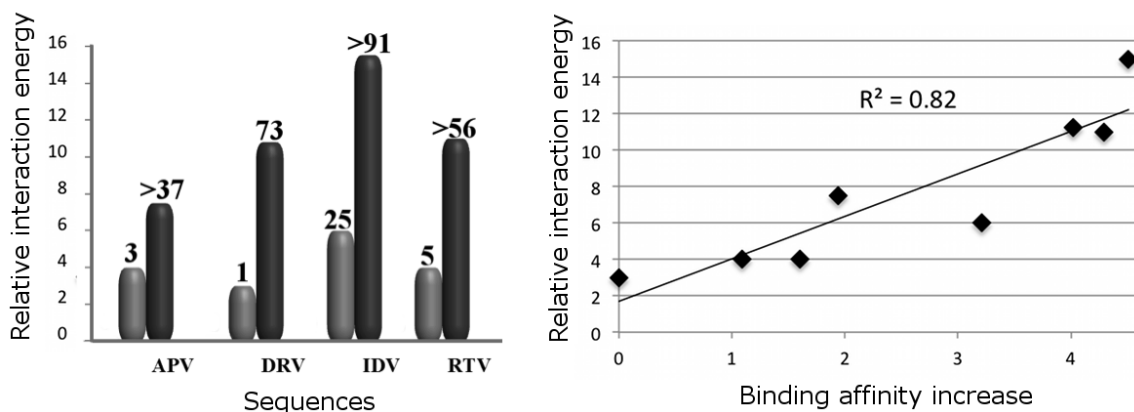

Figure 2. Left panel: PELE's relative binding energies (in kcal/mol) predicted for the high (dark gray) and low (light gray) resistance HIV-1 PR mutants for APV, DRV, IDV and RTV observed by Koh et al.[31] (exact sequences provided in the Methods section). Experimental relative (with respect to NL4-3) increase in IC_{50} is shown with numbers above each bar. Right Panel: Correlation between theoretical and experimental relative binding affinity increase (with respect to NL4-3) along with their coefficient of determination. Values larger than (>) were used as the absolute experimental measure (if we neglect these three values R^2 increases to 0.87)

For a second test, we used a subset of sequences derived from the work of Shenderovich et al.[15] Using known and in-house prepared mutations, these authors developed possibly the most comprehensive computational predictor to date. However, as noticed by the same authors, the quality of predictions correlates negatively with the increase of number of mutations. From their test set, we selected all the sequences with more than 10 mutations for the IDV inhibitor (a total of 15 sequences, which represent a specially difficult set for prediction[15]). Figure 3 shows the correlation between our relative interaction energy estimates, using again NL4-3 as the reference zero value, and the experimental relative binding energy. Additionally we include the estimates derived from Shenderovich et al.[15], also referred to NL4-3. We should notice that in Shenderovich et al., all mutant models were derived from one crystal and that they used simpler side chain (and backbone) sampling algorithms. Despite the difficulty of this set of 15 compounds, the method outlined here shows a coefficient of determination equal to 0.75 (PELE's p-value= $2.5e^{-05}$ and ΔE's p-value = 0.004), improving the previous result significantly.
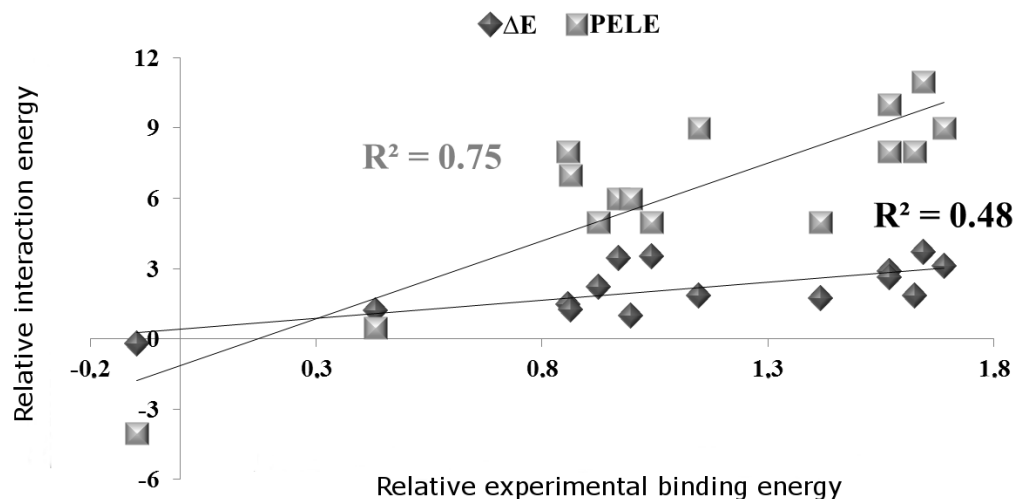
Figure 3. Correlation between modeled changes in relative binding energies (in kcal/mol) obtained in Shenderovich et al.[15] (ΔE, square beads) and PELE (rhomboid beads) with changes from experimental relative binding energy.

Prediction results for APV and DRV

After testing our computational protocol with publicly available data, we proceeded to perform "blind predictions" of the resistance in clinical samples. Importantly, each of these variants, taken from HIV infected patients, contained a large number of mutations in each monomer (10-25) when compared to the reference NL4-3. Therefore, in some of the simulated systems the protease bore as many as 50 mutations, representing an incredible difficult test (see Figure S1 in <u>Supporting Information</u> (SI) for the alignment of each sequence to the subtypeB consensus protease sequence). The data for the clinical isolates was divided (and ordered) based on the resistance scores calculated from expert assessments in HIVdb[49]. Samples were categorized as sensitive (S) when HIVdb resistance scores were below 20, intermediate (I) resistance when scores were between 20 and 60, and resistant (R) when scores were 60 or higher.

Figure 4 shows the results for APV where again we computed PELE's relative binding energies to the reference compound NL4-3. For this compound we had initially run a benchmark with four variants containing known resistance-related mutations: two consensus reference proteins (the sequence from pdb structure 3NU3 and the reference sequence HXB2, which were both considered non-resistant), the single mutation I50V with 30 fold increase in K_i (pdb id 3NU5) and the triple mutant V32I/I47V/V82I with 15 fold increase in K_i (pdb id 3S43). The data from this benchmark has also been included in Figure 4 (left-hand side). As can be seen, the two additional non-resistant reference sequences (the one from 3NU3 and HXB2) show equal or (slightly) better interaction energies to that found for NL4-3. This result shows how our computational protocol

predicts reference sensitive sequences other than NL4-3 as sensitive, and similarly, the two APV-resistant reference sequences (3S43 and 3NU5), with a 15 and 30 fold increase in K_i, respectively, were correctly predicted as interfering drug-binding (note that the impact of the mutation in predicted binding, matches the relative K_i increase found experimentally).

