

# Understanding the Structural and Functional Effects of Mutations in HIV-1 Protease Mutants Using 100ns Molecular Dynamics Simulations

Blue Waters Undergraduate Petascale Education Program

Christopher D. Savoie  
University of New Orleans  
2000 Lakeshore Drive  
New Orleans, LA 70148  
csavoie@uno.edu

David L. Mobley  
University of New Orleans  
2000 Lakeshore Drive  
New Orleans, LA 70148  
dlmobley@uno.edu

## ABSTRACT

The Human Immunodeficiency Virus type 1 protease (HIV-1 PR) performs a vital role in the lifecycle of the virus, specifically in the maturation of new viral particles. Therefore, delaying the onset of AIDS, the primary goal of HIV treatment, can be achieved by inhibiting this protease[2]. However, the rapidly mutating virus quickly develops drug resistance to current inhibitors, thus novel protease inhibitors are needed.

Here, 100ns molecular dynamics (MD) simulations were conducted for the wild type and two mutant proteases to gain insight into the mechanisms by which the mutations confer drug resistance. Several different metrics were used to search for differences between the wild type and mutant proteases including: flap tip distance and root-mean-square deviation (RMSD), mutual information, and Kullback-Leibler divergence. It was found that at the 100ns timescale there were no large differences in the structure, flexibility and motions of the wild type protease relative to the mutants, and longer simulations may be needed to identify how the structural changes imparted by the mutations affect the protease's functionality.

## General Terms

Measurement, Experimentation, Design

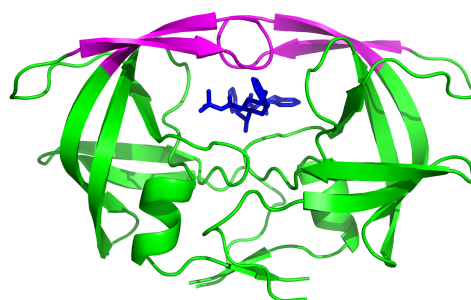
## Keywords

Blue Water Undergraduate Petascale Internship, HIV-1 Protease, Parallel Computing, Molecular Dynamics Simulation

## 1. INTRODUCTION

Currently, there is a considerable effort being put into the development of novel drugs to combat HIV by many researchers around the world. Part of that work focuses on

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Copyright © JOCSE, a supported publication of the Shodor Education Foundation Inc.



**Figure 1: A cartoon representation of the 198 amino acid wild type HIV-1 PR bound to indinavir. The flap tips are colored magenta.**

designing drugs that inhibit the homodimeric, aspartyl HIV-1 PR. This enzyme, one of only a handful of targets for HIV treatment, serves a key role in the lifecycle of HIV by processing gag-pol polyproteins into discrete, functional proteins. Inhibiting the protease prevents it from performing proteolysis, resulting in immature and noninfectious virus particles[2]. Therefore, inhibition of the protease can delay the onset of AIDS; however, mutations rapidly develop in it from one viral generation to the next in response to antiviral treatment[4]. These mutations often lead to the formation of antiviral drug resistance.

By using MD simulations to gain insight into the mechanism by which drug resistance occurs, we hope to aid in the development of adaptive inhibitors which show high activity towards both wild type and mutant proteases. Because a protease inhibitor could be designed to take advantage of conserved regions vital for the protease's function, there is hope adaptive inhibitors can be developed.

After being selected to take part in the Blue Waters Undergraduate Petascale Education Program (BW-UPEP), I,

Christopher Savoie, was given the opportunity to participate in this research project. The BW-UPEP began with a two week intensive training workshop held at the National Center for Supercomputing Applications on the University of Illinois at Urbana Champaign campus. The workshop was both a hands on learning experience and a survey of high performance computing and parallel programming. It provided an introduction to using MPI for running parallel programs along with an overview of how to use programs such as VMD and biomolecular simulation software, among other things. Following the workshop I returned the University of New Orleans where this research project was performed under the guidance of my mentor, Dr. David Mobley.

