"Computational Studies of the HIV-1 Protease Dimer Interface."

by

Jerome J. Quintero

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biophysics)
in The University of Michigan
2011

Doctoral Committee:

Associate Professor Heather A. Carlson, Chair
Professor Hashim M. Al-Hashimi
Associate Professor Jason E. Gestwicki
Associate Professor Kristina I. Håkansson
Associate Professor Kevin J. Kubarych

2011

Acknowledgements

I would like to thank my two advisors, Professors Heather Carlson and Jason Gestwicki, for their advice and guidance throughout the course of my graduate career. I would like to thank Heather, specifically, for introducing me to the world of simulations, protein dynamics, and a microscopic universe where force field parameters are Gospel. Thank you for the patience to deepen my own appreciation for the finer details. I would like to thank Jason for allowing me to wander into his land of misfit toys and to tackle a decades-old problem in completely novel, and unobvious way.

I would like to thank the Biophysics Department for support. The support from the department personnel has helped me at every critical step along my training. I would also like to acknowledge the funding help provided from the Rackham Merit Fellowship and the Molecular Biophysics Training Grant. I would especially like to thank the Center for Women's Studies for providing me funding through an incredibly turbulent time in my life.

A warm thanks to all of my co-workers for it has been a pleasure to work with them. I will always appreciate them for their gracious support along the way, and their everyday companionship that made graduate school a lot more tolerable. I would like to thank all my friends I found along the way; whether it

was time spent in a class, time spent on cool Fall Saturdays cheering for the team, or the many hours spent enjoying a cool evening in a breezeway.

I would like to extend a deep appreciation of all the love and support I have received from my family, no matter the distance from home (I am finally finished Grandma). I want to especially acknowledge my gratitude for the all the love and tremendous amout of support from my fiancée, Thu Ha.

Table of Contents

List of Figures

List of Tables

Abstract

Computational Studies of the HIV-1 Protease Dimer Interface

by

Jerome J. Quintero

Chair: Heather A. Carlson

HIV-1 protease (HIVp) is one of four major drug targets to prevent propagation of the infectious HIV virion. Currently, all ten marketed HIVp drugs are inhibitors that target the HIVp active site. However, these drug therapies provide selective pressure resulting in mutations of the protease that escape drug efficacy. Consequently, the development of inhibitors of HIVp that have new modes of action is necessary. The dimer interface is an attractive target due to its highly conserved nature and its importance in forming an active enzyme.

Until now, all dissociative inhibitors were created by mimicking residues at the dimer interface, and resulted in several non-drug-like compounds. However, we created several receptor-based pharmacophore models of the HIVp

dimer interface, using ensembles of multiple protein structures (MPS). The MPS method was used to map the dimer interface with a series of small-molecule probes – methanol, ethane, and benzene. The maps were translated into pharmacophore models which were used to filter *in silico*, three-dimensional library of small molecules. The MPS method identified several novel small-molecule inhibitors capable of inhibiting dimerization, with several compounds characterized with less than 50 μM-level affinity. In the clinically relevant multi-drug resistant form of HIVp, these compounds maintained dissociative inhibition with nearly identical inhibition rates. Zhang-Poorman kinetic analysis verified the small molecules inhibit HIVp in a dissociative manner.

In addition to creating novel inhibitors, we modeled the protein-ligand interaction of known dissociative inhibitors using Langevin Dynamics. Ten, 10-ns simulations were initiated based on the hypothetical mechanism of ligand binding, but the dynamics simulations showed that the complex was unstable. Although the simulations did not result in a clear mechanism for protein-ligand binding, of the known dimerization inhibitors, they did demonstrate the entropic penalty of the proposed binding mechanism is unfavorable.

Finally, we propose to use hydrogen/deuterium exchange (HDX) - mass spectrometry techniques to obtain new structural information and further characterize the molecular recognition between HIVp and dimer inhibitors. HDX can provide the first structural evidence defining the mechanism of HIVp dissociative inhibition by small molecules. HDX could be broadly applicable for a range of active-site and allosteric inhibitors.

Chapter 1

Introduction

Structure Based Drug Design

Structure-Based Drug Design (SBDD) began with the "Lock-and-Key" hypothesis of protein-ligand binding, proposed by Emil Fischer in 1890.[1] Previous success depended on finding the correct "key", often a small molecule drug mimicking a natural ligand, for a particular "lock", the protein receptor surface. Finding the right key can involve arduous experimental verification of a library of compounds against a target, and often includes a structure-based analysis of the screened small molecules.[2] Although this empirical method reveals information about the properties of the small molecules tested against a protein target, the method does not provide any information about the protein binding pocket.

The lock-and-key hypothesis is significant, albeit simplistic. Fischer correctly recognized the importance of complementary shape between a ligand and the receptor pocket. Fischer also correctly identified the principle understanding that the ligand and a protein pocket should share complementary chemical properties: 1) increase favorable, cooperative electrostatics while minimizing unfavorable interactions (i.e. two positively charged functional groups near each other), and 2) maintain the hydrogen-bonding network between ligand and receptor. X-ray crystallography, NMR spectroscopy and MD

simulations all provide a far more detailed understanding of a protein receptor.[3-5] The information derived from a protein's structure provides a three-dimensional chemical illustration of a protein's surface, allowing the design of small molecules that complement the size and chemical properties of the binding site.[6] The pharmacophore illustration of the protein surface can then be used to filter a small molecule library to identify a smaller set of ligands that allegedly match the receptor and can be experimentally tested against a protein.[7] Several examples of early success in SBDD exist, such as dihydrofolate reductase and methotrexate, elastase with trifluoroacetyl-dipeptide-anilide analogues, and the influenza virus neuraminidase with sialidase-derived inhibitors but are a few example systems.[8-10]

Protein Flexibility

Because we know that the protein and interacting small molecule ligands are not static as previously believed, as in the lock-and-key model, an induced-fit hypothesis proposes an explanation for protein-substrate systems.[11] In the induced-fit hypothesis, a substrate can enter a protein's active site and induce a protein conformational change, often enveloping the ligand. Proteins in solution are naturally dynamic, with or without a ligand present. A protein natively exists as an ensemble of protein conformations, resembling a probability distribution of states.[12] At equilibrium, the population distribution of the conformational ensemble is distributed across several states, based upon how advantageous the free energy of one conformation compares to another. The

population distribution of unbound protein conformations is skewed upon ligand or drug binding to favor the new energetically favorable bound state.[13]

In this view of protein-ligand interactions, and given a diverse set of conformations, one can theoretically disrupt the normal population distribution of states by binding a ligand specifically tailored against a unique binding site of that state.[14] The protein-ligand complex would then stabilize the conformation, resulting in a new distribution of conformation population states. Regulation of enzyme activity through allostery will be discussed further in a later section, including regulation through protein-protein interfaces.

Protein motions are important in drug design. Conformational changes occur upon ligand binding and can be monitored from the femtosecond time-scale for bonded atom vibrations up to seconds for subdomain interactions.[15] Several methods can characterize a protein's motions, from smaller, faster motions (i.e. methyl motions, loop motions, etc.) to larger domain motions and interactions.[16] Ultrafast spectroscopic methods can provide detailed atomic vibrations on the femtosecond time-scale. NMR relaxation techniques provide local dynamic information, from the picosecond timescale to the larger domain motions taking place over seconds. X-ray crystallography, although static, provides the detailed atomic coordinates for computational simulations measuring protein motions in the femtosecond to microsecond range. Hydrogen-deuterium exchange experiments coupled with mass spectrometry can provide information on the millisecond to second time-scale. All of these techniques clarify protein motions to a system of interest.

Allostery

The mechanistic basis of allostery is poorly understood, but it can be described by three different models. The first model, the concerted MWC (proposed by Monod, Wyman and Changeux) model postulated that a conformation change, such as binding of a ligand by one enzyme subunit of a multimeric enzyme, will necessarily induce a conformational change to all other existing subunits. The second model, the sequential KNF model (proposed by Linus Pauling and favored by Koshland, Nemethy and Filmer) states that the conformational change from one subunit, upon binding a ligand, is not necessarily propagated to all other subunits in a multimeric enzyme.[17] The third model is an entropy-driven model, stating that entropy is the force for allosteric change.

A classic case corresponding to the MWC model is the oxygen-transport protein hemoglobin.[18] Hemoglobin (Hb) exists in a multi-subunit quaternary structure containing four heme groups with an coordinated iron in the center to bind oxygen for transport to the cells. Hb has two distinct states: the R (relaxed) state and the T (taut) state.[19] Upon binding of two oxygen molecules to the higher-affinity R state, the ligand induces a conformation change and Hb converts from the R to T state.[20] At the T state, the two remaining available heme groups can bind additional oxygen molecules, but at a lower affinity towards the oxygen molecule.