For the sequences extracted from the patients (1-21), it can be seen that all sensitive sequences (green, columns 1-7) are predicted to be sensitive by our method, and all highly resistant sequences are predicted to be highly resistant (red, columns 14-21). Sequences classified as intermediate resistance (in yellow) offer a less clear differential profile, but there is only one pair (13/14) where PELE will have some difficulty to predict the resistance level produced by the mutation.
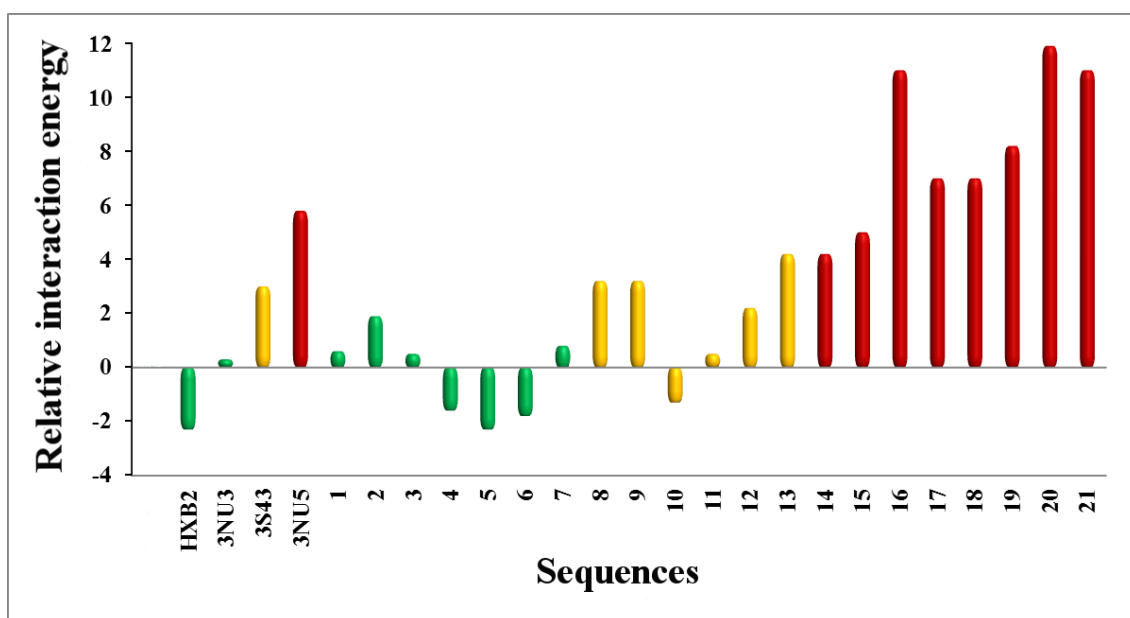

Figure 4. PELE's relative change in APV binding energies (kcal/mol). Sensitive (<20), intermediate (20>= =<60) and resistance (>60) HIVdb values for each sequence are shown in green, yellow and red colors, respectively.

Due to the large number of mutations in each sequence, assigning individual effects is not a trivial task. Figure 5 compares the protein-APV interactions for sequences number 1 and 20, with a ~12 kcal difference in relative affinity (HIVdb scores 0 and 150, respectively). Sequence 1 presents 6 mutations (per chain) compared to NL4-3 while sequence 20 has 16 mutations. Clearly, sequence 20 shows a reduction in the number of interactions. Mutations on residues Val82 and Val84 affect the binding mode and orientation of the inhibitor. In particular, these mutations influence the proper interaction of APV with polar residues Asp25, Asp29 and Asp30, significantly decreasing the affinity of the ligand. Asp25, for

example, plays an important role in ligand recognition by making a strong interaction with the hydroxyl group of almost all inhibitors.
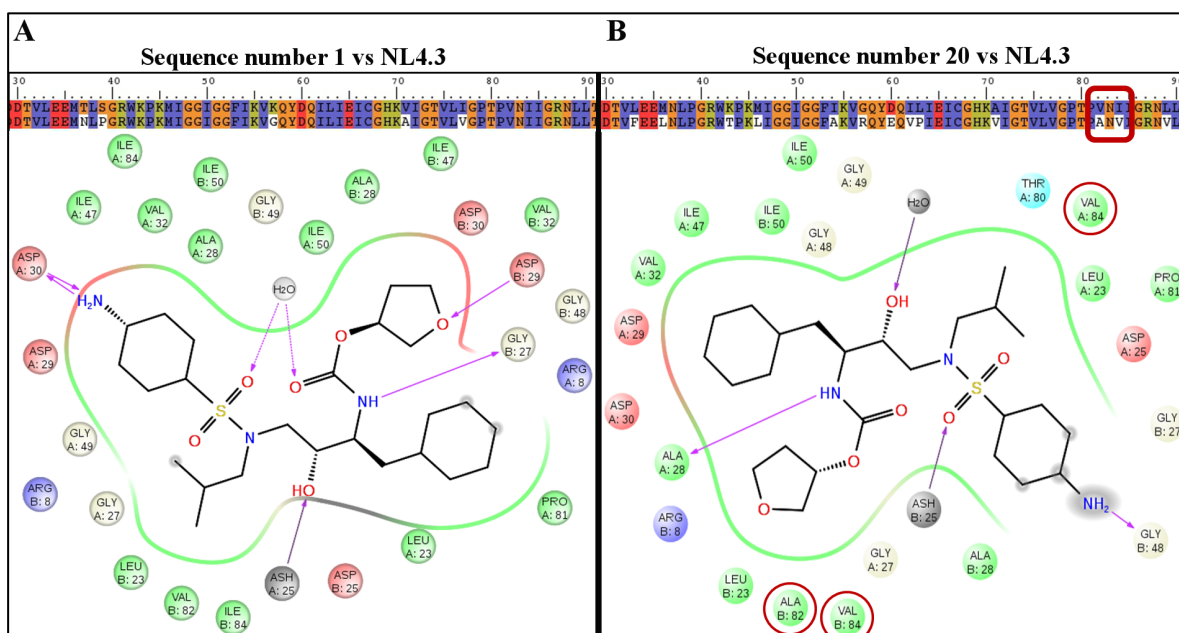

Figure 5. Protein-APV interaction diagram for sequences number 1, panel A, and 20, panel B. Important residues, discussed in the main text, are underlined in red circles. Partial sequence alignment to NL4-3 is also shown for both systems.