## 2. RESEARCH PLAN

### 2.1 Goals and Milestones

Our chief goal is to understand how mutations in the protease's amino acid sequence adversely affect ligand binding. By looking at inhibitors' binding affinities relative to the wild type and the mutants, we commonly see at least an order of magnitude decrease in the mutant's binding affinity, resulting from a single amino acid substitution[15, 11]. To understand why this occurs, we planned to setup, run, and analyze MD simulations of known HIV-1 PR variants with several ligands. Experimentally, each protease and ligand pair we chose shows varying degrees of affinity loss when mutations are introduced in the wild type[11].

To set up the simulations, we began with protein structures taken from the Protein Data Bank (PDB)[3] with mutations introduced as needed using PyMOL's mutagenesis tools [14]. After the models were built, the system was benchmarked to determine an efficient number of nodes to run it on. That is, we wanted to know how many nodes would give us the best value in terms of the wall clock time and the computational expense associated with that wall clock time. Because eventually the additional reduction in wall clock time from using an additional node becomes relatively small, adding more computing resources to the simulation becomes counterproductive at some point.

To accomplish our main goal, we planned to analyze the simulations' trajectories, using several methods to gauge how the mutations alter the function of the wild type protease. First, because we hypothesized that the relatively flexible flap tip region of the protease might show differences across HIV-1 PR variants, we chose to monitor the flap tip distance and RMSD. This flexible flap tip region of the protease is vital for a functional protease as well as ligand binding. As a result, changes in the interactions between the flap tips along with changes in their mobility and flexibility are important in the develop of novel inhibitors.

While hydrophobic interactions are primarily responsible for stabilizing a protein in solution, hydrogen bonds play an important role in fine tuning the structure. Because of this, we also wanted to explore the hydrogen bonding between the flap tips to see what effect mutations have there. Finally, we wanted to study correlated motions in the protease using mutual information, and we planned to see how mutations alter the torsional degrees of freedom at side chain dihedrals by comparing the distributions of angles from the wild type and mutant using Kullback-Leibler (KL) divergence.

Because of the amount of work involved and to provide a measure of progress, we formulated several milestones for the project with the majority of the work taking place during the summer. The four main milestones were as follows: (1) set up the various protease/ligand combinations and scripts for running the simulations, then submit the jobs for the first set of simulations to the queue; (2) write Python scripts utilizing GROMACS tools to perform a structural analysis by making measurements of flap tip hydrogen bonding, distance, and RMSD; (3) while running the second set of simulations, begin structural studies of the first by reviewing the initial analysis results and by producing plots of the data; and (4) perform additional structural studies using KL divergence and mutual information to look for patterns of change introduced by the drug resistance mutations.

### 2.2 Simulation Setup

Simulations were performed on the wild type HIV-1 PR and two variants. The first variant of HIV-1 PR contains the mutation I50V which is located in the protease's flap tips. The second variant has mutations V82F and I84V which are located towards the catalytic dyad in the center of the protein. Moreover, lopinavir and indinavir were simulated in the presence of the wild type and I50V proteases while ritonavir was simulated with the wild type and I82F/I84V protease. Experimentally, these mutants dramatically reduce inhibitor binding affinities[11]. For example, the V82F/I84V HIV-1 PR mutant nearly shows a 400 fold decrease in affinity for ritonavir and the I50V HIV-1 PR mutant shows a roughly 600 fold decrease in affinity for lopinavir[11].

The wild type protease's crystallographic structure was obtained from the PDB file 2BPX. From this structure, all attached water molecules were removed. Next, using PyMOL's Mutagenesis Wizard, the above mentioned mutations were inserted and the two resulting structures stored along with the wild type's. Following this, the three protein structures were protonated using Multi-Conformation Continuum Electrostatics (MCCE)[13, 1]. Following protonation, it was confirmed that both catalytic, aspartyl residues (25D and 25D') retained a neutral protonation state. It has been suggested that correct protonation state prediction at these residues is necessary for accurately calculating binding free energies[18].

