Identification of Features Contributing to Binding Promiscuity of Small-Molecule Inhibitors for Rapidly Mutating Targets

by

Ishan Arora

Bachelor of Technology, Chemical Engineering
Indian Institute of Technology (IIT) Delhi, 2014

Master of Science, Chemical Engineering Practice
Massachusetts Institute of Technology, 2016

Submitted to the Department of Chemical Engineering
in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Chemical Engineering

at the

Massachusetts Institute of Technology

September 2017

Signature of Author ..... Signature redacted .....
Department of Chemical Engineering
August 17, 2017

Certified by ..... Signature redacted .....
Bruce Tidor
Professor of Biological Engineering and Computer Science
Thesis Supervisor

Certified by ..... Signature redacted .....
Hadley D. Sikes
Associate Professor of Chemical Engineering
Thesis Co-advisor

Accepted by ..... Signature redacted .....
Patrick S. Doyle
Robert T. Haslam (1911) Professor of Chemical Engineering
Chairman, Committee for Graduate Students

Identification of Features Contributing to Binding Promiscuity of Small-Molecule Inhibitors for Rapidly Mutating Targets

by

Ishan Arora

Submitted to the Department of Chemical Engineering
on August 17, 2017 in Partial Fulfillment of the
Requirements for the Degree of
Master of Science in Chemical Engineering

Abstract

HIV infection has become a persistent worldwide epidemic despite the continuous development of novel inhibitors. A key challenge in combating HIV and other pandemic viral infections is the ability of the virus to mutate at an enormous rate and rapidly develop resistance to existing drugs. Among the various strategies that have been explored for the design of broadly binding HIV protease inhibitors, the substrate envelope hypothesis which is based on the idea of designing drugs that mimic the structural features of substrates has proved particularly effective. However, studies aimed at probing the substrate envelope hypothesis have found that the substrate envelope is a contributory but not sufficient property for robust binding and hence it is important to develop a better understanding of the other factors that contribute to binding promiscuity. This study investigated the key features which differentiate robust HIV protease inhibitors from susceptible HIV protease inhibitors by examining the interactions of certain known flat and non-flat binders with the different residues of HIV protease in terms of binding energy and number of contacts and correlating this analysis with the information about the mutational space of the virus. It was found that the promiscuous inhibitors, susceptible inhibitors and substrates all interact with the same set of HIV protease residues, some of which are vulnerable to primary mutations. The total contribution to the binding of an inhibitor/substrate to HIV protease from the HIV protease residues that are associated with primary mutations was observed to be a vital attribute separating flat binders from susceptible binders, with a greater contribution to binding from these residues translating into a higher susceptibility of the inhibitor to primary mutations. Certain strategies were proposed for incorporating these inferences in the computational drug design framework in order to generate robust HIV protease inhibitors. Although the analysis in this project was carried out using HIV protease as the model system, it is envisaged that the results obtained here would be generalizable to other rapidly mutating targets and hence these insights would facilitate drug design in the case of the outbreak of new epidemics of highly mutable infectious agents.

Thesis Supervisor: Bruce Tidor

Title: Professor of Biological Engineering and Computer Science

Acknowledgements

I would like to thank my research advisor, Professor Bruce Tidor for his valuable advice and guidance regarding the different aspects of this project. I am grateful to my thesis co-advisor, Professor Hadley Sikes for her insightful suggestions and ideas which helped streamline my work on this project. I wish to thank all the members of the Tidor group for helping me build a strong foundation in this field and giving me thought-provoking ideas regarding several aspects of this project. I owe a debt of gratitude to my parents and other family members for their unconditional support and cooperation throughout my time at MIT.

Contents

Introduction

Drug Resistance

Resistance is the naturally occurring response of any microorganism that faces the selective pressure of inhibitor therapy and is pronounced for rapidly mutating targets like HIV due to the high level of viral replication and low fidelity of the reverse transcriptase enzyme [1]. Antiviral drug resistance is defined by the presence of viral mutations that decrease drug susceptibility compared to the susceptibility of wild-type viruses. In the presence of antiretroviral drugs, variants with reduced susceptibility to these drugs tend to get selected and become the predominant quasispecies [2]. Drug resistant strains can be transmitted from one patient to another or can be acquired as a consequence of the use of antiretroviral inhibitors. Multi-class drug resistance occurs when a virus that is resistant to one drug acquires resistance to another drug [3]. Drug resistance is a grave public health problem as it results in the impairment of therapies and people harboring resistant viruses can transmit them to non-infected persons [1]. Drug resistance can be measured using either genotypic or phenotypic assays. Genotypic assays look directly at the genetic material of HIV in the blood and detect mutations that cause drug resistance. On the other hand, phenotypic assays measure the in-vitro susceptibility of plasma HIV to individual drugs, in comparison with a reference wild-type virus [1].

Primary mutations refer to amino acid substitutions that significantly reduce drug susceptibility by themselves while secondary mutations are those that reduce drug susceptibility in combination with primary mutations or improve the fitness of the viruses carrying primary mutations. Secondary mutations typically have little direct effect on inhibitor binding [4].

The Molecular Virology of HIV

HIV belongs to the category of viruses known as retroviruses which carry genetic information in the form of RNA. The different steps of the HIV life cycle are binding, fusion, reverse transcription, integration, replication, assembly, budding and maturation.

The first stage of the HIV life cycle involves binding of HIV to receptors on the surface of a CD4 cell. The HIV envelope and the CD4 cell membrane fuse to allow HIV to enter the CD4 cell. As the virus enters the cell, its RNA is reverse-transcribed to DNA by the reverse transcriptase enzyme [5]. The viral DNA enters the cell nucleus where it is integrated into the host cell genome by the integrase enzyme. Once integrated into the CD4 cell DNA, HIV starts exploiting the machinery of the cell to make long chains of HIV proteins [6]. The HIV proteins and HIV RNA move to the surface of the cell and assemble into immature HIV which then buds off from the host cell. Finally, during the maturation process, the protease enzyme cuts the long polypeptide chains into smaller units that reassemble into a mature virus which can then infect other cells where the viral life cycle is repeated. The infected cells are eventually destroyed [7].

HIV displays extremely high genetic variability which enables it to evade host immunity and drugs. This large variability is a consequence of the low fidelity of the reverse transcriptase enzyme (the mutation rate of HIV has been estimated to be about 3.4×10^{-5} which corresponds to an average of $\sim$1 mutation per replication cycle), the high rate of viral replication (approximately 10^{10} virions are generated daily in an infected person) and the high rate of reverse transcriptase-mediated recombination events [6, 8]. These features of HIV make the design of promiscuous HIV inhibitors a formidable challenge.

Highly active antiretroviral therapy (HAART) is currently the standard therapy for patients infected with HIV. It consists of a combination of drugs that target HIV at different stages in its life cycle [9]. HIV entry inhibitors interfere with the binding, fusion and entry of HIV into host cells [6]. Nucleoside reverse transcriptase inhibitors (NRTIs) compete with the natural substrates for incorporation into the growing viral DNA chain. However, they lack a 3'-hydroxyl group on the deoxyribose moiety and hence block further extension of the DNA chain (chain termination) [10]. Non-nucleoside reverse transcriptase inhibitors (NNRTIs) bind to the hydrophobic pocket within the p66 subunit of the reverse transcriptase enzyme and alter the structure of the enzyme, thus preventing it from converting viral RNA into DNA [9]. Integrase strand transfer inhibitors (INSTIs)

prevent HIV genetic code from integrating with the human DNA [11]. Protease inhibitors (PIs) interfere with the formation of functionally active virions by binding to the active site of the protease enzyme [12].