Figure 6 shows the equivalent results for DRV. For this compound we also modeled three additional variants containing known resistance-related mutations: two consensus reference proteins (the one taken from pdb structure 2IEN and the reference sequence HXB2) and the double mutant I50L/A71V (3M60), which has 4 fold higher dissociation constant. As in the APV case, reference and sensitive compounds show little or no effect when compared to NL4-3. Again, the most remarkable result is the correlation between the predicted relative binding energy and the estimation of resistance, where all highly resistant variants are clearly identified.
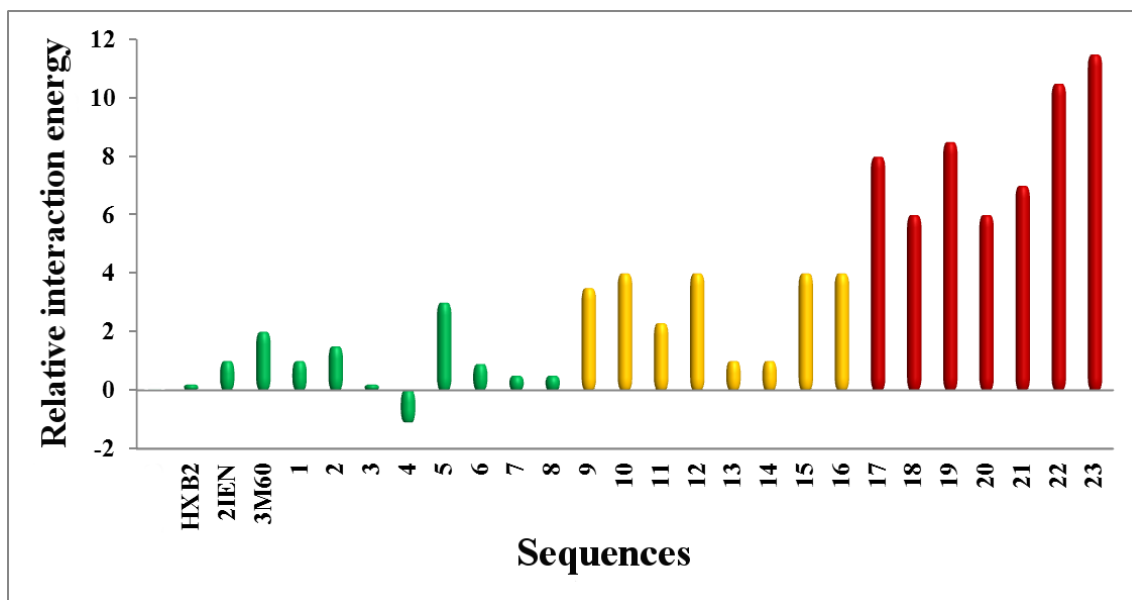
Figure 6. PELE's relative change in DRV binding energies (kcal/mol).). Sensitive (<20), intermediate (20>= =<60) and resistance (>60) HIVdb values for each sequence is shown in green, yellow and red colors, respectively .

Finally, MD/MMPBSA simulations were performed for all patients' sequences in both ligands. While results (Figure S2 in SI) still indicate an overall tendency between the sensitive, intermediate and resistance regimes, wrong predictions are clearly observed in both systems (more in DRV) and differences do not allow now to discriminate high resistance mutants.

DISCUSSION

The primary aim of this study was to develop an automatic computational protocol for rapid discrimination between resistant and sensitive HIV-1 protease variants. To this aim we have used PELE, an atomic-resolution sampling algorithm combining a stochastic Monte Carlo procedure with protein structure prediction techniques, which is specially suited for induced-fit docking problems.[41, 42]

Since most of the mutant structures have not been crystallized, one of the main questions is how to generate an all atom model for each sequence. Since we were comparing the binding energies to the NL4-3 reference sequence, one simple strategy would have been to use its structure as a template (in Modeller,[50, 51] I-TASSER,[52] etc.). However, most sequences have up to 20 mutations (when compared to NL4-3) in each chain, which could introduce large errors in building the models; the large number of HIV-1 X-ray crystal structures enabled better strategies. An obvious one is the use of multiple templates in homology modeling, but our initial analysis indicated that it is better to start from a well-

defined single template than to combine different crystal structures. Thus, we decided to create a subset of several crystal structures to which our sequences were compared and to use the crystal structure with the best sequence alignment to our target (to minimize the number of mutations modelled) as template. By doing so, we reduced significantly the number of mutations for model building (see Table S1 in SI for a list of the reference crystal structures).

Our initial two tests compared the predicted relative binding energy with systems having experimental affinities. As seen in Figures 2 and 3, PELE is capable of distinguishing between resistant and sensitive sequences. Moreover, it shows a good correlation to experimental affinities even for sequences with a large number of mutations, improving existing theoretical prediction tools. Next, we tested our protocol in a blind test with patient sequences for two well-known HIV-1 inhibitors, APV and DRV. The results showed again noteworthy correlations between calculated changes in binding energy and the index obtained by HIVdb scores. In all cases we could distinguish the sequences with strong resistance, having predicted relative binding energies >5kcal/mol. This value is also consistently being observed in the initial two tests. The method showed problems to discriminate between mutants with moderate resistance, but this is the segment where rule-based methods (used here as reference) are expected to show the poorest performance. MD/MMPBSA predictions, on the other hand, have difficulties in predicting high resistance mutants. This behavior, however, is expected since this method works efficiently when addressing small changes (involving mostly single and double mutants).[16, 53]

The majority of resistance-related mutations are conservative substitutions in positions Leu10, Val32, Ile54, Val82, Ile84 and Leu90. Our simulations indicated that these mutations do not introduce large changes in the structure of the complex, but modify subtle van der Waals and hydrogen bond interactions between the ligand and active site amino acids (see Figure 5). Such resistance mechanisms are observed for other systems, [54, 55] and agree with the fact that none of these mutations significantly disrupts enzyme activity. However, even small, the subtle conformational changes induced by mutations are important in defining accurate binding, making it necessary the use of specific homology models and induce fit relaxation. The later point is clearly shown by the fact that PELE's scores after homology modeling (or after only a short exploration of sampling) did not correlate with susceptibility data (Table 1). The quality (and necessity) of the induced fit conformational sampling is also evident when predicting the interaction with a general scoring function. For this, we used Glide XP from Schrodinger.[38] Figure 7 shows relative Glide scores before and after the induced fit for the 21 clinical sequences in APV (plus NL4-3 and HXB2). Each score is the average of 50 structures obtained after clustering PELE's trajectories. Clearly the correlation with HIVdb scores significantly improves after conformational sampling with PELE. While the final Glide scores still have low correlation, 0.33, they can identify most of the high resistance sequences. Remarkably, the

correlation increases to 0.80 when using PELE's all-atom binding energy (as score), revealing the importance of an explicit treatment of all-atom interactions.[30] Notice also that the clustering procedure only reduces the overall PELE's correlation from 0.82 to 0.80.

Table 1. Evolution of the coefficient of determination between PELE's interaction energies and HIVdb values along the conformational sampling for the 21 APV patient sequences.