As a starting point for simulating the bound protease, the binding mode for indinavir, lopinavir, and ritonavir was derived respectively from the PDB files: 2BPX, 1MUI, and 1HXW. Next, AM1-BCC partial charges were assigned for each ligand. Then a Generalized AMBER Force Field (GAFF) was used for ligand parameters[16, 17] and ffamber99sb force field parameters for the protein[6]. After the force fields for the ligand and protein were selected and applied to the system, a simulation box was set up. A dodecahedral shaped box with edges 1.2nm from the solute was used and applied to the models using the GROMACS tool editconf. Finally, bulk TIP3P water was inserted to fill the simulation box using the GROMACS program genbox.

Finally, using GROMACS 4.0.7 in conjunction with MPI, we performed energy minimization, equilibration and dynamics on our system[5]. A steepest decent minimization of 500 steps was performed. Next, constant pressure and constant

volume equilibration was carried out on the system. While this was occurring, the system's temperature was raised to 300K and the pressure was adjusted to 1atm. Equilibration steps were carried out for a total of 1100ps. Finally, during the production stage, we used a 2fs timestep for our simulations which were ran to 100ns, and we recorded data every picosecond. Ultimately, two duplicate simulations were ran for each protein and ligand pair.

### 3. CHALLENGES AND OPPORTUNITIES

As a Chemistry major, I do not have an extensive background in computer programming, having had only one formal programming course in C++. In addition, I had little experience with scientific computing environments such as Linux. To be successful in reaching the set milestones, I had to first become familiar with Linux, learning how to navigate the new environment. In addition, I had to learn Python scripting skills to run most of the analysis tools on the simulation data. At times, some Perl and R were necessary, and a level of understanding had to be achieved to move forward.

Human error guarantees that mistakes will be made when scripting, therefore skills must be developed to find and correct bugs to again move forward. Troubleshooting is not only limited to scripting errors; and occasionally, this project required looking for errors in environmental variables to ensure a script had access to the required modules.

An additional challenge was learning how to deal with such a large problem. One way this was done was using a reductionist approach. Large problems, such as the central question of this project, can't be solved as a single problem. Instead, it was broken down into many smaller problems. Learning how to break large assignments like this down to smaller pieces that could be handled day-to-day was key to success.

Overall, reaching this project's main goal was a challenge, requiring the development of new skills ranging from scripting to troubleshooting. However, it clearly provided many opportunities to advance my skills as a researcher. That is, it fostered my ability reduce large problems down to ones of workable size and taught me how to work with others in a research setting.

## 4. ANALYSIS AND RESULTS

### 4.1 Flap Tip Distance, RMSD and Hydrogen Bonding

Searching for clues to give us insight into the mechanism by which drug resistance arises, we first investigated several structural and chemical features of the flap tips. A list of the tools used in this analysis and the properties they measure are given in Table 1. The flap tips are relatively dynamic compared to the rest of the protease, moving between closed, semi-open, and open states when binding ligands. Thus, their distance of separation may be important for gauging their activity towards a given inhibitor. The flap tip distance was defined as the distance between the centers of mass of residues 50 and 50'. There were not any large differences in the flap tip distance between mutant and wild type pairs in many cases as seen in Figures 2-3. And, there was not

| Tool    | Property                             |
|---------|--------------------------------------|
| g_dist  | Flap tip distance                    |
| g_rms   | Flap tip RMSD                        |
| g_hbond | Number of hydrogen bonds             |
| MutInf  | Mutual information and KL divergence |

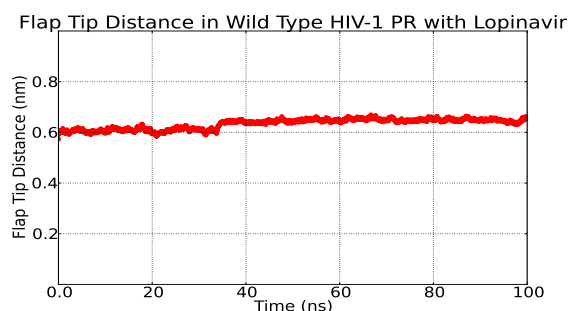
**Table 1: Analysis tools used and properties measured**

a consistent pattern of the mutant or wild type proteases displaying a larger distance of separation between the flap tips on average. Moreover, while the flap tip distance in the mutant in Figure 3 does appear to have more mobility, this is not consistently seen in other simulations.