HIV Protease: Function and Structure

HIV protease plays a vital role in the life cycle of HIV. It cleaves the precursor Gag and Gag-Pol polyproteins into mature, functional enzymes and structural proteins [13]. The timing of these cleavage reactions is critical as these intact polyproteins are necessary early in the life cycle of HIV when they assemble the immature form of the virus but need to be cleaved later at specific sites in order to form functioning and infectious viruses. The indispensable role of HIV protease in viral maturation has made it an attractive target for drug design [14].

HIV protease exists as a homodimer with each subunit consisting of 99 amino acid residues. The two subunits come together in such a way as to form a tunnel where they meet. The active site is at the center of the tunnel and has the characteristic Aspartic Acid-Threonine-Glycine sequence common to aspartic proteases. The central active site cavity is capped by two flexible flap regions that move to allow the substrate to enter the tunnel [15]. Each flap contains three characteristic regions: side chains that extend outward (Met46 and Phe53), hydrophobic chains that extend inward (Ile47 and Ile54) and a glycine rich region (Gly48, Gly49, Gly51 and Gly52). Ile50 remains at the tip of the turn [16]. The secondary structure of each monomer includes one α-helix and two antiparallel β-sheets [17]. Figure 1 shows a schematic representation of the structure of HIV protease along with the different residues that are associated with primary and secondary mutations. The HIV protease dimer is shown in pink ribbons and the clinical inhibitor darunavir is shown in green sticks. The primary mutations are depicted by red spheres while the secondary mutations are represented by blue spheres. The mutations are distributed on both the monomers in order to enhance visibility [15].

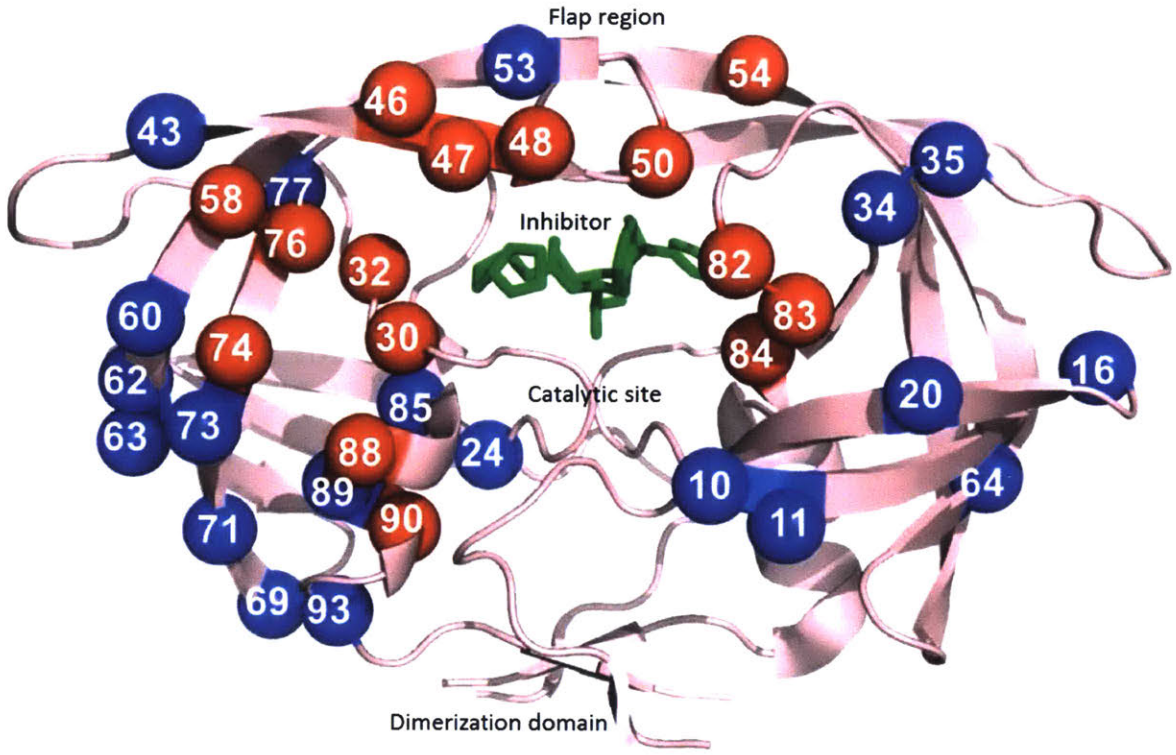

Figure 1: The HIV protease structure in complex with an inhibitor along with the sites of primary and secondary mutations (Figure adapted from [15])

HIV protease inhibitors are competitive inhibitors that tightly bind in the active site with the flaps folded into a closed conformation over the active site. These inhibitors prevent the processing of substrates by keeping the protease enzyme in a locked down state [18]. The clinical inhibitors of HIV protease provide a testimony to the effectiveness of structure-based drug design. The introduction of protease inhibitors into combination antiretroviral therapy has resulted in a marked improvement in clinical outcomes [14]. However, the emergence of drug resistance to protease inhibitors has compromised the effectiveness of treatment of HIV infections. The resistance mutations reduce susceptibility to the protease inhibitor(s) while maintaining protease function. More mutations are selected by protease inhibitors than any other class of antiretroviral drugs [18].

Robust Inhibitor Development

The development of promiscuous HIV protease inhibitors that are immune to drug resistance mutations has been the subject of intensive research for several years. Various strategies have been explored for designing drug resistance resistant HIV protease inhibitors such as using a solvent anchoring approach, utilizing a lysine sulfonamide-based molecular core and maximizing the number of hydrogen bonds with the protease backbone [19-22]. A more comprehensive approach for accomplishing the same is to incorporate the substrate envelope constraints in the computational drug design framework [22]. This strategy is discussed in detail in this section.

For drug resistance to occur, the target needs to evade drug binding but still carry out its biological function. Thus, drugs that bind to the target in a manner similar to the substrates should be less vulnerable to drug resistance as in such a case, mutations that adversely impact drug binding would also be likely to affect substrate binding and hence would not be tolerated [22, 23]. The observation that substrates with high sequence diversity adopt a conserved shape when bound to HIV protease led to the idea of the substrate envelope hypothesis according to which inhibitors that stay within a consensus substrate shape (referred to as the substrate envelope) will be less susceptible to drug resistance compared to those that protrude beyond the substrate envelope and make contacts with residues that are not contacted by the substrates [22-24]. A key feature of the substrate envelope hypothesis is that it does not require any prior knowledge of drug resistant strains [24]. The basic idea behind the substrate envelope hypothesis is depicted in Figure 2.