In KNF model, the substrate binding event in one subunit increases affinity for ligands in adjacent subunits, similar to the MWC model; however, conformational

change in one subunit is independent of other subunits.[21] Allostery generally involves the binding of a molecule, often called a modulator or effector molecule, on a protein surface that is not the active site but can regulate, either positively or negatively, an enzyme's catalytic activity. Binding of the effector molecule can be reversible, as in the case of aspartate carbamoyltransferase, but there are examples of non-reversible effector molecules covalently attaching to the protein surface.[22] An example this can be observed in the covalent modification of phosphoryl groups to several residues, or by modifications to cysteine residues.[23, 24]

Some have argued for the role of entropy as the driving force for allosteric changes. Cooper and Dryden claim that every protein has a set of native motions and dynamics, including the frequency and amplitude of motions.[25] Upon binding a ligand, the motions in the bound subunit are altered, causing a shift in the properties of motions in other subunits. Popovych *et al.* demonstrated the first case of dynamically driven protein allostery characterizing the conformational states of the dimer CAP protein upon binding of cAMP ligand to the CAP dimer.[26] One model of allostery probably does not cover all examples of protein allostery; however, these three categories above can describe most known allosteric regulation mechanisms.

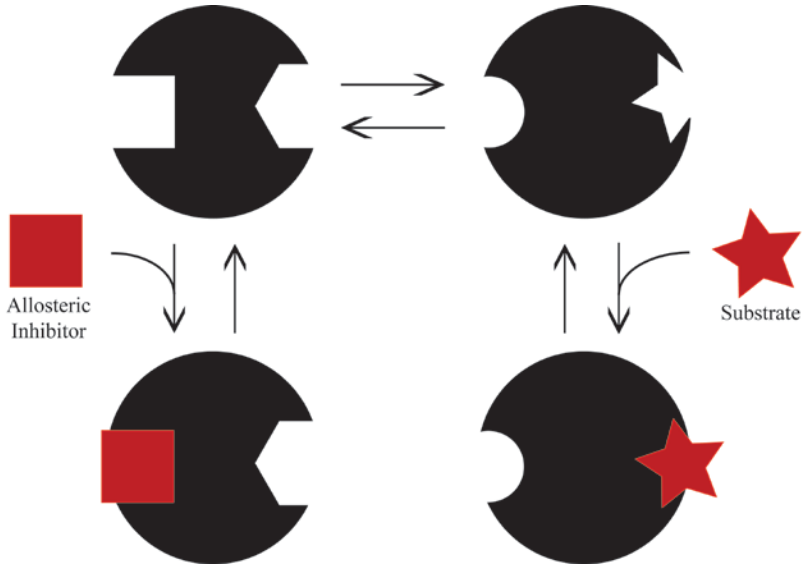

Figure 1.1: Example of Negative Allosteric Regulation. From the unique protein conformation ensemble perspective, an enzyme's catalytic activity can be regulated through an allosteric mechanism. Binding an allosteric inhibitor to this unique conformational state negatively affects the affinity of the substrate to the enzyme active-site.

Our understanding of protein allostery has grown, allowing allostery to be engineered into a protein, as in the case of trypsin specificity.[27] Although allostery was originally based upon observed conformational changes in symmetric proteins, such as hemoglobin or aspartate carbamoyltransferase, the term has been broadened to encompass any ligand binding event that affects another region's ability to bind a ligand, positively or negatively, for either a multimeric or monomeric protein. Nussinov and colleagues have argued that every enzyme can be regulated in an allosteric fashion.[28] Since a protein exists as an ensemble of conformations, flexible protein, with its unique conformation ensemble, can be inhibited through ligand binding against a specific state, locking the protein into one conformation.[29] Figure 1.1 above is an example of negative regulation of a protein's catalytic activity. An effector molecule, the allosteric inhibitor, binds within a region outside of the active-site and reduces substrate affinity to the active site.

Protein-Protein Interactions

The human interactome has an estimated 650,000 protein-protein interactions (PPIs).[30] The sheer size of the interactome indicates a favorable probability that a number of PPIs can be targeted with small molecules. As an example, the Human Immunodeficiency Virus (HIV) exploits the use of host cellular proteins – often called HIV-dependent factors (HDFs) - to reproduce HIV particles. Brass *et al.* screened a library of small interfering RNA (siRNA) to find candidate HDFs for possible drug targets.[31] Brass identified 250 candidate PPIs involved in the life cycle of HIV as candidate targets for future drug design. There are several examples of designed allosteric regulation of protein activity by targeted disruption of a protein-protein interaction.[32-35] The p53-MDM2 PPI is a target of interest for cancer patients.[36] The release of p53 (a protein that regulates the cell cycle in response to cellular stress leading to DNA damage) from MDM2 would aid in the restoration of p53 function in tumor suppression *via* apoptosis. Maraviroc is FDA approved for treatment of HIV infections by blocking the HIV gp120 from binding to a host cell's CCR5 receptor protein, a step needed for virus particle entry into the host CD4 expressing cells.[37]

The average surface area for a protein-protein interaction is 1600 $Å^2$, but can range up to 4000 $Å^2$.[38] That compares to range for an active site protein-ligand interaction at 300 to 1000 $Å^2$.[39, 40] The average weight for small molecules disrupting a protein-protein interaction is approximately 680 MW. The higher molecular weight for PPI inhibitors is a concern with regards to Lipinski's Rule of Five, a set of guidelines

describing the general physical properties of orally bioactive drugs.[41] Lipinski's Rule of Five states that for a drug to be orally bioactive, a drug must be 1) < 500 Da in molecular weight, 2) have ≤ 5 hydrogen -bond donors, 3) ≤ 10 hydrogen -bond acceptors, and 4) have a partition coefficient (or log P) < 5. There are exceptions to Lipinski's Rule of Five, of course, as seen with Navitoclax. Navitoclax, targeting the Bcl-X_L--BAD PPI mediating apoptosis, breaks all but one of Lipinski's rules and is currently in clinical trials.[42, 43]

Ligand efficiency is a metric used to compare a ligand's binding free energy relative to its size – defined specifically as the free energy of binding by the ligand (ΔG_{bound}, kcal/mol units) over the number of non-hydrogen atoms – and can help indicate the druggability of a targeted region.[44] Despite higher molecular weight, ligand efficiencies for ligands targeting PPIs remain comparable, albeit less, than the ligand efficiency of active site inhibitors, such as protein-kinase inhibitors or protease inhibitors.[33, 34] The ligand efficiencies for PPIs can be explained by the observation, that PPIs regularly contain relatively small patches on the protein surface responsible for high affinity binding, commonly referred to as protein "hot spots".[45]

Whether designing a small molecule to target a protein's active site or an allosteric regulatory site to perturb an enzyme's kinetics, SBDD is an iterative process of analysis, design, synthesis and experimental verification that leads to an optimized relationship between a protein surface receptor and its complementary, unnatural ligand. There are a growing number of computational techniques being developed to speed the cycle of SBDD and to expedite the process to bring new drugs to market.

Molecular Docking

Molecular docking programs, which are computational programs that allow broad ease of access for common users and the most generally used, simulate protein-ligand interaction. There are several docking programs available, such as DOCK, AUTODOCK, GOLD, GLIDE, FLEXX among others.[46-51] The programs use varying parameters to search for the best fit between the receptor and the ligand. Parameters can include complementarity of physical interactions, such as the shape between the receptor and ligand with regards to the Van der Waals (VDW) radius and atomic charge, and intermolecular interactions such as hydrogen bonds or hydrophobicity.[52] The ligand poses generated from molecular docking simulations are then scored based on the parameters, ranking a predictive pose's affinity between a receptor and ligand. Receptor flexibility has been increasingly incorporated to help integrate motions occurring in ligand-binding events.