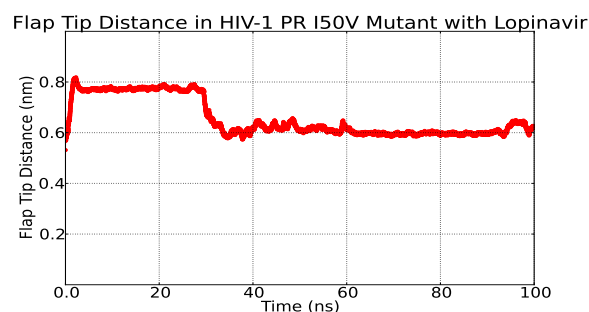
For each protein and ligand pair that was simulated, the RMSD of the flap tips was determined to give us more insight into the behavior of the flap tips. Included in the RMSD measurement were residues numbered 46-55 and 46'-55'. As seen in Figures 4-5 for one particular wild type and mutant set, there is not a large large difference in the ending value of the RMSD. That is, the flap tips in both the mutant and wild type end up displaying nearly the same about of flexibility. While the flap tips do show greater flexibility in some of the mutant cases, this is offset by an equal number of wild type cases showing a higher RMSD.

Another possible effect of mutations on the protease was hypothesized to be a decrease in the interflap hydrogen bonding. Therefore, the number of hydrogen bonds between the flap tips over time was also used as a metric for identifying differences between the wild type proteases and the mutants. Residues 48-53 and 48'-53' were included in the calculation; other residues in the flap are too far away from the opposing flap to form hydrogen bonds. Thus, they were not included in the calculation. Overall, the results do not show either the mutants or wild type proteases having more hydrogen bonds between flap tips on average at this timescale.

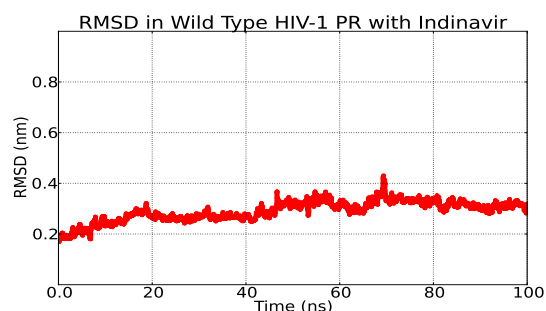
While we did not see a difference in the overall amount of hydrogen bonding between the flap tips in the protease and wild type proteases, we wanted to know if the mutations possibly interfered or changed the network of hydrogen bonds between the flap tips. Moreover, it has been suggested in the literature that the stability of particular hydrogen bonds there plays a key role in governing the transitions of the flap tips from a closed to semi-open conformation[9]. Therefore, the number of hydrogen bonds between specific pairs of interflap residues was calculated as a function of time. From our simulations, however, there is no clear change in the hydrogen bonding pattern in the flap tips. Specifically, across cases containing the wild type protease we do not see consistent hydrogen between particular interflap residues which may suggest the structure and properties of the ligand alter the dynamics there. However, comparing duplicate simulations of the same wild type/ligand pair, we again did not see any agreement in the pattern of interflap hydrogen bonding from one copy to the other. After we completed our analysis using the above three mentioned metrics and having not found significant structural differences between the wild type and mutant proteases, we decided to explore the correlations and patterns of motion within the protease.



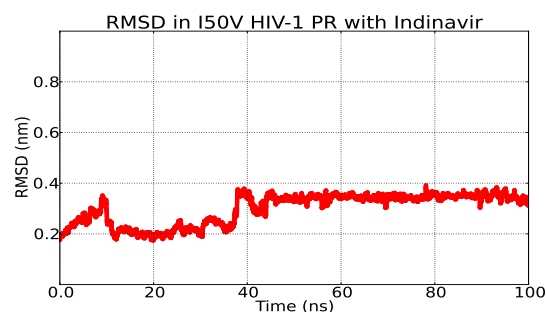
**Figure 2:** A running average of the flap tip distance versus time in the wild type HIV-1 PR with lopinavir bound.