	First-score	Step-50	Step-100	Step-200	Step-300	Total
R^2	0.43	0.64	0.72	0.75	0.81	0.82
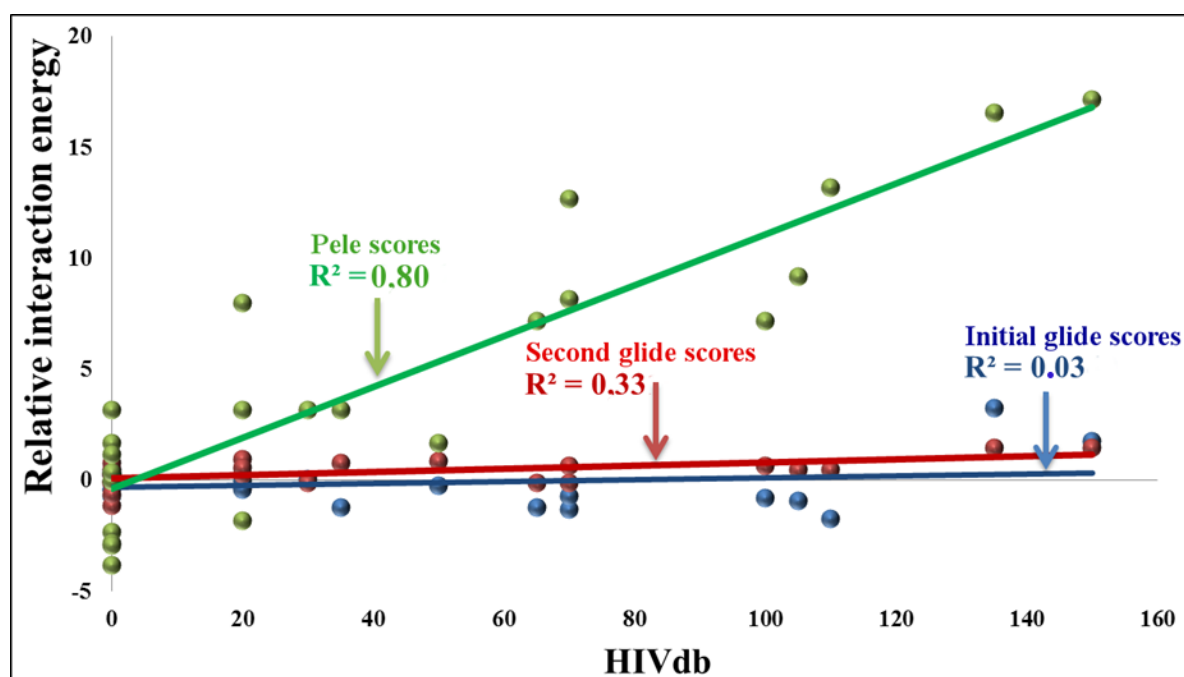
Figure 7. Comparison between PELE and Glide (initial and final) correlation scores, against HIVdb values, for the 21 APV patient sequences (using 50 representative structures after clustering).

Conclusions

Overall, this study shows how computational techniques are capable of quantitatively discriminating resistance variants of HIV-1 protease. Our protocol, combining sequence alignment to current pdb structures and state of the art protein-ligand induced fit sampling algorithms, shows great promise as an automatic tool for a quick prediction. The entire protocol can be run in less than 24 hours on a small commodity workstation, and is based on biophysical first principles. Moreover, it is capable to trace the effect of novel mutations

and the binding of new drugs. In summary, modeling drug-target interactions holds the potential to provide less biased and more accurate assessments of antiretroviral drug resistance, which could improve clinical management of HIV-infected subjects.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: XXX

Figure S1 showing the alignment for the patients sequences with the HIV-1 subtypeB consensus protease sequence. Table S1 showing the crystal structure used for modeling each patient sequence. Figure S2 showing the MD/MMPBSA results.

AUTHOR INFORMATIONS

Corresponding Authors

Victor Guallar, Phone, +34 (93) 4137727; e-mail: victor.guallar@bsc.es;

Roger Paredes, Phone, + 34 (93) 465 6374 (Ext: 162); e-mail: rparedes@irsicaixa.es

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

References

1.   Rodger, A. J.; Lodwick, R.; Schechter, M.; Deeks, S.; Amin, J.; Gilson, R.; Paredes, R.; Bakowska, E.; Engsig, F. N.; Phillips, A.; Grp, E. S., Mortality in well controlled HIV in the continuous antiretroviral therapy arms of the SMART and ESPRIT trials compared with the general population. *Aids* **2013**, 27, 973-979.

2.   Hogg, R. S.; Heath, K. V.; Yip, B.; Craib, K. J. P.; O'Shaughnessy, M. V.; Schechter, M. T.; Montaner, J. S. G., Improved survival among HIV-infected individuals following initiation of antiretroviral therapy. JAMA-*J. am. Med Assoc.* **1998**, 279, 450-454.

3.   Palella, F. J.; Delaney, K. M.; Moorman, A. C.; Loveless, M. O.; Fuhrer, J.; Satten, G. A.; Aschman, D. J.; Holmberg, S. D., Declining Morbidity and Mortality among Patients with Advanced Human Immunodeficiency Virus Infection. *New Engl. J. Med.* **1998**, 338, 853-860.

4.   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.; Emini, E. A., IN-VIVO Emergence of HIV-1 Variants Resistant to Multiple Protease Inhibitors. *Nature* **1995**, 374, 569-571.

5.   Cohen, M. L., Epidemiology of Drug-Resistance - Implications for a Postantimicrobial ERA. *Science* **1992**, 257, 1050-1055.

6.   Larder, B. A.; Kemp, S. D., Multiple Mutations in HIV-1 Reverse-Transcriptase Confer High-Level Resistance to Zidovudine (AZT). *Science* **1989**, 246, 1155-1158.

7.   Rhee, S. Y.; Fessel, W. J.; Zolopa, A. R.; Hurley, L.; Liu, T.; Taylor, J.; Nguyen, D. P.; Slome, S.; Klein, D.; Horberg, M.; Flamm, J.; Follansbee, S.; Schapiro, J. M.; Shafer, R. W., HIV-1 protease and reverse-transcriptase mutations: Correlations with antiretroviral therapy in subtype B isolates and implications for drug-resistance surveillance. *J. Infect. Dis.* **2005**, 192, 456-465.

8.   Gilks, C. F.; Crowley, S.; Ekpini, R.; Gove, S.; Perriens, J.; Souteyrand, Y.; Sutherland, D.; Vitoria, M.; Guerma, T.; De Cock, K., The WHO public-health approach to antiretroviral treatment against HIV in resource-limited settings. *Lancet* **2006**, 368, 505-510.

9.   Keiser, O.; Orrell, C.; Egger, M.; Wood, R.; Brinkhof, M. W. G.; Furrer, H.; van Cutsem, G.; Ledergerber, B.; Boulle, A.; Shcs; Int Epidemiol Databases, E., Public-health and individual approaches to antiretroviral therapy: Township South Africa and Switzerland compared. *Plos Medicine* **2008**, 5, 1102-1111.