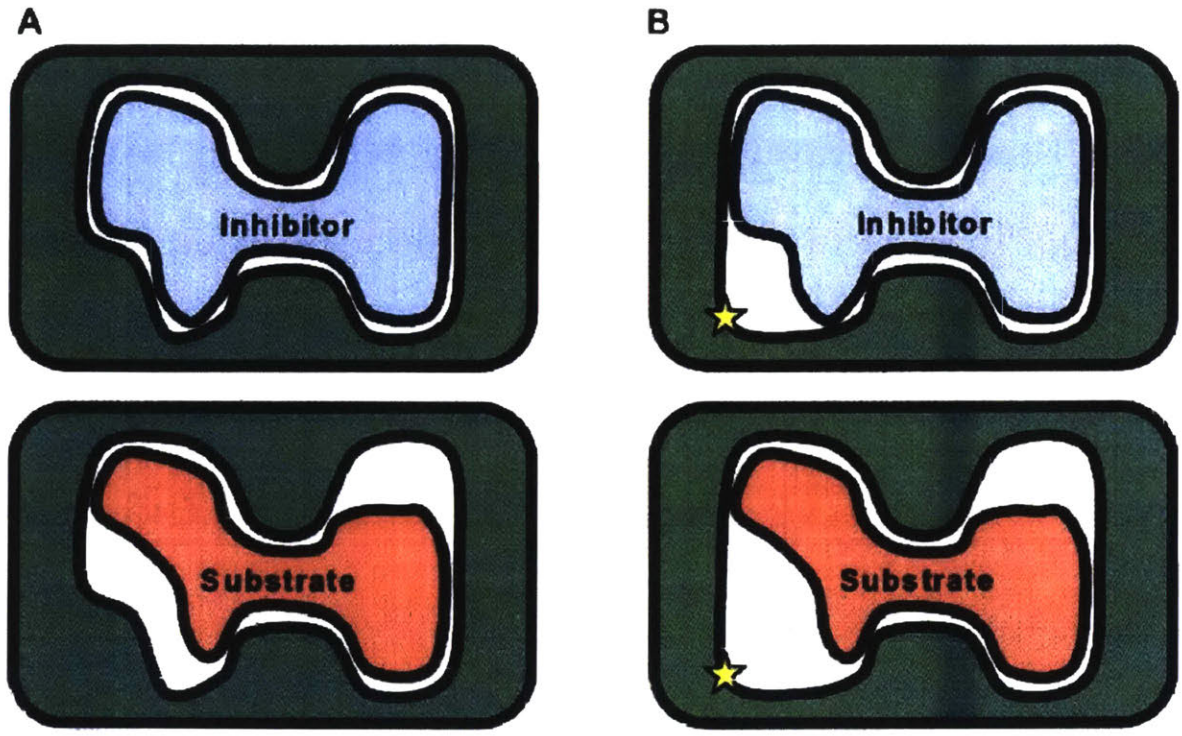

Figure 2: Schematic illustration of the substrate envelope hypothesis (Figure reproduced from [25])

As illustrated in Figure 2, in the wild-type target, the inhibitor (A, top) occupies a greater portion of the binding site compared to the substrate (A, bottom). In the mutated target, the inhibitor (B, top) loses contacts and consequently binding affinity while there is negligible effect on the binding of the substrate (B, bottom) as the substrate did not have any binding interactions with the residue(s) that mutated [24].

Extensive work has been carried out to investigate the implications of the substrate envelope hypothesis for the design of broadly binding HIV protease inhibitors that have low vulnerability to drug resistance mutations [22-28]. One such study involved the application of a computational inverse small-molecule design strategy to generate HIV protease inhibitors predicted to have favorable binding energetics while remaining inside the substrate envelope [24]. The promising candidates from this computational analysis were synthesized, assayed and characterized which

led to the identification of subnanomolar inhibitors to drug resistant variants [24]. When tested against a panel of four drug resistant protease variants: M1 (L10I, **G48V**, **I54V**, **V82A**), M2 (**D30N**, **N88D**), M3 (L10I, A71V, G73S, **I84V**, **L90M**) and M4 (**I50V**, A71V) (the primary mutations are in boldface), it was found that some of these compounds had little susceptibility to drug resistance mutations with a maximum affinity loss of 14-16 fold (flat binders), certain compounds were moderately susceptible to drug resistance mutations with an affinity loss of up to 35-80 fold (moderate binders) and some compounds were highly vulnerable to these mutations, losing more than 100 fold in affinity (susceptible binders) [24]. The substrate envelope hypothesis was further tested by designing, synthesizing and assaying derivatives of these compounds that progressively protruded outside the substrate envelope [23]. The strategy here was to design paired inhibitors of HIV protease that were very similar to one another but one respected the substrate envelope while the other violated it [23]. When tested against the same mutant panel that was used in the previous study [24], it was found that the envelope-respecting inhibitors were flat binders across the mutant panel while the envelope-violating inhibitors were susceptible to at least one drug resistant protease variant [23]. Similar results were obtained in another study which involved the design of several HIV protease inhibitors using three different approaches, two of which incorporated the substrate envelope constraints (the same scaffold was employed in all the three design approaches) [27]. It was found that the third approach (without the substrate envelope constraints) led to the design of high-affinity inhibitors but these inhibitors were highly susceptible to drug resistance mutations while the inhibitors that stayed inside the substrate envelope (from the first two approaches) had relatively flat binding profiles [27].

The scaffold that was used in these studies was adopted from the clinical inhibitors amprenavir and darunavir and is shown in Figure 3. This scaffold fits predominantly within the substrate envelope and has three sites (denoted as R^1, R^2 and R^3 in Figure 3) to which various substituent groups can be attached [27].

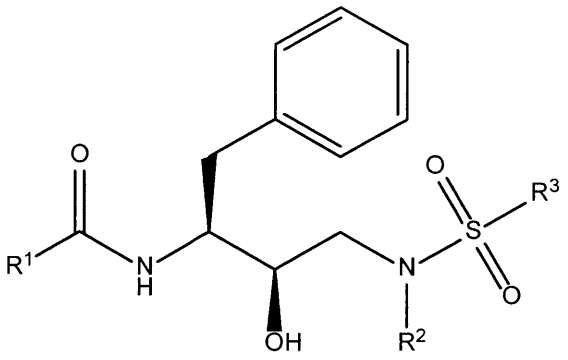

Figure 3: (*R*)-(hydroxyethylamino)sulfonamide isostere (amprenavir/darunavir scaffold)

The substrate envelope hypothesis was also assessed in a computational framework for inhibitors of Abl kinase, chitinase, thymidylate synthase, dihydrofolate reductase and neuraminidase and it was demonstrated that the total volume of an inhibitor molecule that projects outside the substrate envelope correlates with average mutation sensitivity [29].

The results of these studies suggest that the substrate envelope hypothesis is a useful strategy for the development of inhibitors that can combat the problem of drug resistance but other factors need to be considered as some of the inhibitors that were designed based on the substrate envelope hypothesis were also found to be susceptible to primary mutations [24, 30]. The overarching goal of this study was to identify the key features which differentiate flat binders from susceptible binders (with HIV protease as the model system) as this knowledge would be invaluable for the development of computational strategies for the design of drug resistance resistant inhibitors for rapidly mutating targets.

Methods

A 1.2-Å crystal structure of darunavir-bound wild-type HIV protease was obtained from the Protein Data Bank (PDB) (accession code 1T3R). Crystal structures of four inactivated HIV protease-substrate peptide complexes were also obtained from the PDB (accession codes 1F7A, 1KJ7, 1KJF and 1KJG). In all these structures, the terminal side-chain dihedral angle for

asparagine, glutamine and histidine residues was considered for a 180° rotation if this would improve the geometry of local hydrogen bonding [24]. In the protease-substrate complexes, all water molecules were removed except for five that were highly conserved across all structures [31]. In the darunavir-bound structure, any crystal water that made two hydrogen bonds was retained. Furthermore, the conserved flap water molecule was retained. Hydrogen atoms were added to all the structures using the HBUILD module in CHARMM using the CHARMM22 parameter force field. Missing side chain atoms were built back into the structures using CHARMM and the CHARMM22 parameter force field. Hydrogen atoms and computationally added heavy atoms were minimized using steepest descent method for 1000 steps followed by 250 rounds of Newton-Raphson energy minimization [32].

The INTE command in CHARMM was used for computing the van der Waals interaction energy between the various residues of HIV protease and the different inhibitors and substrates included in this study. The COOR DIST command in CHARMM was used for calculating the number of contacts less than a certain specified distance (3 Å, 5 Å and 7 Å) made by the various HIV protease residues with different inhibitors and substrates.