Molecular Mechanics

Molecular mechanics (MM) is the simulation of molecular system using classical Newtonian mechanics. Quantum mechanics (QM) is more accurate than MM and may also be employed; however, QM comes a penalty, as computational hours rapidly grow proportional to the number of atom simulations, thus limiting the number of atoms which may be simulated to computational resources.[53] The atom limit on QM simulations designates Newtonian mechanics as better suited for larger molecular systems, such as a

solvated protein. There are a number of MM programs including Chemistry at HAvard Molecular Mechanics (CHARMM), Assisted Model Building with Energy Refinement (AMBER), GROningen MOlecular Simulation (GROMOS), and Not (just) Another Molecular Dynamics (NAMD) program.[54-57]

Force-field parameters were developed to approximate the Newtonian mechanics the atomic forces experience within a molecular system. The energy within a system is typically calculated as the summation of several atomic forces: 1) bonded atomic forces, such as bonded atom vibrations; 2) bonded angles amongst atoms and dihedrals; and 3) non-bonded forces, such as electrostatic and VDW terms. The current AMBER force field is shown in the following equation:[58]

$$V(r) = \sum_{bonds} K_b(b - b_o)^2 + \sum_{angles} K_\theta(\theta - \theta_o)^2 + \sum_{dihedrals} (V_n/2)(1 + cos[n\phi - \delta]) + \sum_{nonbij} \left[\left(A_{ij}/r_{ij}^{12}\right) - \left(B_{ij}/r_{ij}^{6}\right) + \left(q_i q_j/\varepsilon r_{ij}\right)\right]$$

In AMBER's force field implementation, the first three summation terms are the bonded terms. Bond stretching is noted in the first term, where K_b is the stretch constant between two atoms, and b and b_o are the bond length and equilibrium bond length at equilibrium, respectively. Angles are shown in the second term, K_θ is the angle constant, and θ and θ_o are the bonded angle and the bonded angle at equilibrium. Dihedrals are summed in the next term, here V_n is the dihedral constant, and ϕ and δ are the dihedral and the dihedral phase angle, respectively. The non-bonded terms, representing the VDW modeled with a

12-6 Lennard-Jones potential and the charge-charge interactions are each calculated as point-charge interactions. The 12-6 Lennard-Jones potential accounts for both the short range repulsive interactions. The Pauli repulsion, indicated by the $r^{(-12)}$ term, describes favorable or unfavorable ionic interactions. The attractive long range terms, such as VDW, are accounted in the $r^{(-6)}$ term. Interactions between two point charges (q_i and q_j) are calculated as a Coulombic potential so that r_{ij} is the radius between the two point charges and ε is the dielectric influence.

Force-field testing optimizes MM calculations, reflecting empirical measurements. In 1995, Cornell *et al.* published the AMBER94 Force Field serves as the basis for AMBER simulations for over a decade.[59] There have been several iterations of the AMBER94 Force Field, including refinement of parameters (such as AMBER99 Stony Brook, or AMBER99SB), additions in the force field to include for fluctuations in polarizability within residues, force field parameters for sugar molecules, or to include Generalized AMBER Force Field (GAFF) parameters for small molecules.[60-64]

Molecular Dynamics

Molecular Dynamics (MD) is a technique combining the MM force fields and Newton's Second Law, $F = m * a$, to study large biomolecular systems along the course of a trajectory. The trajectory length relies on available computational resources and the size of the biomolecular system; it resembles QM in this regard, although with less of a penalty. Efforts have been made to parallelize and accelerate the calculation to

allow trajectories into the microsecond range. For our studies, the AMBER package was chosen as the MM software package.

Although there is not one general, explicit method, there are general principles and protocols ubiquitous to system setups. Coordinates for initial atom position are typically provided from experimental X-ray or NMR sources. Parameters for atoms are then assigned dependent upon force field (AMBER99SB for our studies), and atom initial velocities are assigned by a Maxwellian distribution determined from the user defined seed number. Typically, MD equilibration is a series of steps intended to relax the biomolecular system and, depending on size of system, can last into the production portion of an MD trajectory. Proper equilibration of a system has been proven important to a system, as equilibration overcomes secondary effects of crystallization (i.e. crystal packing effect causes improper bond distances or dihedral). Crystal packing effects are particularly troublesome, as they may result in an unconventional structural pose or bias the protein conformational ensemble towards one state over a more energetically favorable conformation.[65] Water molecules are added to a system explicitly in MD simulations in a crystalline form and must be equilibrated to ensure complete solvent contact along the solvent accessible surface area (SASA) of the biomolecular system.[66] TIP3P and TIP4P are common parameters used for waters and are treated as a rigid molecule to simplify the calculation.[67, 68]

After equilibration, the MD trajectory continues into the production phase (also simply referred to as the "trajectory") of the study. Structural and energetic information, such as atom coordinates, velocities, energies and temperature, can be outputted from the trajectory at user-specified intervals. Analysis of the trajectory is guided by the

information the user seeks. Program packages, similar to the Multiscale Modeling Tools for Structural Biology (MMTSB) and AMBER's ptraj analysis program, are available to assist in data analysis.[55, 69] One common analysis, a root mean square difference (RMSD), can be used to quantitatively measure protein fluctuations compared to a reference frame, often the start of the production run or an averaged position within a portion of the trajectory. Pertinent to studies of protein-ligand interactions, a ligand's contact information onto a protein surface can be deduced from the MMTSB's contact script. The MMTSB contact script defines a "contact" as the minimum inter-residue distance between every heavy atom pair less than 4.2 Å. The Dictionary of Secondary Structure Program (DSSP) assigns secondary structure (i.e. β-sheet, α-helix) on a per residue basis depending on the geometric similarities of the residue's hydrogen-bonding network to known secondary structures.[70] Normal Mode Analysis, Correlated Dynamics, bond angles (i.e. phi- / psi-angles for Ramachandran Plots) can all be derived from MD trajectories based upon the information needs of the user.

Implicit Solvation

Implicit solvation is a sampling technique used to reduce the number of atoms in a system, and therefore reduce the computational demand for its model. An implicit solvation simulation models the solvent as a continuous medium used to approximate the solvent's mean free energy influence in a solute-solvent system as calculated by the Poisson-Boltzmann (PB) equation, instead of modeling the solvent explicitly with individual water molecules.[71] The generalized Born (GB) is an approximation of PB

making GB more generally applicable due to its computational efficiency.[72] Although the medium is continuous, its properties can be changed (i.e. the dielectric within a lipid bilayer). Implicit solvation is considered a sampling technique, since the system of interest can sample more protein conformational ensembles than a standard, explicit solvent MD trajectory can produce within an equivalent amount of computational effort.[73] Since this is an approximation of the solvent's free energy, extensive testing of the continuous medium must be performed. For our studies, GB(II)SA-OBC was the preferred solvent parameter.[74, 75]

Langevin Dynamics (LD) simulations are simulations, often implicitly solvated but which can be explicit, with the Langevin equation specifically applied. The Langevin equation is:

$$M\ddot{X} = -\nabla U(X) - \gamma M\dot{X} + \sqrt{2\gamma k_b TMR(t)}$$

Here, U(X) is the potential of an atom interacting with the solvent medium. ∇ is the gradient of the potential giving -∇U(X) the force of the potential. $\dot{X}$ and $\ddot{X}$ are the first and second derivatives, respectively, of atoms at a given position X at a time-point in the trajectory (X_t); these give the velocity and acceleration of those atoms, for each. M is the mass of the atom, k_b is the Boltzmann constant, T is temperature and R(t) is the Gaussian Process. The Langevin equation can be applied to hold the temperature of a system constant, but for implicit solvation, the benefit of the Langevin equation is given in γ, a damping constant. The damping constant gives the continuous medium a viscosity term, so the atoms will "collide" similar to a solvated environment and not in a gas-phase environment.

Monte Carlo Method

The Metropolis Monte Carlo (MC) method for molecular mechanics is a sampling method based upon the Monte Carlo algorithm that relies on a stochastic approach to produce its designed result.[76] For MD, the designed result is the movement of atoms. Although it is a force field-based method similar to MD, the MC method transposes a small number of atoms in a time-step, unlike single atom displacements in MD, to a new state (or geometric position). MC then calculates the potential energy at the new state. The decision to accept, or reject, the new state is decided upon a Boltzmann distribution of the potential energy from the new position and is compared against a randomly generated number in the range of 0 to 1.

MC is a markov chain process; wherein exists a finite number of states and transitions from one state (i) to the next (i+1) is not influenced by the previous (i) state. The MC method is an iteration of several steps: 1) specify initial atom coordinates, 2) transpose by random displacement, 3) calculate differences in potential energy, 4) determine if the change in potential energy is less than 0, then go to step 2, otherwise 5) if the Boltzmann distribution of the potential energy is less than a randomly generated number, the new coordinates are accepted and proceeds back to step 2 (if greater or equal to the randomly generated number, original coordinates are retained and proceeds to step 2). The number of MC iterations is user defined.

The Biochemical and Organic Simulation System (BOSS) is a package of MM force field based programs including the MC method.[77] The OPLS (Optimized Potentials

for Liquid Simulations) parameters are used to simulate common solvents.[78] The potential energy formula is similar to that of AMBER. The BOSS program package is used by our laboratory to minimize solvent molecules along a protein surface.