10.   Hamers, R. L.; Wallis, C. L.; Kityo, C.; Siwale, M.; Mandaliya, K.; Conradie, F.; Botes, M. E.; Wellington, M.; Osibogun, A.; Sigaloff, K. C. E.; Nankya, I.; Schuurman, R.; Wit, F. W.; Stevens, W. S.; van Vugt, M.; de Wit, T. F. R.; Paser, HIV-1 drug resistance in antiretroviral-naive individuals in sub-Saharan Africa after rollout of antiretroviral therapy: a multicentre observational study. *Lancet Infect. Dis.* **2011**, 11, 750-759.

11.   Hamers, R. L.; Sigaloff, K. C. E.; Wensing, A. M.; Wallis, C. L.; Kityo, C.; Siwale, M.; Mandaliya, K.; Ive, P.; Botes, M. E.; Wellington, M.; Osibogun, A.; Stevens, W. S.; de Wit, T. F. R.; Schuurman, R.; PharmAccess African Studies, E., Patterns of HIV-1 Drug Resistance After First-Line Antiretroviral Therapy (ART) Failure in 6 Sub-Saharan African Countries: Implications for Second-Line ART Strategies. *Clin. Infect. Dis.* **2012**, 54, 1660-1669.

12.   Organization, W. H., *WHO HIV Drug Resistance Report 2012*. World Health Organization: 2012.

13.   Johnson, V. A.; Calvez, V.; Gunthard, H.; Paredes, R.; Pillay, D.; Shafer, R. W.; Wensing, A. M.; Richman, D. D., Update of the drug resistance mutations in HIV-1: March 2013. *Top Antivir. Med.* **2013**, 21, 6-14.

14.   Killian, B. J.; Kravitz, J. Y.; Somani, S.; Dasgupta, P.; Pang, Y.-P.; Gilson, M. K., Configurational Entropy in Protein–Peptide Binding:: Computational Study of Tsg101 Ubiquitin E2 Variant Domain with an HIV-Derived PTAP Nonapeptide. *J. Mol. Biol.* **2009**, 389, 315-335.

15.   Shenderovich, M. D.; Kagan, R. M.; Heseltine, P. N. R.; Ramnarayan, K., Structure-based phenotyping predicts HIV-1 protease inhibitor resistance. *Prot Sci* **2003**, 12, 1706-1718.

16.   Hao, G.-F.; Yang, G.-F.; Zhan, C.-G., Computational Mutation Scanning and Drug Resistance Mechanisms of HIV-1 Protease Inhibitors. *J. Phys. Chem. B* **2010**, 114, 9663-9676.

17.   Lapins, M.; Wikberg, J. E. S., Proteochemometric Modeling of Drug Resistance over the Mutational Space for Multiple HIV Protease Variants and Multiple Protease Inhibitors. *J. Chem. Inf. Model.* **2009**, 49, 1202-1210.

18.   Saigo, H.; Uno, T.; Tsuda, K., Mining complex genotypic features for predicting HIV-1 drug resistance. *Bioinformatics* **2007**, 23, 2455-2462.

19.   Yu, X.; Weber, I. T.; Harrison, R. W., Prediction of HIV drug resistance from genotype with encoded three-dimensional protein structure. *BMC Genomics* **2014**, 15, S1-S1.

20.   Masso, M.; Vaisman, I. I., Sequence and structure based models of HIV-1 protease and reverse transcriptase drug resistance. *BMC Genomics* **2013**, 14, S3-S3.

21.   Hou, T.; Zhang, W.; Wang, J.; Wang, W., Predicting drug resistance of the HIV-1 protease using molecular interaction energy components. *Proteins* **2009**, 74, 837-846.

22.   Kim, E. E.; Baker, C. T.; Dwyer, M. D.; Murcko, M. A.; Rao, B. G.; Tung, R. D.; Navia, M. A., Crystal-structure of HIV-1 Protease in Complec with VX-478, a Potent and Orally Bioavailable Inhibitor of the Enzyme. *J. Am. Chem. Soc.* **1995**, 117, 1181-1182.

23.   Turner, S. R.; Strohbach, J. W.; Tommasi, R. A.; Aristoff, P. A.; Johnson, P. D.; Skulnick, H. I.; Dolak, L. A.; Seest, E. P.; Tomich, P. K.; Bohanan, M. J.; Horng, M. M.; Lynn, J. C.; Chong, K. T.; Hinshaw, R. R.; Watenpaugh, K. D.; Janakiraman, M. N.; Thaisrivongs, S., Tipranavir (PNU-140690): A potent, orally bioavailable nonpeptidic HIV protease inhibitor of the 5,6-dihydro-4-hydroxy-2-pyrone sulfonamide class. *J. Med. Chem.* **1998**, 41, 3467-3476.

24.   Tie, Y. F.; Boross, P. I.; Wang, Y. F.; Gaddis, L.; Hussain, A. K.; Leshchenko, S.; Ghoshl, A. K.; Louis, J. M.; Harrison, R. W.; Weber, I. T., High resolution crystal structures of HIV-1 protease with a potent non-peptide inhibitor (UIC-94017) active against multi-drug-resistant clinical strains. *J. Mol. Biol.* **2004**, 338, 341-352.

25.   Tantillo, C.; Ding, J. P.; Jacobomolina, A.; Nanni, R. G.; Boyer, P. L.; Hughes, S. H.; Pauwels, R.; Andries, K.; Janssen, P. A. J.; Arnold, E., Locations of ANTI-AIDS Drug-Binding Sites and Resistance Mutations in the 3-Dimensional Structure of HIV-1 Reverse-Transcriptase - Implications for Mechanisms of Drug-Inhibition and Resistance. *J. Mol. Biol.* **1994**, 243, 369-387.

26.   Gubareva, L. V.; Bethell, R.; Hart, G. J.; Murti, K. G.; Penn, C. R.; Webster, R. G., Characterization of mutants of influenza A virus selected with the neuraminidase inhibitor 4-guanidino-Neu5Ac2en. *J. Virol.* **1996**, 70, 1818-1827.

27.   Nijhuis, M.; Wensing, A. M. J.; Bierman, W. F. W.; de Jong, D.; Kagan, R.; Fun, A.; Jaspers, C. A. J. J.; Schurink, K. A. M.; van Agtmael, M. A.; Boucher, C. A. B., Failure of Treatment with First-Line Lopinavir Boosted with Ritonavir Can Be Explained by Novel Resistance Pathways with Protease Mutation 76V. *J. Infec. Dis.* **2009**, 200, 698-709.

28.   Friend, J.; Parkin, N.; Liegler, T.; Martin, J. N.; Deeks, S. G., Isolated lopinavir resistance after virological rebound of a ritonavir/lopinavir-based regimen. *AIDS* **2004**, 18, 1965-1966.

29.   Borrelli, K. W.; Vitalis, A.; Alcantara, R.; Guallar, V., PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *J. Chem. Theo. Comput.* **2005**, 1, 1304–1311.

30.   Borrelli, K. W.; Cossins, B.; Guallar, V., Exploring hierarchical refinement techniques for induced fit docking with protein and ligand flexibility. *J. Comp. Chem.* **2010**, 31, 1224-1235.