Results and Discussion

In order to identify the key attributes that separate robust HIV protease inhibitors from susceptible HIV protease inhibitors, the interactions of certain inhibitors designed previously in the Tidor group and substrates with the various residues of HIV protease were studied in terms of binding energy and number of contacts. Five HIV protease inhibitors were included in this analysis: KB-83, AF-68, AF-77, AD-93 and KC-08. These inhibitors were designed previously in the Tidor group based on the substrate envelope hypothesis using the same scaffold (shown in Figure 3). However, it was found from experimental studies that while KB-83, AF-68 and AF-77 are broadly binding inhibitors (flat binders), AD-93 is a borderline inhibitor (moderate binder) and KC-08 is a susceptible inhibitor (non-flat binder) [23, 24]. Four substrates were included in this study: CA-p2, p2-NC, p1-p6 and RT-RH. Substrates are in essence natural flat binders because

when the virus tries to develop mutations to evade the drugs, it still needs to be able to bind to the substrates.

The van der Waals interaction energies between the various residues of HIV protease and the different inhibitors were calculated using CHARMM. The analysis of the results suggested that the top fourteen residues of HIV protease in terms of binding energy contribution are the same for all these inhibitors even though some of these inhibitors are flat binders and some are susceptible binders. All these residues contribute at least 1% of the total van der Waals energy of interaction between HIV protease and the individual inhibitors. Figure 4 compares the binding of the five inhibitors: KB-83, AF-68, AF-77, AD-93 and KC-08 to HIV protease. Only the top fourteen residues for binding are shown in this graph.

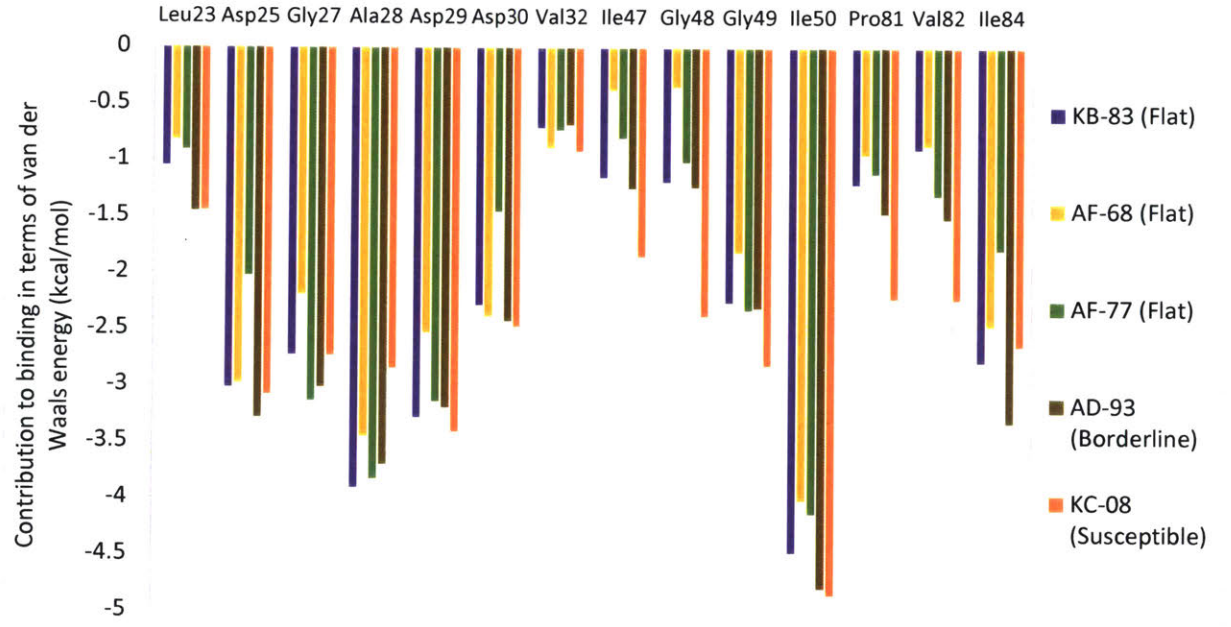

Figure 4: Comparison of binding of the five inhibitors: KB-83, AF-68, AF-77, AD-93 and KC-08 to HIV protease

Figure 5 depicts the results of a similar analysis for the four substrates: CA-p2, p2-NC, p1-p6 and RT-RH. Only the top fourteen residues for binding are shown in this graph.

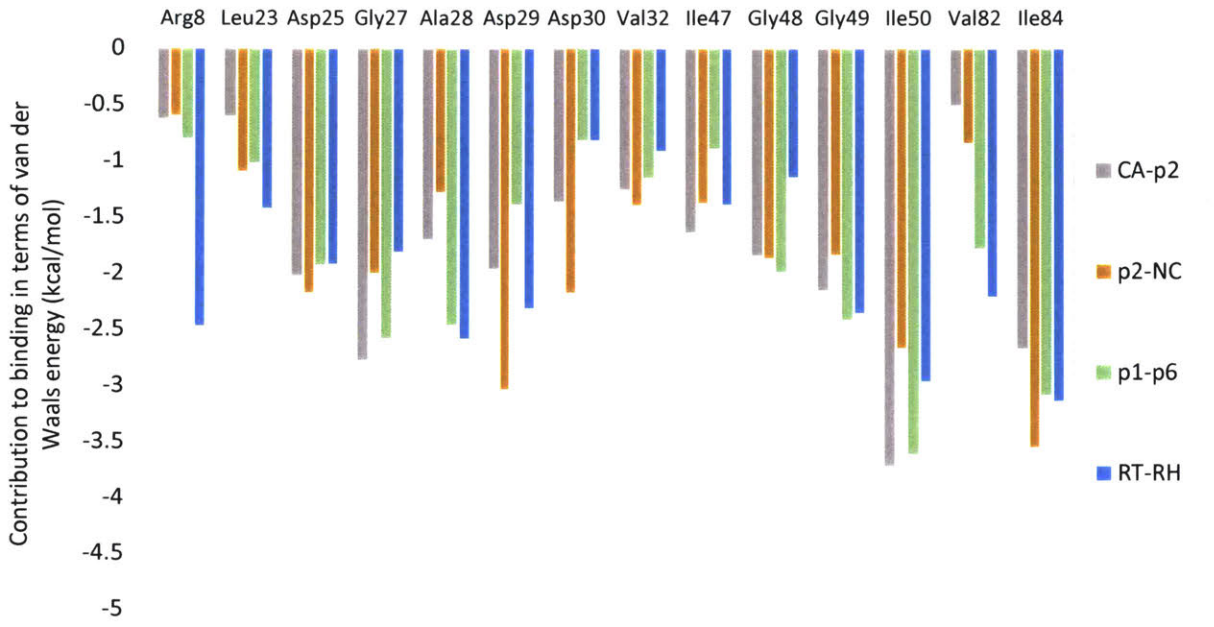

Figure 5: Comparison of binding of the four substrates: CA-p2, p2-NC, p1-p6 and RT-RH to HIV protease

It can be observed from Figure 4 and Figure 5 that thirteen of the top fourteen residues for binding are the same for all the inhibitors and substrates considered in this project. The only difference is that Arg8 features in this list for the substrates in place of Pro81 in the case of the inhibitors. Thus, the substrates and broadly binding inhibitors don't interact with different residues (in comparison with the susceptible inhibitors) but interact differently with the same set of residues.