Protein Surface Mapping

One technique to map the chemical surface of a protein for SBDD is to saturate crystallized proteins in a chemical fragment solvent prior to X-ray crystallography.[79] The crystal soaking technique has a fairly high throughput capacity and has a high resolution to spatially differentiate the chemical properties along a protein surface.[80] In crystal soaking, a previously crystallized protein in a unbound (apo) state is saturated in a series of solutions composed of a chemical fragment that represents a physical steric constraint (i.e. aromatic), or a physiochemical property of interest (i.e. hydrophobicity, hydrogen-bonding). The chemical fragments affix themselves in regions along the protein surface that are energetically and conformationally favorable to the fragment's chemical property. The fragment map of the protein surface can then be used to build ligands with features comparable to the targeted receptor. Chemical fragment crystal soaking has been used for numerous targets; however, crystal soaking has its own disadvantages. Crystals are often sensitive to the solvents of interest, which leads to poor diffraction patterns. In addition to the sensitivity of the crystal lattice to the chosen solvent, the crystal soaked map of the protein surface still only offers a static picture of the protein and is not representative of the protein motions. Lastly, electron density on protein

surface is ambiguous and can be indeterminate whether the density is for water or a chemical fragment.

There are computational techniques that mimic the crystallographic soaking method, such as the Multiple Copy Simultaneous Search (MCSS) method developed by Karplus.[81] This method places multiple copies (on the magnitude of hundreds to thousands) of a chemical fragment near the targeted protein region. Each fragment is permitted to interact with the protein surface, but not with other fragments. The fragment probes are then minimized to the protein surface, converging to positions both energetically and conformationally favorable for the chemical fragment. The protein itself is represented by either a static structure or a small degree of dynamic information can be illustrated using a restrained MD simulation. The MCSS method can result in too many energy minima positions, in which case arbitrary energetic cut-offs are often used to filter the minimized probes, to retain the more energetically favorable probes.[82]

Multiple Protein Structures

The multiple protein structures (MPS) method is an evolution of the MCSS method to computationally map protein surfaces, but MPS was conceptualized to incorporate greater conformational sampling of the flexible protein target. A study by Carlson *et al.* pioneered the MPS method against the apo Human Immunodeficiency Virus type 1 (HIV-1) integrase protein.[83] The method was refined by Meagher, *et al.* and made robust over several protein systems.[84]

The first step of MPS includes mapping several protein conformations' with chemical fragments. Several chemical probes are individually mapped to each static structure. Methanol, ethane, and benzene fragments flood the target region and represent hydrogen-bonding, hydrophobic and aromatic interactions, respectively. The probes are minimized to the protein surface by the Multi-Unit Simulation of Interacting Chemicals (MUSIC) program, which is part of the BOSS programming package. MUSIC performs a low-temperature MC minimization that allows the probes to search for energetically favorable geometric positions along the protein surface. The mapped conformational structures can be derived from several sources: NMR, MD or X-ray crystallography. Although one could use any initial atomic coordinate source, NMR has demonstrably providing the most conformational sampling, followed by MD, and then crystallographic sources.[85]

After probe minimization along the protein surface, clusters of probes are observed. A single probe, the lowest energy minimum within the cluster, is used to represent the cluster (called a parent probe). Each protein structure, with minimized parent probes, is then overlaid to a single structure. Protein structure alignment produces new clusters of parent probes. The clustered parent probes are then compared across the multiple conformations to identify chemical and spatial consensus across multiple conformations. The consensus cluster defines characteristics of the pharmacophore model and represents the average position and RMSD of the clustered parent probes.

The MPS pharmacophore model represents the chemical and spatial complementary a small molecule must have to bind against the protein pocket. The pharmacophore can filter, *in silico*, a library of small molecule conformations based upon

the properties of the pharmacophore model. MPS provides enrichment in successful hits superior to experimental chemical library screening methods. The MPS technique has demonstrated experimental success for filtering small molecules able to inhibit its intended target. Refinement of the MPS method was required to create a robust method over several protein systems.

The MPS method was validated through demonstration, selectively discriminating small molecule inhibitors of HIV-1 protease (HIVp), dihydrofolate reductase (DHFR) and murine double mutant (MDM2).[86-88] In the case of DHFR, the MPS method selectively filtered species-specific inhibitors, discriminating high-affinity DHFR inhibitors over more general, lower-affinity inhibitors. Incorporating longer DHFR MD trajectories, and therefore more conformational sampling, produced better-performing pharmacophore models, which alludes to the importance of protein flexibility in SBDD. The MPS technique identified five novel small molecules, providing new diverse chemical scaffolds, targeting the p53-MDM2 protein-protein interaction. The MPS procedure was also performed to target the allosteric "eye" site on HIVp.[89]

HIV-1 Protease as a Drug Target

Human Immunodeficiency Virus (HIV) was first diagnosed in the United States in 1981, and today, we have a cocktail of drugs – known as Highly Active Antiretroviral Therapy, or HAART – that target four HIV protein targets: protease, integrase, reverse transcriptase, and the gp120-CD4 protein-protein interaction.[90] These four targets are necessary for the maturation and propagation of the active HIV particle. Genomic viral replication has a low fidelity rate, creating a high error rate in viral replication; whereas

mutant HIV proteins allow the variants to escape drug efficacy. The effect of mutations is reflected in high rates of drug resistance to nearly 20% of marketed HIV inhibitors, and a drug-resistant HIV transmission rate near 9%.[91] The 2008 UNAIDS reports an approximate 33 million HIV infected patients, resulting in 2 million deaths and a number of individuals harboring HIV. Therefore, a strong need for development of new therapeutics to attack HIV in a novel fashion is crucial.[92]

The HIV genome is stored on a single double stranded RNA and is split into three major polyproteins: the Gag, Pol and Env polyproteins. The Gag and Env polyproteins primarily encode structural proteins necessary for the viral particle. The Gag polyprotein splits into four major domains that contain critical viral proteins: the matrix, capsid, nucleocapsid, and the p6 protein.[93] The matrix domain contains the Gag protein, when myristoylated, binds to the host cell membrane.[94, 95] The myristoylated, membrane-bound Gag protein then creates neighboring Gag-Gag protein-protein interactions. These PPIs create the viral particle budding. The Env glycoproteins, such as gp120 and gp41, are recruited to the matrix to provide the budding viral particle its envelope proteins. [96] The Gag nucleocapsid domain contains proteins that bind to replicated double-stranded viral RNA, which anchors the RNA to the viral matrix. The capsid provides an inner compartment for viral proteins (protease, integrase, and reverse transcriptase) necessary for function. Along with six other accessory proteins, such as Nef and Vif, which have roles at different stages of viral replication, the HIV particle can be reduced to a molecular entity of 15 total proteins and RNA.[97]

Upon release from the cell, the immature HIV particle is proteolyzed by HIV protease (HIVp) in a series of events, transforming it into an infectious viral particle.[98]

This series of proteolytic events is critical to the HIV life cycle, making regulation of HIVp catalysis a viable therapeutic target. Down regulation of HIVp catalysis is achieved by preventing the natural substrate from entering the active site. This is the traditional method to regulate HIVp activity, thereby halting HIV particle maturation.[99, 100]

Competitive Inhibitors of HIV-1 Protease

The first HIVp structure was characterized by Navia and colleagues, deposited in the Protein Data Bank (PDB) with the identification (ID) of 2HVP.[101] HIVp is a homodimer with C_2-symmtry, 99 amino acids, and shares a similar fold to other characterized apartyl proteases.[102] As of publication of this dissertation, there are hundreds of structures in the PDB of unbound HIVp, ligands complexed with HIVp, mutant variants of HIVp, and combinations of these categories. The repository includes a tethered form of HIVp with several Gly residues connecting the two monomer units through the N- and C-termini.[103] PDB 4HVP is the first structure to characterize the relationship between HIVp bound to an inhibitor, which blocks the HIVp active site from substrate catalysis. Since 4HVP publication, 10 different drugs are U.S. Food and Drug Administration (FDA) approved and released to market to target the inhibition of the HIVp proteolysis *via* direct competition of the active site over its natural peptide substrate.[104]

HIVp performs proteolytic cleavage along the amino acid backbone through a general acid (GA)/ general base (GB) reaction involving two catalytic Asp25 (and

Asp25') residues and a shared water molecule, "cutting" the peptide substrate into two. The reaction mechanism empirically described by Bihani and colleagues uses crystallographic techniques to trap two short peptides complexed with HIVp after cleavage.[105, 106] PDB ID 3DOX describes the role of the water molecule in the GA/GB reaction between the HIVp catalytic Asp residues and the cleaved substrate. Although this is the first empirical evidence of the GA/GB reaction, Hosur and colleagues were not the first to propose the GA/GB mechanism. Several groups proposed a GA/GB mechanism in lieu of a direct nucleophilic catalytic mechanism, applying computational methods to argue the GA/GB reaction is more energetically favorable over the direct nucleophilic mechanism.[107, 108] The cyrstallographic snapshot provided in 3DOX is conclusive empirical evidence of the GA/GB reaction mechanism for HIVp proteolysis.