31.   Koh, Y.; Nakata, H.; Maeda, K.; Ogata, H.; Bilcer, G.; Devasamudram, T.; Kincaid, J. F.; Boross, P.; Wang, Y. F.; Ties, Y. F.; Volarath, P.; Gaddis, L.; Harrison, R. W.; Weber, I. T.; Ghosh, A. K.; Mitsuya, H., Novel bis-tetrahydrofuranylurethane-containin nonpeptidic protease inhibitor (PI) UIC-94017 (TMC114) with potent activity against multi-PI-resistant human immunodeficiency virus in vitro. *Antimicrob. Agents Ch.* **2003**, 47, 3123-3129.

32.   Lindberg, J.; Pyring, D.; Lowgren, S.; Rosenquist, A.; Zuccarello, G.; Kvarnstrom, I.; Zhang, H.; Vrang, L.; Classon, B.; Hallberg, A.; Samuelsson, B.; Unge, T., Symmetric fluoro-substituted diol-based HIV protease inhibitors - Ortho-fluorinated and meta-fluorinated P1/P1 '-benzyloxy side groups significantly improve the antiviral activity and preserve binding efficacy. *Eur. J. Biochem.* **2004**, 271, 4594-4602.

33.   Shen, C.-H.; Wang, Y.-F.; Kovalevsky, A. Y.; Harrison, R. W.; Weber, I. T., Amprenavir complexes with HIV-1 protease and its drug-resistant mutants altering hydrophobic clusters. *FEBS J.* **2010**, 277, 3699-3714.

34.   Tie, Y.; Wang, Y.-F.; Boross, P. I.; Chiu, T.-Y.; Ghosh, A. K.; Tozser, J.; Louis, J. M.; Harrison, R. W.; Weber, I. T., Critical differences in HIV-1 and HIV-2 protease specificity for clinical inhibitors. *Protein Sci.* **2012**, 21, 339-350.

35.   Mittal, S.; Bandaranayake, R. M.; King, N. M.; Prabu-Jeyabalan, M.; Nalam, M. N. L.; Nalivaika, E. A.; Yilmaz, N. K.; Schiffer, C. A., Structural and Thermodynamic Basis of Amprenavir/Darunavir and Atazanavir Resistance in HIV-1 Protease with Mutations at Residue 50. *J. Virol.* **2013**, 87, 4176-4184.

36.   Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic local alignment search tool. *J. Mol. Biol.* **1990**, 215, 403-410.

37.   Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W., Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided Mol. Des.* **2013**, 27, 221-234.

38.   Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, 47, 1750-1759.

39.   Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W., Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J. Chem. Theo. Comp.* **2010**, 6, 1509-1519.

40.   Yu, Z. Y.; Jacobson, M. P.; Josovitz, J.; Rapp, C. S.; Friesner, R. A., First-shell solvation of ion pairs: Correction of systematic errors in implicit solvent models. *J. Phys. Chem. B* **2004**, 108, 6643-6654.

41.   Espona-Fiedler, M.; Soto-Cerrato, V.; Hosseini, A.; Lizcano, J. M.; Guallar, V.; Quesada, R.; Gao, T.; Perez-Tomas, R., Identification of dual mTORC1 and mTORC2 inhibitors in melanoma cells: prodigiosin vs. obatoclax. *Biochem. Pharmacol.* **2012**, 83, 489-96.

42.   Hosseini, A.; Espona-Fiedler, M.; Soto-Cerrato, V.; Quesada, R.; Pérez-Tomás, R.; Guallar, V., Molecular interactions of prodiginines with the BH3 domain of anti-apoptotic Bcl-2 family members. *PloS one* **2013**, 8.

43.   Case, D.; Darden, T.; Cheatham III, T. E.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K., AMBER 12. *University of California, San Francisco* **2012**, 1.

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

45.   Miyamoto, S.; Kollman, P. A., Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comp. Chem.* **1992**, 13, 952-962.

46.   Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, 98, 10089-10092.

47.   Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, 81, 3684-3690.

48.   Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theo. Comp.* **2012**, 8, 3314-3321.

49.   Liu, T. F.; Shafer, R. W., Web resources for HIV type 1 genotypic-resistance test interpretation. *CLIN INFECT DIS* **2006**, 42, 1608-1618.

50.   Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M.-y.; Pieper, U.; Sali, A. Comparative Protein Structure Modeling Using Modeller. In *Curr Protoc Bioinformatics*; John Wiley & Sons, Inc.: 2002.

51.   Fiser, A.; Do, R. K. G.; Šali, A., Modeling of loops in protein structures. *Protein Sci.* **2000**, 9, 1753-1773.

52.   Mitra, P.; Shultis, D.; Brender, J. R.; Czajka, J.; Marsh, D.; Gray, F.; Cierpicki, T.; Zhang, Y., An Evolution-Based Approach to De Novo Protein Design and Case Study on Mycobacterium tuberculosis. *PLoS Comput. Biol.* **2013**, 9, e1003298.

53.   Genheden, S.; Ryde, U., The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery* **2015**, 10, 449-461.

54.   Ali, A.; Bandaranayake, R. M.; Cai, Y.; King, N. M.; Kolli, M.; Mittal, S.; Murzycki, J. F.; Nalam, M. N. L.; Nalivaika, E. A.; Özen, A.; Prabu-Jeyabalan, M. M.;

Thayer, K.; Schiffer, C. A., Molecular Basis for Drug Resistance in HIV-1 Protease. *Viruses* **2010**, 2, 2509-2535.

55.   Skálová, T.; Dohnálek, J.; Dušková, J.; Petroková, H.; Hradílek, M.; Souček, M.; Konvalinka, J.; Hašek, J., HIV-1 Protease Mutations and Inhibitor Modifications Monitored on a Series of Complexes. Structural Basis for the Effect of the A71V Mutation on the Active Site†. *J. Med. Chem.* **2006**, 49, 5777-5784.

TOC
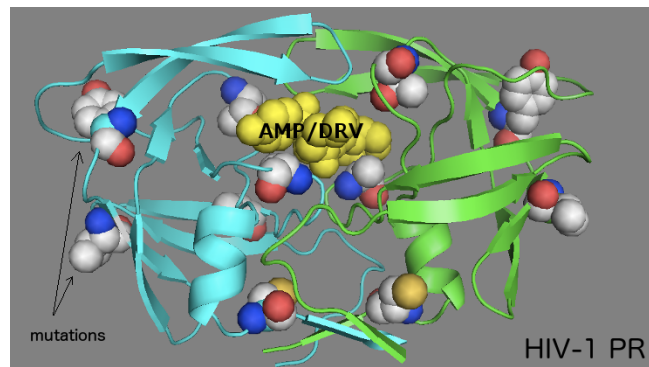
Supplementary Material

Computational Prediction of HIV-1 Resistance to Protease Inhibitors

Ali Hosseini[†], Andreu Alibés[†,‡], Marc Noguera-Julian[§,‖,⊥], Victor Gil[†], Roger Paredes[§,‖,⊥*], Robert Soliva[†], Modesto Orozco[†,‡,#] and Victor Guallar[†,O*]

[†] Joint BSC-CRG-IRB Research Program in Computational Biology. Barcelona Supercomputing Center. c/ Jordi Girona 29, 08034 Barcelona, Spain.