Figure 6 compares the binding of KB-83 (robust inhibitor), KC-08 (susceptible inhibitor), CA-p2 (substrate) and p2-NC (substrate) to HIV protease.

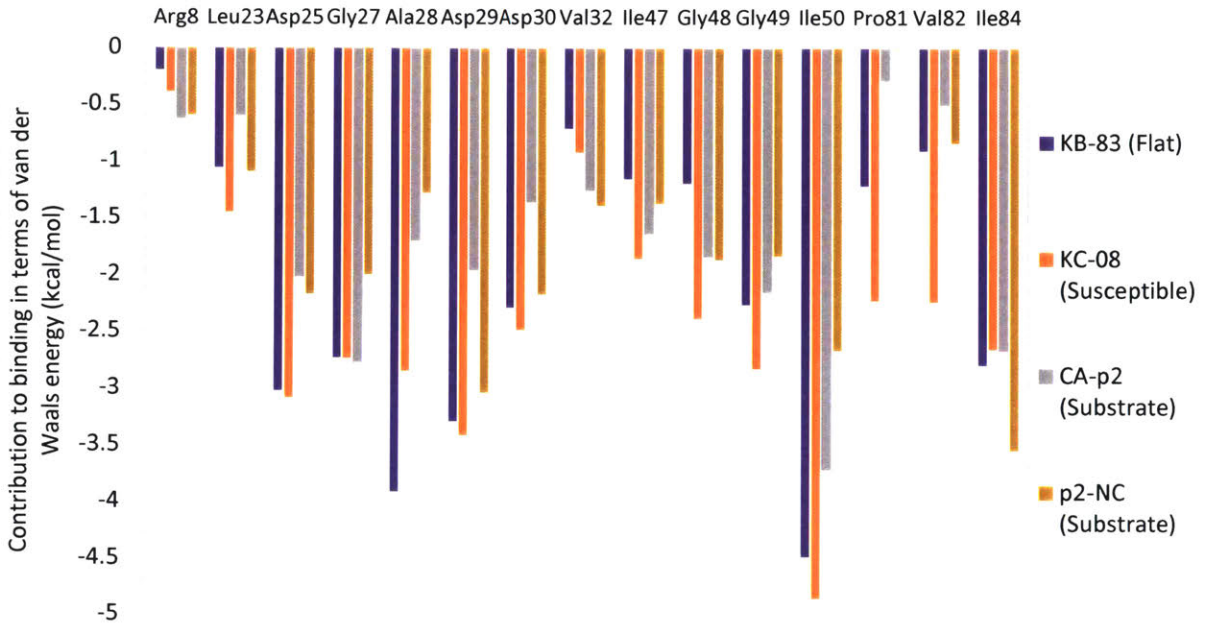

Figure 6: Comparison of binding of different inhibitors (KB-83 and KC-08) and substrates (CA-p2 and p2-NC) to HIV protease

Arg8, Pro81 and the other top thirteen residues of HIV protease for binding of inhibitors and substrates are included in Figure 6. Out of these fifteen residues, seven are associated with primary mutations: Asp30, Val32, Ile47, Gly48, Ile50, Val82 and Ile84. As substrates and broadly binding inhibitors also have significant interactions with these residues, it can be inferred that a binder can be promiscuous even if it has important interactions with certain residues that are associated with primary mutations.

Thus, both flat binders and non-flat binders interact with the same set of residues, some of which are vulnerable to primary mutations. In an effort to understand what differentiates flat binders from non-flat binders, the total contribution to the interaction energy between the HIV protease monomer and different inhibitors and substrates from residues associated with primary mutations was calculated. The results of the same are depicted in Figure 7.

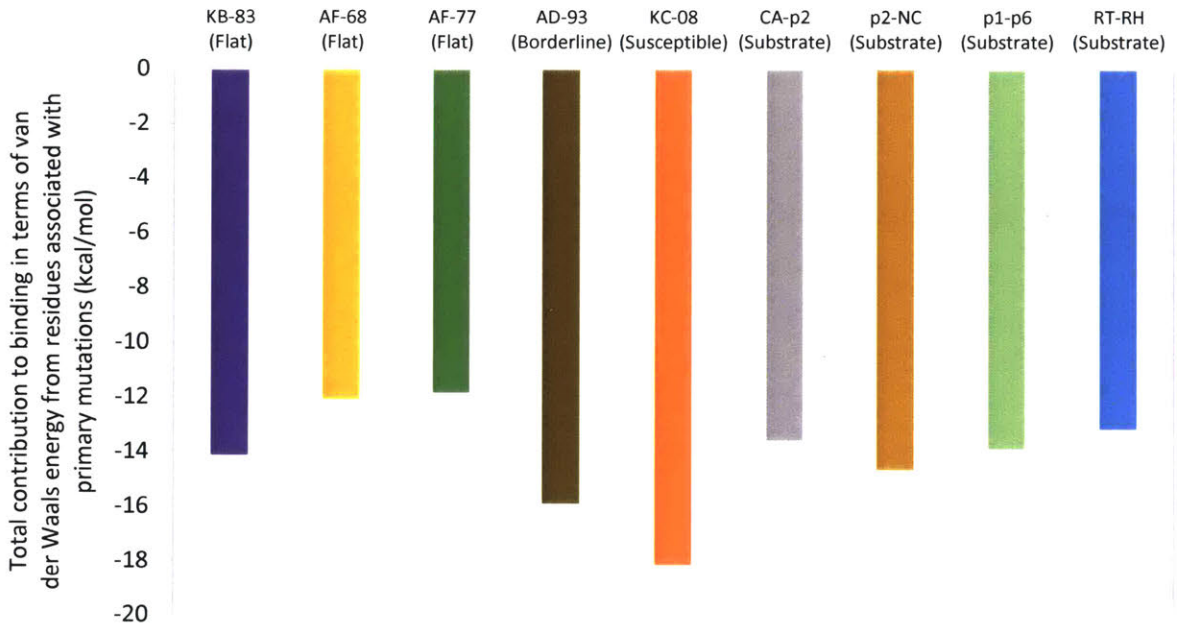

Figure 7: Total contribution to the van der Waals interaction energy between the HIV protease monomer and different inhibitors and substrates from the HIV protease residues that are associated with primary mutations

It can be observed from Figure 7 that the substrates and robust inhibitors have a lower contribution to binding from residues associated with primary mutations as compared to the susceptible inhibitors.

The worst-fold affinity loss for the different inhibitors against the mutant panel used previously in the Tidor group is presented in Table 1 along with the total contribution to the van der Waals interaction energy of these inhibitors with the HIV protease monomer from sites associated with primary mutations [23, 24, 27]. The mutant panel that was employed in these three studies [23, 24, 27] consisted of four drug resistant variants: M1 (L10I, **G48V**, **I54V**, **V82A**), M2 (**D30N**, **N88D**), M3 (L10I, A71V, G73S, **I84V**, **L90M**) and M4 (**I50V**, A71V). Here, the primary mutations are in boldface. The worst-fold loss is defined as the ratio between the K_i (inhibitory activity) values against the weakest-binding mutant and the wild-type protease. Thus, the worst-fold affinity loss is a measure of the susceptibility of an inhibitor to primary mutations.

Table 1: Worst-fold affinity loss and binding energy contribution from the HIV protease residues that are vulnerable to primary mutations for different inhibitors

Inhibitor	**Total van der Waals interaction energy of the inhibitor with the HIV protease monomer contributed by the HIV protease residues that are associated with primary mutations (kcal/mol)**	**Worst-fold affinity loss** [23, 27]
KB-83	-14.09	10
AF-68	-12.04	7
AF-77	-11.79	6
AD-93	-15.84	41
KC-08	-18.10	113

It can be observed from Table 1 that the higher the total contribution to binding (in terms of van der Waals energy) from residues associated with primary mutations, the greater the worst-fold affinity loss for that inhibitor and hence the more susceptible that inhibitor to primary mutations.