Currently, the 10 drugs on the market that target the active site of HIVp inhibit catalysis through direct competition by mimicking the natural substrate. With the exception of Tipranavir, all HIVp drugs are peptidomimetic derivatives of the natural substrate. Peptidomimetics have poor pharmacokinetic properties and toxic side-effects.[109] In addition, all 10 of the HIVp drugs are susceptible to degradation by cytochrome P_{450} enzymes, particularly CYP3A4.[110] Often, Retinovir is coadministered with another HIVp inhibitors to reduce metabolism, but coadministration of two peptidomimetics exacerbates the toxicity and side effects to maintain a high drug concentration, preventing HIVp catalysis.[111] Marinec *et al.* improved the pharmacokinetic properties of HIVp drugs by covalently attaching the natural ligand FK506-binding protein (FKBP) to Amprenavir, an FDA approved HIVp drug.[112] Bifunctionalizing Amprenavir with FK506 permitted greater partitioning of the

bifunctional drug into the cell and reduces drug metabolism. Improvements to the pharmacokinetic properties of HIVp drugs allow lower dosing, thereby reducing side effects of HIVp drugs.

HIVp escapes drug efficacy through selective pressure caused by the active site inhibitor, evolving the protease. Mutations in HIVp frequently lie outside the active site and do not interact with it, nor do they affect ligand binding directly. The mutations are often conservative mutations, so that a mutation at a single specific residue will not significantly alter the chemical properties of that residue (i.e., Ile15Val). Conservative mutation(s) results within the HIVp core a phenomenon termed "hydrophobic sliding," described by Celia Schiffer.[113] A small hydrophobic residue within the core mutates to another small hydrophobic residue, so as not to generate steric clashes between residues within the core. A steric clash within the HIVp core would physically constrain the core to transition between the different states (i.e., open to closed). Although mutations do not directly affect ligand binding, they do alter the shape of the receptor pocket as the protease transitions from one state to another, indirectly affecting the complementary between the ligand and receptor.

First-generation of HIVp inhibitors, such as Saquinavir, Ritonavir and Indinavir, were designed with strong entropic consideration to the HIVp active site, even at the expense of unfavorable enthalpic interactions. Second-generation HIVp inhibitors, such as Lopinavir, Atazanavir and Darunavir, posses comparable, or slightly less favorable, entropic interactions, but a stronger emphasis on increased enthalpic interactions when bound to HIVp. The different binding mechanisms between the two generations were investigated by Freire and colleagues using Isothermal Titration Calorimetry (ITC). The

second-generation inhibitors increased the enthalpic contribution by creating a hydrogen-bonding network with the peptide backbone, as opposed to hydrogen-bonding to amino acid side chains. KNI-764, a second-generation inhibitor precursor, revealed another important compound feature, which is that it must be flexible at crucial positions to account for new motions that occur due to mutations. Although the compounds KNI-764 and KNI-262 are chemically similar and have comparable binding affinities against the wild-type HIVp, the affinity of KNI-262 reduces against the mutant V82F/I84V HIVp mutant due to unfavorable flexibility within the compound. Freire emphasizes that for certain targets, such as wild-type HIVp and variants, it is important for the compound to be adaptively flexible to account for new receptor pocket motions caused by mutants.

Allosteric Regulation of HIV-1 Protease

Overcoming the loss of drug efficacy caused by mutant strains of HIVp is a desirable goal in the development of future therapeutics. Researchers are searching for and manipulating allosteric regions in HIVp to inhibit multi-drug resistant strains of HIVp. Perryman *et al.* used molecular constraints in MD simulations and HIVp as an example of allosteric control.[114] Perryman began with a bound, closed HIVp structure for both the wild-type and mutant HIVp (PDB 1KZK and 1D4S, respectively) and then deleted the enclosed KNI-764 ligand in both structures. Perryman continued to place a distance (7.7 and 10.5 Å) force restraint on the alpha-carbon atoms between residues Gly40 and Gln61 of HIVp; prior to performing MD simulations. In the restrained MD simulations, Perryman observed that the flaps never transitioned to the open

conformation. In the larger distant restraint (10.5 Å), the flaps failed to return to the semi-open conformation despite lacking a bound ligand to stabilize the closed state. The restrainted simulations suggested a means of allosteric control of flap dynamics by restraining the cheeks through the "allosteric groove" between residues Gly40 and Gln61.

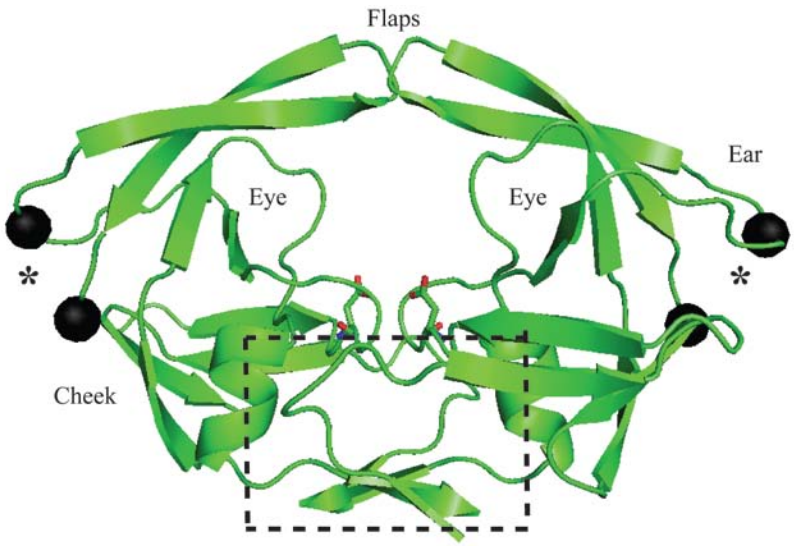

Figure 1.2: HIV-1 protease topology. HIVp is a 99 amino acid homodimer with C2 symmetry. Common names for regions in the HIVp are indicated as defined by Perryman, *et al.* Black spheres indicate the alpha-carbons for Gly40 and Gln61, which distance restraints were placed for MD simulations. Asterisks (*) describe the location of the allosteric groove between Gly40 and Gln61. The dashed box indicates the dimer interface of HIVp, defined as the region bordered by the two α-helices, the β-sheet, and the backbone of the active-site. The two catalytic aspartic acids are shown as sticks, image generated *via* PyMol.

The X-ray crystallographic structure PDB 1TW7 demonstrated an empirical representation of putative allosteric control *via* the allosteric groove Perryman described. 1TW7 showed the HIVp flaps in a novel conformation, terming the conformation as "wide-open" due to the novel and skewed position of the flaps.[115] Layten *et al.* deleted the wedged flap tip and performed MD simulations of the skewed conformation of 1TW7.[116] Their simulations demonstrate that 1TW7 quickly reverts to a semi-open

conformation without the wedge flap tip; whereas the skewed 1TW7 is stabilized by the crystal packing effect. Although Layten's simulations show 1TW7 was stabilized by crystal packing effects, they do not necessarily prove that the flap tip can bind in the allosteric groove. Studies by Lexa, *et al.*, showed a peptidomimetic based upon the wedged flap tip residues could not maintain contact over the course of their MD simulations, and therefore suggesting that flap tip stability to the allosteric groove is unfavorable.[117]

Other regions outside the HIVp allosteric groove are of interest to researchers, in their efforts to prevent HIVp proteolytic turnover. In 2008, Damm *et al.* utilized the MPS method to discover a small molecule with properties that match the mapped complementary space of the eye site, and the small molecule was determined to inhibit HIVp catalysis.[89] Damm performed MD simulations of the eye small molecule complexed to HIVp to demonstrate the mechanism of protein-ligand binding; however, there is no current structural evidence confirming the ligand's binding mechanism. Although there is no structural evidence of Damm's bound eye site ligand, a growing amount of corroborating evidence supports the site's ability to bind a ligand.

Concurrent with Damm, Böttcher *et al.* co-crystallized a bound HIVp with a pyrrolidine-derived inhibitor (PDB 3BC4).[118] Unique to 3BC4, the ligand stabilizes the open HIVp conformation by positioning the naphthyl rings moieties of the ligand underneath the flap near the eye site. PDB 3BC4 crystallized the pyrrolidine ligand bound in two conformations (α and β). The authors speculate the β conformation is an artifact of crystallization, stabilized by neighboring crystal contacts. Lexa and Carlson performed MD simulations of HIVp bound with the ligand in an α, β and α + β

conformation.[119] Their simulations demonstrated that the α ligand conformation of 3BC4 is the most stable of the three. The simulations reported a ligand conformational change in the trajectory. While one of the ligand's naphthyl ring stayed bound to the eye site, the other naphthyl ring flips position to bind into the S1/S1' or to the S2/S2' pocket.