[‡] Institute for Research in Biomedicine (IRB Barcelona), Baldiri i Reixac 8, Barcelona 08028, Spain

[§] IrsiCaixa AIDS Research Institute, Hospital Universitari Germans Trias i Pujol, Badalona, Catalonia, Spain

[‖] Universitat Autònoma de Barcelona, Catalonia, Spain

[⊥] Universitat de Vic – Universitat Central de Catalunya, Vic, Catalonia, Spain

[#] Departament de Bioquímica. Universitat de Barcelona. Avgda Diagona 647. 08029, Spain

[O] Institució Catalana de Recerca i Estudis Avançats, Passeig Lluís Companys 23, 08010 Barcelona, Spain

```
ProtConsB PQITLWQRPL VTIKIGGQLK EALLDTGADD TVLEEMNLPG KWKPKMIGGI  50
   APV_1  ...--G---- ---------- ---------- ----D-D--- R---------
   APV_2  ...---E--- -P-R------ ---------- ------S--- R---------
   APV_3  ---------- ---------- ---------- ------T-S- R---------
   APV_4  ---------- --V-V----- ---------- ------SM-- R---------
   APV_5  ---------I ------E--- ---------- ------S--- R---------
   APV_6  ---------- ---------- ---------- ----D-S--- ----------
   APV_7  ---------- --V-----Q- ---------- ---------- R---------
   APV_8  ...------- -I--V----- ---------- ----D-E--- R---------
   APV_9  ---------- ---------- ---------- -I-------- R---------
  APV_10  ---------- ---------- ---------- ------D--- R----I----
  APV_11  -----E---- ---------- ---------- ------D--- R---------
  APV_12  ---------I ---------- ---------- ---------- ----------
  APV_13  ---------V -IV--A---- ---------- --F------- R-------M--
  APV_14  ---------V ---------- ---------- ----DI---- ----------
  APV_15  ...------I --V------R ---------- --F-DI---- R----L----
  APV_16  ...------I -P-------- ---------- --F--L---- R-T--L----
  APV_17  ---------V -IV--A---- ---------- --F------- R-------M--
  APV_18  ---P-----F --V------VR ---------- --I-DIE--- R-T--I----
  APV_19  ---------I ---------- ---------- ----DID--- ----------
  APV_20  ---------- ---------- ---------- -I-------- R----IV---
  APV_21  ---------F ---------R ---------- ----DID--- R---RI----
   DRV_1  ...------- -I--V----- ---------- ----D-E--- R---------
   DRV_2  ...--G---- ---------- ---------- ----D-D--- R---------
   DRV_3  ...---E--- -P-R------ ---------- ------S--- R---------
   DRV_4  ---------- ---------- ---------- ------T-S- R---------
   DRV_5  -----G---I ---------- --------N ----DIS--- R----I----
   DRV_6  ---------- --VR----I- ---------- -----I---- ----------
   DRV_7  ---------- --V-V----- ---------- ----D-S--- R---------
   DRV_8  ---------- -N--V---I- ---------- ----D----- R---------
   DRV_9  ...------I --V------R ---------- --F-DI---- R----L----
  DRV_10  ...------I -P-------- ---------- --F--L---- R-T--L----
  DRV_11  ---------- ---------- ---------- -I-------- R----IV---
  DRV_12  ---------V -IV--A---- ---------- --F------- R-------M--
  DRV_13  ---------I --V-V----T ---------- N-FQDIQ--- R-T--IV---
  DRV_14  ---------F --V------T ---------- --F-DLA--- R----L----
  DRV_15  ------Q-F --V------R ---------- -II-DIT-S- R----I----
  DRV_16  ---------I --V-V----R ---------- -IF-DIT--- --T--IV---
  DRV_17  ...------- --V-V----- ---------- -IFDDI---- --T--LV---
  DRV_18  ...------I -----A--V- ---------- -IFQ------ R-T--IV---
  DRV_19  ---------I -----A--V- ---------- -IFQ------ R-T--IV---
  DRV_20  ---------F L-V-V----- ---------- -IXXD-D--- R---------
  DRV_21  ---------V --V------VR ---------- -II-DI---- R-T--IV--V
  DRV_22  ---------- --V-V----- ---------- -IFDDI---- --T--LV---
  DRV_23  ---------I ---------IR ---------- --F-DID--- R----I----
```

```
ProtConsB GGFIKVRQYD QILIEICGHK AIGAVLVGPT PVNIIGRNLL TQIGCTLNF_  100
    APV_1  ---------- -VPL------ ---T------ ---------- --L-------
    APV_2  ---------- --TVD----- ---T--I--- ---------- ----------
    APV_3  ------K--- ---------- V--T--I--- ---------- --L-------
    APV_4  ------K--- -VP------- V--T------ ---------- --L-------
    APV_5  -------E-- --T------- -V-T--I--- ---------- ----------
    APV_6  ---------- --P------- ---T------ ---------- ----------
    APV_7  ---------- --P------- ---T------ ---------- ----------
    APV_8  ---V--K--- --P-----K- T--T------ -A-------- --L-------
    APV_9  ---------- --P------- ---T------ ---------- ----------
   APV_10  ---------- --P------- V-ST--I--- ---------M --L-------
   APV_11  ---------- --P------- TV-T--I--- ---------M ----------
   APV_12  ---------- --P------- ---T------ ---------M --L-------
   APV_13  ----S-----E -VP------- VLST------ -S-------- ----------
   APV_14  -------K--- -VSM----Y- VT-T--I--- ---------M --L-------
   APV_15  ----V----- --P------- VV-T-V---- ----V----- --L-------
   APV_16  ----A-----E -VP------- V--T------ -A-V----V- -K--------
   APV_17  ----S-----E -VP------- V--T------ -A-V------ ----------
   APV_18  ----V--K--E -VP------- IMST------ -A-------M --L-------
   APV_19  ---------- -VHV------ ---S------ ----V----- ----------
   APV_20  ---------- --P------- ---T------ ---------- ----------
   APV_21  ---------- -VP------- TTST------ -T-V------ --L-------
    DRV_1  ---V--K--- --P-----K- T--T------ -A-------- --L-------
    DRV_2  ---------- -VPL------ ---T------ ---------- --L-------
    DRV_3  ---------- --TVD----- ---T--I--- ---------- ----------
    DRV_4  -------K--- ---------- V--T--I--- ---------- --L-------
    DRV_5  ---------- --T-----Q- V--T--I--- ---------- --L-------
    DRV_6  ---------- --P------- ---T------ ---------- ----------
    DRV_7  ---------- -VP------- ---T------ ---------- ----------
    DRV_8  ---------- --P------- ---T------ ---------- ----------
    DRV_9  ----V----- --P------- VV-T-V---- ----V----- --L-------
   DRV_10  ----A-----E -VP------- V--T------ -A-V----V- -K--------
   DRV_11  ---------- --P------- ---T------ ---------- ----------
   DRV_12  ----S-----E -VP------- V--T------ -A-V------ ----------
   DRV_13  --L-R----- -VP------- TLAPI-I--- ---V----D-M --L-------
   DRV_14  ----V---E-- -VP------- V--P------ ---V-----M S---------
   DRV_15  ----V--KE-- -VPV------ V-ST------ ---V-----V- ----------
   DRV_16  --LVR-K--- -VP------- I-TT------ -T--V---SM ----------
   DRV_17  ----MR----- -VP------- VFNT------ -L-------M ----------
   DRV_18  ----MR----E --P------- VKCT------ -A-V-----M -E--------
   DRV_19  ----MR----E --P------- VKCT------ -A-V-----M -E--------
   DRV_20  ----M--KE-- --P------- I-ST--T--- ---V-----VM --L------|
   DRV_21  --LLR----- -VP------- V--P------ -M--V----M SR--------
   DRV_22  ----MR----- -VP------- VFNT------ -L-------M ----------
   DRV_23  ----A---E-- -VP------- VV-T------ -A-------M -KL-F-----
```