As another metric of binding affinity, the number of contacts less than 5 Å that different inhibitors and substrates have with the various residues of HIV protease were computed using CHARMM. Figure 8 illustrates the key results of this analysis for KB-83 (robust inhibitor), KC-08 (susceptible inhibitor), CA-p2 (substrate) and p2-NC (substrate). Similar to Figure 6, Arg8, Pro81 and the other top thirteen residues of HIV protease for binding of inhibitors and substrates are included in Figure 8.

A similar analysis was carried out for the number of contacts less than 3 Å and 7 Å that different inhibitors and substrates have with the various residues of HIV protease. The key results of this analysis are shown in Figure 9 and Figure 10.

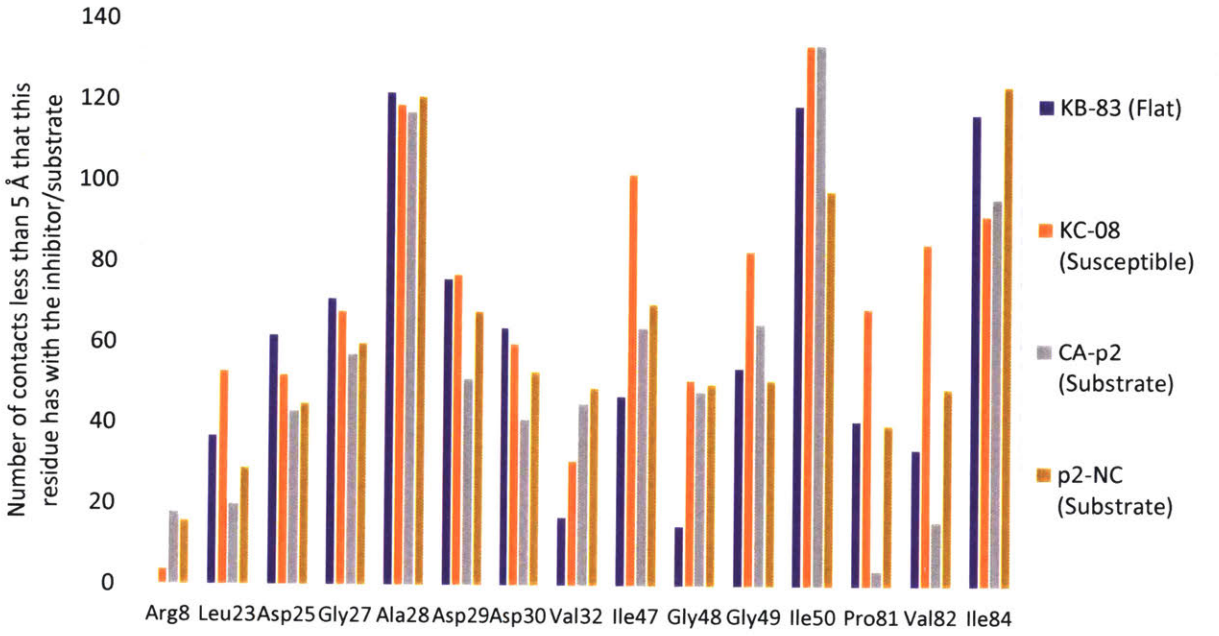

Figure 8: Comparison of binding of different inhibitors (KB-83 and KC-08) and substrates (CA-p2 and p2-NC) to the various residues of HIV protease in terms of number of contacts less than 5 Å

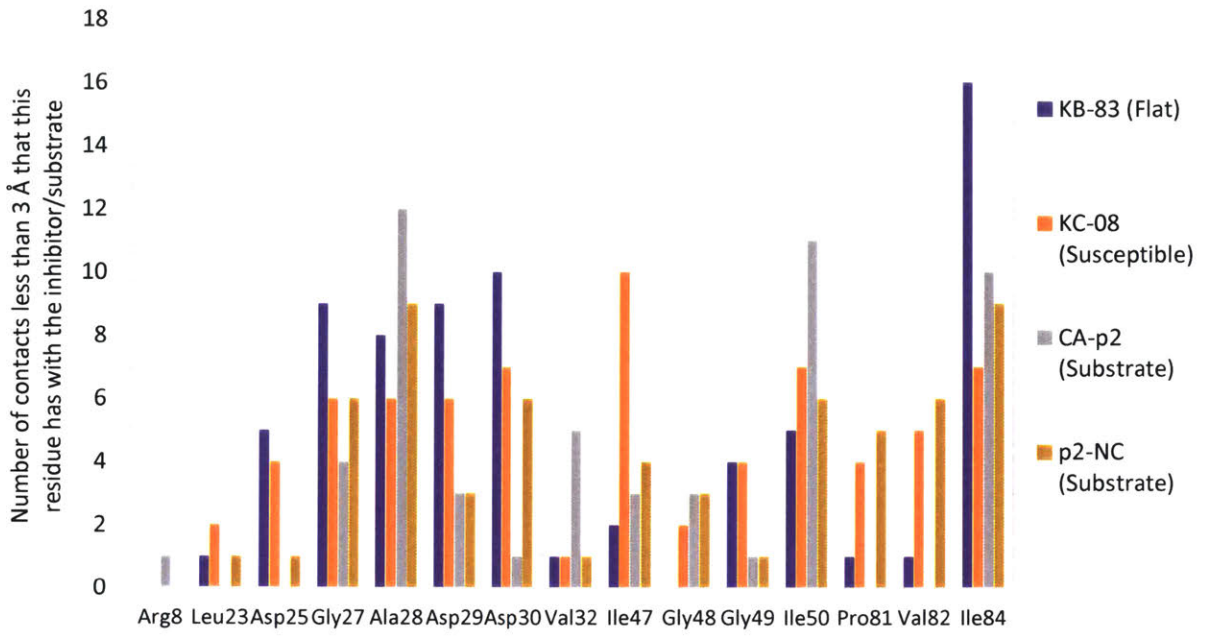

Figure 9: Comparison of binding of different inhibitors (KB-83 and KC-08) and substrates (CA-p2 and p2-NC) to the various residues of HIV protease in terms of number of contacts less than 3 Å