The favorability of this pyrrolidine derivative conformation, with one naphthyl ring in the eye site and the other in the active-site pocket, corroborates EPR studies of the HIVp performed by Torbeev *et al.* Torbeev measured the distance between the MTSL spin-labeled residues at position 55 and 55', and compared the distance to the equivalent residue positions in MD simulations.[120] The result of Torbeev's EPR work reported HIVp is asymmetric when bound by the three traditional inhibitors (MVT-101, KVS-1, JG-365). The HIVp inter-residue distance supports a "semi-open/closed" flap conformation, unlike the "closed/closed" flap conformations seen in bound HIVp crystal structures. The asymmetric "semi-open/closed" flap conformation would allow the eye site to exist, asymmetrically, in a ligand-bound HIVp. Perryman *et al.* soaked crystal of HIVp with several chemical fragments. [121] One of the fragments minimized to HIVp, a 5-nitroindole fragment, was found to be stably bound at the eye site corroborating the existence of the allosteric site. The fragment was bound to only one monomer of the symmetric HIVp dimer. Asymmetric binding by the chemical fragment further supports the asymmetric binding requirement for this mode of inhibition.

HIV-1 Protease Dimer Interface

HIVp monomers must dimerize to allow catalytic activity, which distinguishes it from other aspartic proteases. Because HIVp is an obligatory dimer, and due to the highly conserved nature of the dimer interface, the dimer interface becomes a highly attractive target to prevent formation of the active catalytic site. The dimer region of HIVp is approximately 1800 $Å^2$ in Solvent Accessible Surface Area (SASA), about 1000 $Å^2$ larger than an average SASA dimer interface for homodimeric proteins.[122] 75% of the Gibbs free energy resulting from dimerization of the HIVp monomers lies within the β-sheet region.[123] The interdigitating β-sheet region accounts for 50% of the SASA dimer interface contacts.[124] Mature HIVp is an extremely stable dimer. ITC experiments by Todd *et al.* measured a dissociation rate constant, K_d, (sub nanomolar) at equilibrium at [123] pH 5. Xie and colleagues performed sedimentation equilibrium studies of the HIVp dimer at neutral pH.[125] At a more physiological pH, Xie reported a K_d three orders of magnitude higher (5.8 μM) for the pseudo-wild-type, autolysis-resistant HIVp mutant (Q7K/L33I/L63I) in their studies.

Dimer Interface Inhibitors

Zhang *et al.* performed one of the earliest examples of dissociative inhibition of HIVp.[126] By creating a peptidomimetic based on the last four residues of the C-terminus, Zhang showed through kinetic analysis and sedimentation equilibrium experiments that the peptidomimetic C-terminus can inhibit HIVp activity *via* a dissociative mechanism. Their kinetic analysis, termed the Zhang-Poorman assay, is a fluorescent-based kinetic

analysis that distinguishes dissociative inhibitors from competitive. The Zhang-Poorman assay is now commonly performed by other investigators to determine the inhibition mechanism and rate for dissociative inhibitors. The assay corroborates other biophysical techniques (sedimentation equilibrium, liquid chromatography/mass spectroscopy) to characterize the dissociative mechanism of HIVp inhibition.[127]

Currently there are two major strategies (monovalent and bivalent) to construct ligands against the dimer interface; however, both strategies are rooted in principle by mimicking short sequences of the N- and C-terminal peptides.[128] Althought the N- and C-termini can inhibit HIVp, monovalent inhibitors are derived from the C-terminus due to its position between the monomer's termini. Bivalent peptidomimetic inhibitors contain fragments that complement the N- and C-termini sequences with a linking region that connects the two peptide chains.

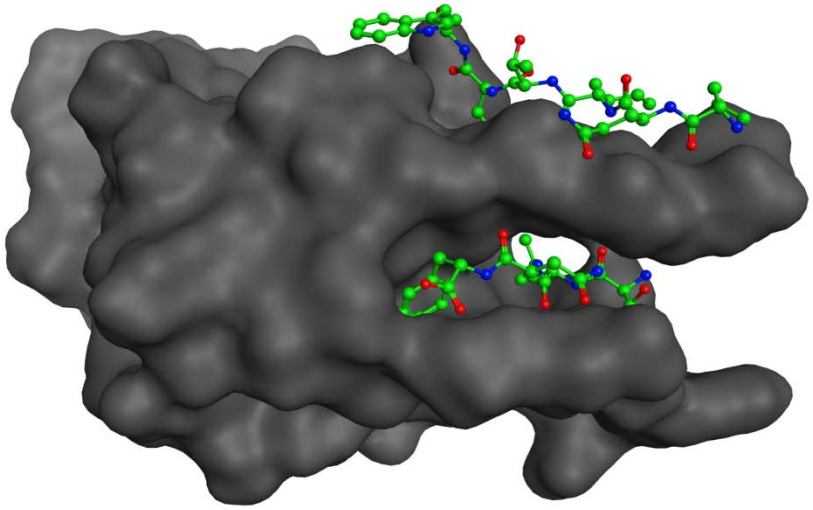

Figure 1.3: Design of Dissociative Inhibitors. One HIVp monomer molecular surface is shown while symmetric residues 1-5 and 95-99 are represented in ball and sticks. Current generation of HIVp dissociative inhibitors are designed to mimic the residues naturally compromising the dimer interface.

Schramm and colleagues targeted the dimer region using short peptidomimetics derived primarily from the C-terminus.[128] Schramm continued the development of his C-terminal derived monovalent peptidomimetic inhibitors by determining the core sequence necessary for dissociative inhibition and added a palymitic acid chain to the N-terminus of the inhibitor.[129, 130] Bannwarth *et al.* varied the length of the carbon chain in the palymitic acid in Schramm's lipoprotein moiety.[131] They also demonstrated that dimer inhibitor efficacy was comparable against mutant strains of HIVp, unlike other competitive site inhibitors. Caflisch *et al.* theoretically mapped the dimer interface using the MCSS method.[132] The MCSS map of the HIVp monomer determined that the complementary Phe99 position is large and can accommodate a variety of groups. The addition of a thyroxine group to Schramm's peptidomimetic corresponding to the complementary Phe99 position increased potency of the monovalent compound, and

validated Caflisch's MCSS dimer interface map.[133] Breccia *et al.* synthesized a monovalent series based on Schramm's compounds, including a positively charged guanidinium group onto the N-terminus of peptidomimetic.[134] Breccia intended for the positively charged guanidinium group to interact favorably with the negatively charged C-terminus.

Jean Chmielewski pioneered the development of bivalent peptidomimetics targeting the HIVp dimer region, linking the N- and C-termini with an alkyl chain of carbons.[135] She optimized the length and flexibility of the alkyl linker region, determining an effective linker length of 16 carbons.[136] The ligand size was reduced to a core peptidomimetic sequence necessary to prevent dimerization, and has increased the strength of bivalent compounds.[137-139] Bouras *et al.* recognized the unfavorable entropy penalty associated with a flexible linker and constructed a bivalent series with more constricted linkers.[140] Bouras substituted Chmielewski's 16-carbon alkyl chain with several aromatic groups (resorcinol, pyridinediol and naphthalenediol) to create a series of bivalent molecules called "molecular tongs". Derivatives of the molecular tongs include substitutions to the aromatic group to increase favorable interactions and the insertion of chemical fragments to replace the peptidomimetic amino acids.[141, 142]

Currently, a large body of empirical evidence indicates dissociative inhibition by dimer inhibitors. Dimer inhibition is largely ascertained by means of Zhang-Poorman kinetic analysis. Unfortunately, the fundamental relationship between the dimer inhibitors and the HIVp dimer interface is not understood. Frutos *et al.* performed 2D [^{1}H-X]-HSQC NMR using a labeled [2-^{13}C]-Trp incorporated into the pseudo-wild-type HIVp.[143] Trp was labeled primarily because there are only two residues (Trp6 and

Trp42), with one (Trp6) in the dimer interface. Frutos reported chemical shifts at the Trp6 residue, which suggested ligand binding near the residue. The authors then performed docking simulations, using AutoDock3. Frutos centered the grid between residues 96-99 of monomeric HIVp, and performed molecular docking to simulate the receptor-ligand binding mechanism. Their study indicates inhibitor binding near Trp6; however, by selecting a grid center on the C-terminus of the monomeric HIVp, the authors restricted their docking simulations to the area occupied by the complementary C-terminus, which biased their result.

The greatest challenge in designing better, non-peptidomimetic inhibitors to target the HIVp dimer interface is the lack of structural information. Greater structural information would clarify the relationship between the HIVp dimer interface and dissociative ligands bound, presumably, at the dimer interface. This recalls the problem associated with the lock-and-key hypothesis; we currently have many keys, but little understanding behind the relationship between the lock and the key.