**Figure 3:** A running average of the flap tip distance versus time in the I50V HIV-1 PR mutant with lopinavir bound.



**Figure 4:** A running average of the flap tip RMSD versus time in the wild type HIV-1 PR with indinavir bound.



**Figure 5:** A running average of the flap tip RMSD versus time in the I50V HIV-1 PR mutant with indinavir bound.

#### 4.1.1 Mutual Information

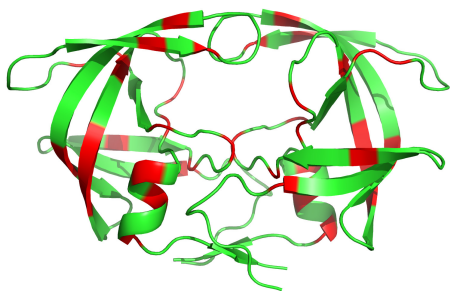
One way we sought to quantify the degree of correlation between a pair of residues was to measure their mutual information. This metric communicates to us the degree to which two random variables are linked; for example, high mutual information indicates a low uncertainty in one random variable given information about the other, and zero mutual information between two random variables implies the two variables are independent[8].

## 4.2 Correlation patterns

Using MutInf, a tool for measuring the mutual information between pairs of residues, we were able to put together mutual information matrices showing the degree to which one residue's dihedral movements are correlated to another's [10]. This was done by comparing the distribution of angles for each torsional degree of freedom for each residue to those of each other residue. Once we identified which residues were correlated in each simulation with MutInf and put together matrices showing that information using the R statistical package, we sought to see how the correlations were affected by the presence of mutations[12].

Identifying differences between wild type and mutant mutual information matrices cannot be achieved by simply taking the difference between two matrices because of a high level background noise. This was found by taking the difference between matrices from duplicate simulations of the same protease and ligand pair. As a result, we found it more useful to look at residues exhibiting a correlation above a cut off of 1.2kT in each case. This cutoff was the maximum correlation in the I82F/I84V HIV-1 PR variant bound to ritonavir in the first set of simulations, and it was the smallest correlation we found in all the cases we examined. This helped by eliminating correlated, adjacent residues from consideration which are not very interesting. One would expect two adjacent residues with bulky side chains to show some degree of correlation since steric hindrance by one residue's side chain may influence when the other's can and cannot move.

Once we identified the pairs of residues showing a high degree of correlation, the ones sharing a common residue were grouped together. We then compared the groups of correlated residues in the wild type protease to those in the mutant protease. We saw a high degree of variability in the location of the groups as well as the number of groups in



**Figure 6: HIV-1 PR with residues found to have a 5% or greater mutation probability during drug treatment colored red[11].**

most cases. Moreover, comparing the results from copies of the same case to one another, we again do not see much overlap between the locations of the groups of correlated residues. Further, in some cases, we had very few groups to use as a basis for comparison because of the chosen cut off value.

#### 4.2.1 Mutation Sites and Correlated Residues

We also thought it interesting to explore whether drug induced mutations in the protease arise in correlated residues more often than not. It has been determined experimentally that 44 residues in the protease show a mutation probability greater than 5% in patients undergoing treatment for AIDS with protease inhibitors[11]. These 44 residues are shown in Figure 6. We calculated a 39% probability that a residue in a random picked pair would be at a mutation site. Next, when we determined the percentage of residues in correlated groups that are at mutation sites for each case, and we did not find a large deviation from that percentage. This suggests that these mutations do not occur at residues with correlated motions more often than other sites.

### 4.3 Kullback-Leibler divergence

A final analysis was done using Kullback-Leibler (KL) divergence. This analysis compared the overlap between a distribution of dihedral angles for each residue in the wild type protease to the corresponding residue in the mutant protease, according to equation 1. Ideally, this would allow one to identify areas where mutations alter the flexibility and motions of the protease. Before we performed this analysis, the simulation trajectory data from duplicate simulations of the same protein/ligand case was combined.

$$D_{KL}(P||L) = \int P(i) \log \frac{P(i)}{Q(i)} \quad (1)$$

Our results show large differences between the wild type and mutants' distributions of dihedral angles for residues at the

mutation sites as expected. However, the KL divergence at other residues, which were not mutated, is more interesting. In the second HIV-1 PR variant with mutations V82F and I84V, we still see a significant divergence in several of the flap tip residues as shown in Figure 7. Perhaps, the motions of these residues are transferred through the ligand to the residues at the flap tips.