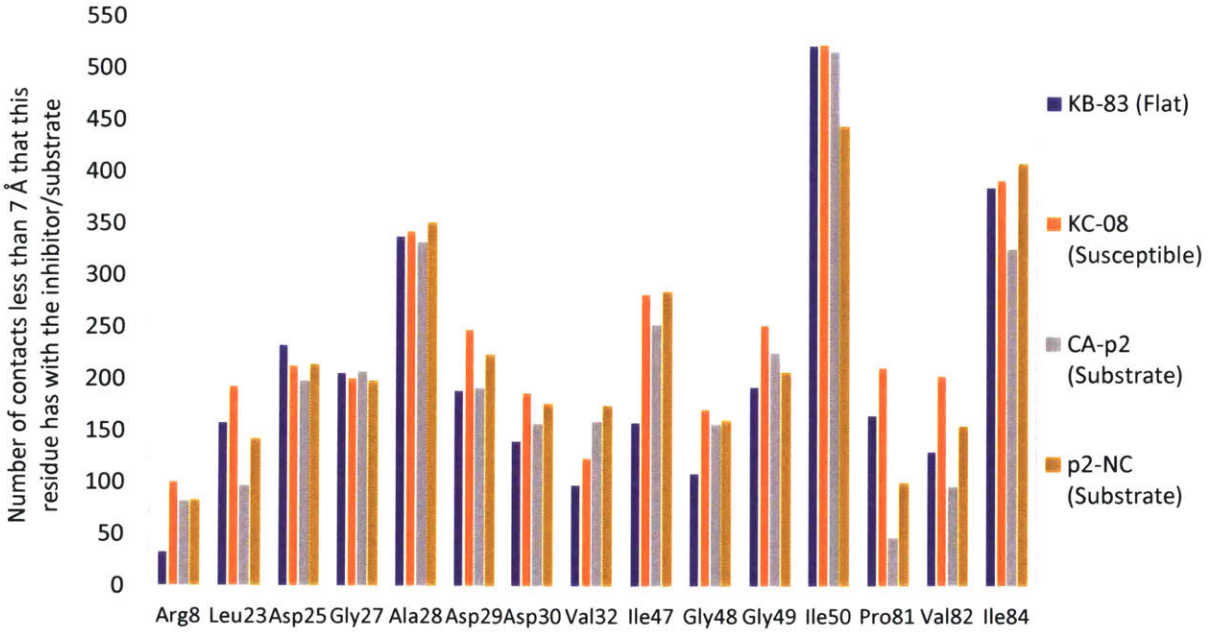

Figure 10: Comparison of binding of different inhibitors (KB-83 and KC-08) and substrates (CA-p2 and p2-NC) to the various residues of HIV protease in terms of number of contacts less than 7 Å

Based on the results in Figure 8, the top five residues of HIV protease for binding of different inhibitors and substrates (in terms of number of contacts less than 5 Å) are highlighted in Table 2 along with the number of contacts less than 5 Å.

Table 2: Top five contributors to binding of different inhibitors (KB-83 and KC-08) and substrates (CA-p2 and p2-NC) to HIV protease in terms of number of contacts less than 5 Å

KB-83 (Robust inhibitor)	**KC-08 (Susceptible inhibitor)**	**CA-p2 (Substrate)**	**p2-NC (Substrate)**
Ala28 122	Ile50 134	Ile50 134	Ile84 124
Ile50 119	Ala28 119	Ala28 117	Ala28 121
Ile84 117	Ile47 102	Ile84 96	Ile50 98
Asp29 76	Ile84 92	Gly49 65	Ile47 70
Gly27 71	Val82 85	Ile47 64	Asp29 68

In Table 2, the residues associated with primary mutations are indicated in purple. It can be observed from Table 2 that four of the top five residues for binding (in terms of number of contacts less than 5 Å) of KC-08 to HIV protease are associated with primary mutations while for CA-p2 and p2-NC, three of the top five residues for binding are prone to primary mutations and for KB-83, two of the top five residues for binding are associated with primary mutations. Thus, it can be inferred that the susceptible inhibitors interact more strongly with the HIV protease

residues that are associated with primary mutations as compared to the promiscuous inhibitors and substrates.

The DEE Drug Design software was developed in the Tidor group to explore the space of small molecules compatible with a binding site on a macromolecule like a protein. This inverse small-molecule design algorithm involves several sequential steps that are implemented in CHARMM [25]:

- Sphere packing of protein site to create a target shape that serves as a hard constraint on the size and shape of designed ligands as well as an approximation to the true molecular shape in solvation calculations
- Calculation of binding potentials on a grid within the target shape for fast energy evaluation during the design procedure
- Placement of molecular scaffolds throughout the target shape in discrete conformations and orientations
- Guaranteed combinatorial search using dead-end elimination (DEE) and A* algorithms to determine the global minimum binding energy configuration of functional groups for each scaffold placement as well as a ranked list of configurations with increasing energy
- Compilation of the structures generated across all scaffold placements into a rank-ordered list and the use of more sophisticated binding energy models to re-score the top ranked structures

The DEE Drug Design software is based on the principle of the substrate envelope hypothesis and currently does not incorporate any information about the mutational space of the virus. Based on the insights gained in this study, an effective strategy for the design of robust HIV protease inhibitors would be to use the total interaction energy contributed by the HIV protease residues that are associated with primary mutations as a filter after the generation of potentially promising compounds using the DEE Drug Design software, with the highest preference given to those compounds which have the lowest total interaction energy with these residues. In this approach, equal importance would be given to all the residues that are vulnerable to primary

mutations irrespective of the propensity of occurrence of mutations at those positions. This methodology could be refined by assigning weights to the different residues depending on the frequency of occurrence of primary mutations at those positions (based on the information available in the HIV Drug Resistance Database). However, these approaches don't take into account the distribution of the interaction energy among the various residues that are associated with primary mutations. A useful strategy for doing so would be to assign a threshold value for the interaction energy (of the inhibitor) with each of these residues and use that as a filter in drug design. This threshold value could be determined based on the interactions of substrates and known broadly binding inhibitors with the HIV protease residues that are prone to primary mutations. A conservative way of defining this threshold value for each residue that is associated with primary mutations would be to take the average of the interaction energies of the different substrates and robust inhibitors (CA-p2, p2-NC, p1-p6, RT-RH, KB-83, AF-68 and AF-77 in this project) with that residue.

Conclusions

The fundamental challenge in combating rapidly mutating targets like HIV is the problem of drug resistance that results from mutations in a macromolecular drug target which allow it to maintain function while no longer being efficiently inhibited by drugs [27]. The substrate envelope hypothesis has proved reasonably successful in providing a rational approach for the design of promiscuous HIV protease inhibitors that have low susceptibility to drug resistance mutations. However, it has been observed that certain inhibitors that were designed by incorporating the substrate envelope constraints in the drug design methodology also suffered a significant loss of affinity when tested against a panel of drug resistant protease variants [24, 30].

This project involved examining the interactions of substrates and certain known robust and susceptible HIV protease inhibitors with the different residues of HIV protease in terms of binding energy and number of contacts less than 3 Å, 5 Å and 7 Å in order to identify the vital features which separate flat binders from susceptible binders. It was found that both the flat binders and

the non-flat binders interact with the same set of HIV protease residues, some of which are associated with primary mutations. The observation that even the flat binders have important interactions with certain HIV protease residues that are vulnerable to primary mutations suggests that in order to come up with robust HIV protease inhibitors, it is not necessary to completely eliminate interactions of the designed compounds with these residues (in fact, attempting to do so might result in inhibitors that have very weak binding affinity to HIV protease). It was observed that the total contribution to binding from the HIV protease residues that are associated with primary mutations (both in terms of binding energy and number of contacts) is a key feature separating robust inhibitors from susceptible inhibitors, with the non-flat binders interacting more strongly with these residues compared to the flat binders.

Based on the insights obtained from this work, certain strategies were proposed for the computational design of promiscuous HIV protease inhibitors. Future work on this project would involve the computational design of inhibitors using these approaches, synthesis of the top ranking compounds that emerge from these drug design methodologies and experimental testing of the synthesized compounds to study their inhibitory activity against the wild-type protease and different drug resistant protease variants.

It would be interesting to carry out a similar analysis for other rapidly mutating targets like HIV reverse transcriptase, hepatitis C virus (HCV) NS3/4A protease and hepatitis B virus (HBV) reverse transcriptase. This would help enhance the generality of this analysis so that it could be employed for drug design in the case of the outbreak of new epidemics of highly mutable infectious agents.

Bibliography

1. S. Vella and L. Palmisano. The global status of resistance to antiretroviral drugs. *Clinical Infectious Diseases*, 41:S239-S246, 2005.

2. J. E. Campo, C. Jamjian and C. Goulston. HIV antiretroviral drug resistance. *Journal of AIDS & Clinical Research*, S5:002, 2012.