Figure S1. Sequence alignment with the HIV-1 subtypeB consensus protease sequence.

Amprenavir			Darunavir		
patient	crystal	identity (%)	patient	crystal	identity(%)
1	1W5Y	93	1	4NJS	92
2	1W5Y	93	2	1W5Y	93
3	1W5Y	94	3	1W5Y	93
4	1W5Y	92	4	1W5Y	94
5	1W5Y	94	5	1W5Y	89
6	1W5Y	97	6	1W5Y	93
7	1W5Y	96	7	1W5Y	95
8	4NJS	92	8	1W5Y	94
9	1W5Y	97	9	1SGU	89
10	2WL0	91	10	1RPI	90
11	1W5Y	93	11	1W5Y	95
12	1YTG	94	12	1IZI	89
13	1IZI	86	13	2FDD	76
14	3D3T	87	14	1SGU	87
15	1SGU	89	15	1SGU	84
16	1SGU	86	16	2FDD	82
17	1SGU	87	17	3TTP	88
18	2FDD	86	18	1SGU	83
19	1W5Y	91	19	1SGU	84
20	1W5Y	95	20	2B60	80
21	2WL0	88	21	2FDD	85
			22	3U7S	85
			23	2FDD	90

Table S1. Reference crystal structure used for each patients' sequence together with its per chain identity percentage. Structures deposited after 2014 were not considered at the time (beginning) of the study.
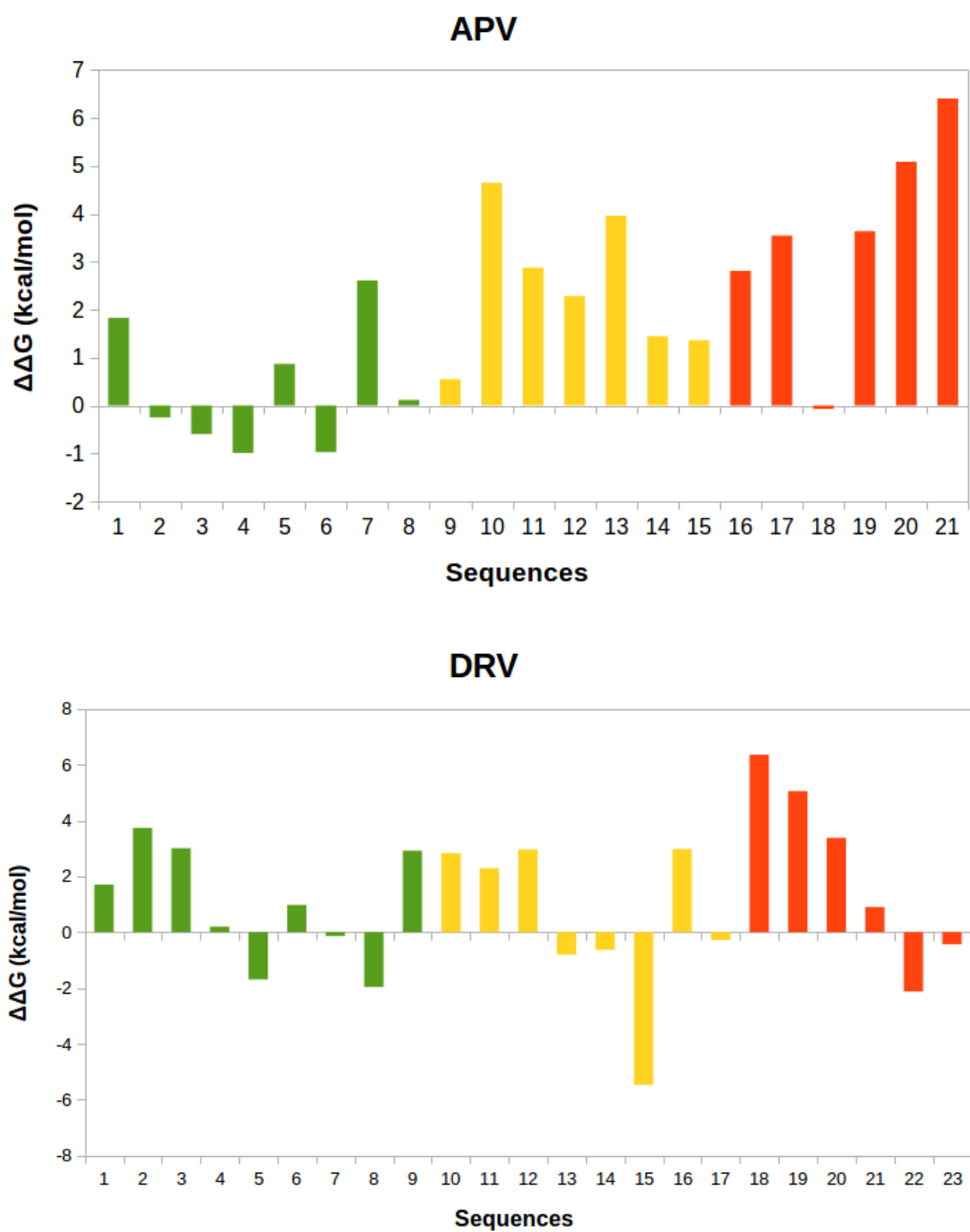
Figure S2. MD/MMPBSA relative change in APV (top panel) and DRV (bottom panel) binding energies (kcal/mol). Sensitive (<20), intermediate (20>= =<60) and resistance (>60) HIVdb values for each sequence are shown in green, yellow and red colors, respectively.