## 5. CONCLUSIONS

In summary, we do not see significant structural differences between the wild type and mutant proteases at the 100ns timescale. Specifically, both mutants and the wild type do not largely differ with respect to their flap tip distance and RMSD on average. Also, we did not observe any patterns in the interflap hydrogen bonding in the same case

duplicate simulations, making it hard to identify alterations in the hydrogen bonding network induced by mutations. Next, using the results of a mutual information analysis to group correlated residues, we did not observe consistent patterns in the location of the correlated groups. And, there was not a correlation between the location of correlated residues and mutation sites. Finally, we performed a KL divergence analysis which showed the mutated residues do have significantly different distributions of torsional angles. However, only in one case involving the V82F/I84V HIV-1 PR variant do we see the mutations possibly altering the motions of residues elsewhere in the protease. Moreover, it is seen that the residues in the flap tips also show a high KL divergence from the wild type in this case, suggesting the motions of residues 82, 84, 82', and 84' may be transferred through the ligand to the flap tips.

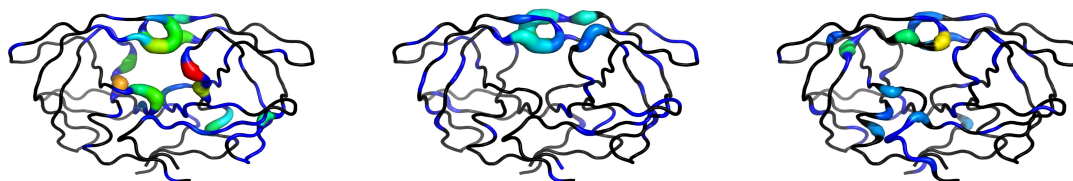
Despite the lack of large structural differences between wild type and mutant strains of the HIV-1 PR observed in our simulations, there is hope that future studies may reveal key changes conferred by the mutations onto the protease thereby lending insight into the mechanism by which resistance arises. One strategy may be to run simulations of the unliganded HIV-1 PR variants at the same timescale, since large scale flap tips conformational change has been observed to occur spontaneously on the nanosecond timescale by others[7]. Alternately, for the ligated enzyme, longer simulations on the microsecond timescale may be needed.

## 6. IMPACTS

This undergraduate research internship has helped motivate my decision to go to graduate school to study computational chemistry. While I considered going to graduate school a possibility prior to participating in this internship, the internship helped reaffirm my interest in it by providing the opportunity to participate in real research, proving it can be quite a rewarding experience.

In graduate school, I hope to continue my study of computational chemistry and learn more about different simulation methodologies and how they can be applied to solving real world problems. To further my abilities in this area, I plan to take courses in statistical mechanics, protein chemistry, and chemical kinetics and dynamics.





**Figure 7: Putty representation of HIV-PR showing KL divergence for three variants. From left to right: a.) V82F/I84V, b.) I50V, and c.) wild type. Thicker putty indicates a greater divergence. Color also is used to indicate each residue's KL divergence with black indicating zero divergence followed by blue (low divergence) progressing to red (high divergence).**

## 7. RECOMMENDATIONS

The Blue Waters Undergraduate Petascale Education Program provided a great introduction to the world of scientific computing. At the workshop, each intern was provided an abundance of resources, ranging from notes on different HPC architectures to information on topics specific for that interns project. Moreover, this workshop served a vital role in providing skills necessary to successfully utilize high performance computing resources. Furthermore, the BW-UPEP provided excellent training for someone with little prior experience in HPC, making parallel computing concepts readily understood.

Unfortunately, the limited time of the two week workshop only allows so many topics to be covered. One way to maximize the benefit to each intern may be to schedule times where the interns are separated into groups based on their project's area. For example, groups could be centered around the life sciences, physical sciences and engineering, and computer science. When the groups meet, topics taught would relate only to the relevant subject area.