3. P. S. Pennings. HIV drug resistance: problems and perspectives. *Infectious Disease Reports*, 5:21-25, 2013.

4. R. W. Shafer. Genotypic testing for human immunodeficiency virus type 1 drug resistance. *Clinical Microbiology Reviews*, 15:247-277, 2002.

5. A. Brik and C.-H. Wong. HIV-1 protease: mechanism and drug discovery. *Organic & Biomolecular Chemistry*, 1:5-14, 2003.

6. S. B. Chabria, S. Gupta and M. J. Kozal. Deep sequencing of HIV: clinical and research applications. *The Annual Review of Genomics and Human Genetics*, 15:295-325, 2014.

7. S. B. Laskey and R. F. Siliciano. A mechanistic theory to explain the efficacy of antiretroviral therapy. *Nature Reviews Microbiology*, 12:772-780, 2014.

8. J. Balzarini, M.-J. Camarasa, M.-J. Pérez-Pérez, A. San-Félix, S. Velázquez, C.-F. Perno, E. D. Clercq, J. N. Anderson and A. Karlsson. Exploitation of the low fidelity of human immunodeficiency virus type 1 (HIV-1) reverse transcriptase and the nucleotide composition bias in the HIV-1 genome to alter the drug resistance development of HIV. *Journal of Virology*, 75:5772-5777, 2001.

9. E. L. Asahchop, M. A. Wainberg, R. D. Sloan and C. L. Tremblay. Antiviral drug resistance and the need for development of new HIV-1 reverse transcriptase inhibitors. *Antimicrobial Agents and Chemotherapy*, 56:5000-5008, 2012.

10. F. Esposito, A. Corona and E. Tramontano. HIV-1 reverse transcriptase still remains a new drug target: structure, function, classical inhibitors and new inhibitors with innovative mechanisms of actions. *Molecular Biology International*, 2012:586401, 2012.

11. C. Hicks and R. M. Gulick. Raltegravir: the first HIV type 1 integrase inhibitor. *Clinical Infectious Diseases*, 48:931-939, 2009.

12. S. Rusconi and O. Viganò. New HIV protease inhibitors for drug resistant viruses. *Therapy*, 3:79-88, 2006.

13. J. Pokorná, L. Machala, P. Řezáčová and J. Konvalinka. Current and novel inhibitors of HIV protease. *Viruses*, 1:1209-1239, 2009.

14. Z. Lv, Y. Chu and Y. Wang. HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV/AIDS – Research and Palliative Care*, 7:95-104, 2015.

15. I. T. Weber, D. W. Kneller and A. Wong-Sam. Highly resistant HIV-1 proteases and strategies for their inhibition. *Future Medicinal Chemistry*, 7:1023-1038, 2015.

16. M. A. S. Perez, P. A. Fernandes and M. J. Ramos. Drug design: new inhibitors for HIV-1 protease based on nelfinavir as lead. *Journal of Molecular Graphics and Modelling*, 26:634-642, 2007.

17. Y. Mao. Dynamical basis for drug resistance of HIV-1 protease. *BMC Structural Biology*, 11:31, 2011.

18. I. T. Weber and J. Agniswamy. HIV-1 protease: structural perspectives on drug resistance. *Viruses*, 1:1110-1136, 2009.

19. B. R. Stranix, G. Sauvé, A. Bouzide, A. Coté, G. Sévigny and J. Yelle. Lysine sulfonamides as novel HIV-protease inhibitors: optimization of the Nepsilon-acyl-phenyl spacer. *Bioorganic & Medicinal Chemistry Letters*, 13:4289-4292, 2003.

20. T. Cihlar, G.-X He, X. Liu, J. M. Chen, M. Hatada, S. Swaminathan, M. J. McDermott, Z.-Y Yang, A. S. Mulato, X. Chen, S. A. Leavitt, K. M. Stray and W. A. Lee. Suppression of HIV-1 protease inhibitor resistance by phosphonate-mediated solvent anchoring. *Journal of Molecular Biology*, 363:635-647, 2006.

21. A. K. Ghosh, B. D. Chapsal, I. T. Weber and H. Mitsuya. Design of HIV protease inhibitors targeting protein backbone: an effective strategy for combating drug resistance. *Accounts of Chemical Research*, 41:78-86, 2008.

22. M. N. L. Nalam, A. Ali, G. S. K. K. Reddy, H. Cao, S. G. Anjum, M. D. Altman, N. K. Yilmaz, B. Tidor, T. M. Rana and C. A. Schiffer. Substrate envelope-designed potent HIV-1 protease inhibitors to avoid drug resistance. *Chemistry & Biology*, 20:1116-1124, 2013.

23. Y. Shen, M. D. Altman, A. Ali, M. N. L. Nalam, H. Cao, T. M. Rana, C. A. Schiffer and B. Tidor. Testing the substrate-envelope hypothesis with designed pairs of compounds. *ACS Chemical Biology*, 8:2433-2441, 2013.

24. M. D. Altman, A. Ali, G. S. K. K. Reddy, M. N. L. Nalam, S. G. Anjum, H. Cao, S. Chellappan, V. Kairys, M. X. Fernandes, M. K. Gilson, C. A. Schiffer, T. M. Rana and B. Tidor. HIV-1 protease inhibitors from inverse design in the substrate envelope exhibit subnanomolar binding to drug-resistant variants. *Journal of the American Chemical Society*, 130:6099-6113, 2008.

25. M. D. Altman. *Computational ligand design and analysis in protein complexes using inverse methods, combinatorial search and accurate solvation modeling*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2006.

26. S. Chellappan, G. S. K. K. Reddy, A. Ali, M. N. L. Nalam, S. G. Anjum, H. Cao, V. Kairys, M. X. Fernandes, M. D. Altman, B. Tidor, T. M. Rana, C. A. Schiffer and M. K. Gilson. Design of mutation-resistant HIV protease inhibitors with the substrate envelope hypothesis. *Chemical Biology & Drug Design*, 69:298-313, 2007.

27. M. N. L. Nalam, A. Ali, M. D. Altman, G. S. K. K. Reddy, S. Chellappan, V. Kairys, A. Özen, H. Cao, M. K. Gilson, B. Tidor, T. M. Rana and C. A. Schiffer. Evaluating the substrate-envelope hypothesis: structural analysis of novel HIV-1 protease inhibitors designed to be robust against drug resistance. *Journal of Virology*, 84:5368-5378, 2010.

28. M. L. Parai, D. J. Huggins, H. Cao, M. N. L. Nalam, A. Ali, C. A. Schiffer, B. Tidor and T. M. Rana. Design, synthesis, biological and structural evaluations of novel HIV-1 protease inhibitors to combat drug resistance. *Journal of Medicinal Chemistry*, 55:6328-6341, 2012.

29. V. Kairys, M. K. Gilson, V. Lather, C. A. Schiffer and M. X. Fernandes. Toward the design of mutation-resistant enzyme inhibitors: further evaluation of the substrate envelope hypothesis. *Chemical Biology & Drug Design*, 74:234-245, 2009.

30. Y. Shen, M. L. Radhakrishnan and B. Tidor. Molecular mechanisms and design principles for promiscuous inhibitors to avoid drug resistance: lessons learned from HIV-1 protease inhibition. *Proteins*, 83:351-372, 2015.

31. M. Prabu-Jeyabalan, E. Nalivaika and C. A. Schiffer. Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. *Structure*, 10:369-381, 2002.

32. R. R. Srinivas. *Computational engineering of small molecules to treat infectious diseases*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2017.