The studies presented in this dissertation will attempt to elucidate this fundamental relationship. The proceeding chapter details the use of the MPS procedure to map the dimer interface of HIVp, creating a pharmacophore to filter a small molecule library matching the complementary space. The chemical map of the complementary space filtered, *in vitro*, a three dimensional library of small molecule inhibitors able to selectively inhibit HIVp in an allosteric region *via* the dimer interface. Chapter three describes an attempt to computationally model the putative binding mechanism between current dissociative inhibitors complexed to the dimer interface. Although these models do not converge to a consensus binding mechanism, they do report the hypothesized

binding mechanism is unfavorable and is likely not the binding mechanism. Chapter four explains an effort to characterize critical protein-protein interactions involved in creating the HIVp dimer. These crucial interactions can be incorporated into pharmacophore models to filter small molecules capable of dissociating the HIVp dimer.

Finally, the last chapter is a completely novel idea to empirically distinguish the binding mechanism of dissociative inhibitors. Hydrogen-deuterium exchange can determine the location dissociative inhibitors bind against the HIVp monomer, at concentrations and conditions similar to Zhang-Poorman kinetic assay. The dissociative inhibitors will block natural hydrogen exchange within a deuterated solution, allowing us to determine the location of small molecule binding. Coupled with high-resolution MD simulations, we can determine energetically favorable binding mechanism to block dimerization. Medicinal chemistry can then be applied to find stronger inhibitors and allow the development of a potential generation of drugs targeting the HIVp dimer interface.

References

1. Bohacek, R.S.; McMartin, C.; Guida, W.C., The art and practice of structure-based drug design: a molecular modeling perspective. *Med Res Rev* **1996**, *16*, 3-50.

2. Yasgar, A.; Shinn, P.; Jadhav, A.; Auld, D.; Michael, S.; Zheng, W.; Austin, C.P.; Inglese, J.; Simeonov, Compound management for quantitative high-troughput screening. *JALA Charlottesv Va.* **2008**, *13*, 79-89.

3. Glusker, J.P., X-ray crystallography of proteins. *Methods Biochem Anal* **1994**, *37*, 1-72.

4. Kay, L.E., NMR methods for the study of protein structure and dynamics. *Biochem Cell Biol* **1997**, *75*, 1-15.

5. Karplus, M.; McCammon, J.A., Molecular dynamics simulations of biomolecules. *Nat Struct Biol* **2002**, *9*, 646-652.

6. Kier, L.B., Molecular orbital calculation of preferred conformations of acetylcholine, muscarine, and muscarone. *Mol Pharmacol* **1967**, *3*, 487-494.

7. Khedkar S.A.; Malde, A.K.; Coutinho, E.C.; Srivastava, S., Pharmacophore modeling in drug discovery and development: an overview. *Med Chem* **2007**, *3*, 187-197.

8. Kuyper, L.F; Roth, B.; Baccanari, D.P.; Ferone, R.; Beddell, C.R.; Champness, J.N.; Stammers, D.K.; Dann, J.G.; Norrington, F.E., Receptor-based design of dihydrofolate reducatse inhibitors: comparison of cyrstallographically determined enzyme binding with enzyme affinity in a series of carboxy-substituted trimethoprim analogs. *J Med Chem* **1985**, *28*, 303-311.

9. Mattos, C.; Rasmussen, B.; Ding, X.; Petsko, G.A.; Ringe, D., Analogous inhibitors of elastase do not always bind analogously. *Nat Struct Biol* **1994**, *1*, 55-58.

10. von Itzstein, M.; Wu, W.Y.; Kok, G.B.; Pegg, M.S.; Dyason, J.C.; Jin, B.; Van Phan, T.; Smythe, M.L.; White, H.F.; Oliver, S.W.; Colman, P.M.; Varghese, J.N.; Ryan, D.M.; Woods, J.M.; Bethell, R.C.; Hotham, V.J.; Cameron, J.M.; Penn, C.R., Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **1993**, *363*, 418-423.

11. Koshland Jr., D.E., Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci* **1958**, *44*, 98-104.

12. Freire, E., Statistical thermodynamic linkage between conformational and binding equilibria. *Adv Protein Chem* **1998**, *51*, 255-279.

13. Carlson, H.A.; McCammon, J.A., Accommodating protein flexibility in computational drug design. *Mol Pharmacol* **2000**, *57*, 213-218.

14. Laskowski, R.A.; Gerick, F.; Thornton, J.M., The structural basis of allosteric regulation in proteins. *FEBS Lett* **2009**, *583*, 1692-1698.

15. Carlson, H.A., Protein flexibility and drug design: how to hit a moving target. *Curr Opin Chem Biol* **2002**, *6*, 447-452.

16. Henzler-Wildman, K.; Kern, D., Dynamic personalities of proteins. *Nature* **2007**, *450*, 964-972.

17. Pauling, L., The oxygen equilibrium of hemoglobin and its structural interpretation. *Proc Natl Acad Sci* **1935**, *21*, 186-191.

18. Monod, J.; Wyman, J.; Changeux, J.P., On the nature of allosteric transitions: a plausible model. *J Mol Biol* **1965**, *12*, 88-118.

19. Perutz, M.F., The Croonian Lecture, 1968: The Haemoglobin molecule. *Proc R Soc Lond B* **1969**, *173*, 113-140.

20. Edelstein, S.J., Extensions of the allosteric model for haemoglobin. *Nature* **1971**, *230*, 224-227.

21. Koshland Jr., D.E.; Nemethy, G.; Filmer, D., Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **1965**, *5*, 365-385.

22. Kantrowitz, E.R.; Lipscomb, W.N., Escherichia coli, aspartate transcarbamoylase: the molecular basis for a concerted allosteric transition. *Trends Biochem Sci* **1990**, *15*, 53-59.

23. Salazar, C.; Hofer, T., Allosteric regulation of the transcription factor NFAT1 by multiple phoshorylation sites: a mathematical analysis. *J Mol Biol* **2003**, *327*, 31-45.

24. Wobbe, L.; Blifernez, O.; Schwarz, C.; Mussgnug, J.H.; Nickelsen, J.; Kruse, O., Cysteine modification of a specific repressor protein controls the translational status of nucleus-encoded LHCII mRNAs in *Chlamydomonas*. *Proc Natl Acad Sci* **2009**, *106*, 13290-13295.

25. Cooper A.; Dryden, D.T.F., Allostery without conformational change. *Eur Biophys J* **1984**, *11*, 103-109.

26. Popovych, N.; Sun, S.; Ebright, R.H.; Kalodimos, C.G., Dynamically driven protein allostery. *Nature Struct Mol Biol* **2006**, *13*, 831-838.

27. Page, M.J.; Carrell, C.J.; Di Cera, E., Engineering protein allostery: 1.05 A resolution structure and enzymatic properties of a Na^{+}-activated trypsin. *J Mol Biol* **2008**, *378*, 666-672.

28. Gunasekaran, K.; Ma, B.; Nussinov, Is allostery an intrinsic property of *all* dynamic proteins? *Proteins* **2004**, *57*, 433-443.

29. Hilser, VJ, An ensemble view of allostery. *Science* **2010**, *327*, 653-654.

30. Stumpf, M.P.H.; Thorne, T.; de Silva, E.; Stewart, R.; An, H.J.; Lappe, M.; Wiuf, C., Estimating the size of the human interactome. *Proc Natl Acad Sci* **2008**, *105*, 6959-6964.

31. Brass, A.L.; Dykxhoorn, D.M.; Benita, Y.; Yan, N.; Engelman, A.; Xavier, R.J.; Lieberman, J.; Elledge, S.J., Identification of Host Proteins Required for HIV Infection Through a Functional Genomic Screen. *Science* **2008**, *319*, 921-926.

32. Arkin, M.R.; Wells, J.A., Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat Rev Drug Discov* **2004**, *3*, 301-317.

33. Wells, J.A.; McClendon, C.L., Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **2007**, *450*, 1001-1009.

34. Buchwald, P., Small-molecule protein-protein interaction inhibitors: therapeutic potential in light of molecular size, chemical space and ligand binding efficiency considerations. *IUBMB Life* **2010**, *62*, 724-731.

35. Domling, A., Small molecular weight protein-protein interaction antagonists - an insurmountable challenge? *Curr Opin Chem Biol* **2008**, *12*, 281-291.

36. Chene, P., Inhibiting the p53-MDM2 interaction: an important target for cancer therapy. *Nat Rev Cancer* **2003**, *3*, 102-109.

37. MacArthur, R.D.; Novak, R.M., Reviews of anti-infective agents: maraviroc: the first of a new class of antiretroviral agents. *Clin Infect Dis* **2008**, *47*, 236-241.

38. Conte, L.L.; Chothia, C.; Janin, J., The atomic structure of protein-protein recognitions sites. *J Mol Biol* **1999**, *285*, 2177-2198.