## 8. ACKNOWLEDGMENTS

I would like to thank Dr. David Mobley of the University of New Orleans for his guidance throughout this project and for the time he spent teaching me about computational chemistry. Also, thanks to Chris McClendon at UCSF for his technical help and insight with his mutual information and KL divergence code. Finally, thanks to Shodor and NCSA for supporting this project.

## 9. REFERENCES

- [1] E. Alexov and M. Gunner. Incorporating Protein Conformational Flexibility into pH- Titration Calculations: Results on T4 Lysozyme. *Biophys. J.*, 74:2075–2093, 1997.
- [2] P. Ashorn, T. J. McQuade, S. Thaisrivongs, A. G. Tomasselli, W. G. Tarpley, and B. Moss. An Inhibitor of the Protease Blocks Maturation of Human and Simian Immunodeficiency Viruses and Spread of Infection. *PNAS*, 87(19):7472–7476, 1990.
- [3] H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov, and P. Bourne. The Protein Data Bank. *Nucl. Acids Res.*, 28:235–242, 2000.
- [4] L. Galiano, F. Ding, A. M. Veloro, M. E. Blackburn, C. Simmerling, and G. E. Fanucci. Drug Pressure Selected Mutations in HIV-1 Protease Alter Flap Conformations. *J. Am. Chem. Soc.*, 131:430–431, 2009.
- [5] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008.
- [6] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins Struct. Funct. Bioinf.*, 65:715–725, 2006.
- [7] V. Hornak, A. Okur, R. C. Rizzo, and C. Simmerling. HIV-1 Protease Flaps Spontaneously Open and Reclose in Molecular Dynamics Simulations. *PNAS*, 103(4):915–920, January 2006.
- [8] P. E. Latham and Y. Roudi. Mutual Information. *Scholarpedia*, 4(1):1658, 2009.
- [9] D. Li, B. Ji, K. Hwang, and Y. Huang. Crucial Roles of the Subnanosecond Local Dynamics of the Flap Tips in the Global Conformational Changes of HIV-1 Protease. *J. Phys. Chem. B*, 114:3060–3069, 2010.
- [10] C. L. McClendon, G. Friedland, D. L. Mobley, H. Amirkhani, and M. P. Jacobson. Quantifying Correlations Between Allosteric Sites in Thermodynamic Ensembles. *J. Chem. Theory Comput.*, 5(9):2486–2502, 2009.
- [11] H. Ohtaka, S. Muzammil, A. Schon, A. Velazquez-Campoy, S. Vega, and E. Freire. Thermodynamic Rules for the Design of High Affinity HIV-1 Protease Inhibitors with Adaptability to Mutations and Highly Selectivity Towards Unwanted Targets. *Int. J. Biochem. Cell Biol.*, 36:1787–1799, 2004.
- [12] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation

- for Statistical Computing, Vienna, Austria, 2011.  
ISBN 3-900051-07-0.
- [13] G. R.E, A. E.G, and G. M.R. Combining Conformational Flexibility and Continuum Electrostatics for Calculating pKa's in Proteins. *Biophys J*, 83:1731–1748, 2002.
- [14] Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.3r1. August 2010.
- [15] A. Valazquez-Campoy, S. Vega, E. Fleming, U. Bacha, Y. Sayed, H. W. Dirr, and E. Freire. Protease Inhibition in African Subtypes of HIV-1. *AIDS Rev.*, 5:165–171, 2003.
- [16] J. Wang, W. Wang, K. P. A., and D. A. Case. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.*, 25:247–260, 2006.
- [17] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and Testing of a General AMBER Force Field. *J. Comput. Chem.*, 25:1157–1174, 2004.
- [18] K. Wittayanarakul, S. Hannongbua, and M. Feig. Accurate Prediction of Protonation State as a Prerequisite for Reliable MM-PB(GB)SA Binding Free Energy Calculations of HIV-1 Protease Inhibitors. *A. Comput. Chem.*, 29:673–685, 2007.