39. Cheng, A.C.; Coleman, R.G.; Smyth, K.T.; Cao, Q; Soulard, P.; Caffrey, D.R.; Salzberg, A.C.; Huang, E.S., Structure-based maximal affinity model predicts small-molecule druggability. *Nat Biotechnol* **2007**, *25*, 71-75.

40. Smith, R.D.; Hu, L.; Falkner, J.A.; Benson, M.L.; Nerothin, J.P.; Carlson, H.A., Exploring protein-ligand recognition with Binding MOAD. *J Mol Biol* **2006**, *24*, 414-425.

41. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Del Rev* **2001**, *46*, 3-26.

42. Gandhi, L.; Camidge, D.R.; Ribeiro de Oliveira, M.; Bonomi, P.; Gandara, D.; Khaira, D.; Hann, C.L.; McKeegan, E.M.; Litvinovich, E.; Hemken, P.M.; Dive, C.; Enschede, S.H.; Nolan, C.; Chiu, Y.-L.; Busman, T.; Xiong, H.; Krivoshik, A.P.; Humerickhouse, R.; Shapiro, G.I.; Rudin, C.M., Phase I study of Navitoclax (ABT-263), a novel Bcl-2 family inhibitor, in patients with small-cell lung cancer and other solid tumors. *JCO* **2011**, *29*, 909-916.

43. Tse, C.; Shoemaker, A.R.; Adickes, J.; Anderson, M.G.; Chen, J.; Jin, S.; Johnson, E.F.; Marsh, K.C.; Mitten, M.J.; Nimmer, P.; Roberts, L.; Tahir, S.K.; Xiao, Y.; Yang, X.; Zhang, H.; Fesik, S.; Rosenberg, S.H.; Elmore, S.W., ABT-263: a potent and orally bioavailable Bcl-2 family inhibitor. *Cancer Res* **2008**, *68*, 3421-3428.

44. Kuntz, I.D.; Chen, K.; Sharp, K.A.; Kollman, P.A., The maximal affinity of ligands. *Proc Natl Acad Sci* **1999**, *96*, 9997-10002.

45. Bogan, A.A. and Thorn, K.S., Anatomy of hot spots in protein interfaces. *JMB* **1998**, *280*, 1-9.

46. Shoichet, B.K.; Kuntz, I.D., Matching chemistry and shape in molecular docking. *Protein Eng* **1993**, *6*, 723-732.

47. Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comp Chem* **1998**, *19*, 1639-1662.

48. Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R., Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **1997**, *267*, 727-748.

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

50. 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.

51. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G., A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* **1996**, *261*, 470-489.

52. Kuntz, I.D.; Blaney, J.M.; Oathley, S.J.; Langridge, R.; Ferrin, T.E., A geometric approach to macromolecule-ligand interactions. *J Mol Biol* **1982**, *161*, 269-288.

53. Leach, A.R., Molecular Modelling: Principles and Applications; Prentice Hall; 2001

54. Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M., CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comp Chem* **1983**, *4*, 187-217.

55. Pearlman, D.A.; Case, D.A.; Caldwell, J.W.; Ross, W.S.; Cheatham III, T.E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P., AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp Phys Commun* **1995**, *91*, 1-41.

56. Scott, W.R.P; Hunenberger, P.H.; Tironi, I.G.; Mark, A.E.; Billeter, S.R.; Fennen, J.; Torda, A.E.; Huber, T.; Kruger, P.; van Gunsteren, W.F., The GROMOS biomolecular simulation program package. *J Phys Chem A* **1999**, *103*, 3596-3607.

57. Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J Comp Chem* **2005**, *26*, 1781-1802.

58. Case, D.A.; Darden, T.A.; Cheatham III, T.E.; Simmerling, C.L.; Wang, J.; Duke, R.E.; Luo, R.; Merz, K.M.; Wang, B.; Pearlman, D.A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J.W.; Ross, W.S.; Kollman, P.A. (2004), AMBER 8, University of California, San Francisco

59. Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz Jr., K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* **1995**, *117*, 5179-5197.

60. Wickstrom, L.; Okur, A.; Simmerling, C., Evaluating the performance of the ff99SB force field based on NMR scalar coupling data. *Biophys J* **2009**, *97*, 853-856.

61. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M.C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P., A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comp Chem* **2003**, *24*, 1999-2012.

62. Woods, R.J.; Dwek, R.A.; Edge, C.J.; Fraser-Reid, B., Molecular mechanical and molecular dynamic simulations of glycoproteins and oligosaccharides. 1. GLYCAM_93 parameter development. *J Phys Chem* **1995**, *99*, 3832-3846.

63. Case, D.A.; Cheatham III, T.E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr., K.M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J., The AMBER biomolecular simulation programs. *J Comp Chem* **2005**, *26*, 1668-1688.

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

65. Jacobson, M.P.; Friesner, R.A.; Xiang, Z.; Honig, B., On the role of the crystal environment in determining protein side-chain conformations. *J Mol Biol* **2002**, *320*, 597-608.

66. Meagher, K.L.; Carlson, H.A., Solvation influences flap collapse in HIV-1 protease. *Proteins* **2005**, *58*, 119-125.

67. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L., Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **1983**, *79*, 926-935.

68. Petrenko, V.F.; Ryzhkin, I.A., Electron energy spectrum of ice. *Phys Rev Lett* **1993**, *71*, 2626-2629.

69. Feig, M.; Karanicolas, J.; Brooks III, C.L. (2001), MMTSB Tool Set, MMTSB NIH Research Resource, The Scripps Research Institute

70. Kabsch, W. and Sander, C., Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577-2637.

71. Feig, M.; Brooks III, C.L., Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr Opin Struct Biol* **2004**, *14*, 217-224.

72. Still, W.C.; Tempczyk, A.; Hawley, R.C.; Hendrickson, T., Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc* **1900**, *112*, 6127-6129.

73. Chen, J.; Brooks, C.L.; Khandogin, J., Recent advances in implicit solvent-based methods for biomolecular simulations. *Curr Opin Struct Biol* **2008**, *18*, 140-148.

74. Onufriev, A.; Bashford, D.; Case, D.A., Modification of the generalized Born model suitable for macromolecules. *J Phys Chem B* **2000**, *104*, 3712-3720.

75. Feig, M.; Onufriev, A.; Lee, M.S.; Im, W.; Case, D.A.; Brooks III, C.L., Performance comparison of generalized born and poisson methods in the calculation of electrostatic solvation energies for protein structures. *J Comp Chem* **2003**, *25*, 265-284.

76. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E., Equation of state calculations by fast computing machines. *J Chem Phys* **1953**, *21*, 1087-1092.

77. Jorgensen, W.L.; Tirado-Rives, J., Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J Comp Chem* **2005**, *26*, 1689-1700.

78. Jorgensen, W.L.; Maxwell, D.S.; Tirado-Rives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* **1996**, *118*, 11225-11236.

79. Allen, K.N.; Bellamacina, C.R.; Ding, X.; Jeffery, C.J.; Mattos, C.; Petsko, G.A.; Ringe, D., An experimental approach to mapping the binding surfaces of crystalline proteins. *J Phys Chem* **1996**, *100*, 2605-2611.

80. Nienaber, V.L.; Richardson, P.L.; Klighofer, V.; Bouska, J.J.; Giranda, V.L.; Greer, J., Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nat Biotechnol* **2000**, *18*, 1105-1108.

81. Miranker, A. and Karplus, M., Functionality of maps of binding sites: a multiple copy simultaneous search method. *Proteins* **1991**, *11*, 29-34.

82. Schubert, C.R.; Stultz, C.M., The multi-copy simultaneous search methodology: a fundamental tool for structure-based drug design. *J Comput Aided Mol Des* **2009**, *23*, 475-489.

83. Carlson, H.A.; Masukawa, K.M.; Rubins, K.; Bushman, F.D.; Jorgensen, W.L.; Lins, R.D.; Briggs, J.M.; McCammon, J.A., Developing a dynamic pharmacophore model for HIV-1 integrase. *J Med Chem* **2000**, *43*, 2100-2114.

84. Meagher, K.L.; Lerner, M.G.; Carlson, H.A., Refining the MPS pharmacophore method: consistency across three independent HIV-1 protease models. *J Med Chem* **2006**, *49*, 3478-3484.

85. Damm, K.L.; Carlson, H.A., Exploring experimental sources of multiple protein conformations in structure-based drug design. *J Am Chem Soc* **2007**, *129*, 8225-8235.

86. Bowman, A.L; Lerner, M.G.; Carlson, H.A., Protein flexibility and species specificity in structure-based drug discovery: dihydrofolate reductase as a test system. *J Am Chem Soc* **2007**, *129*, 3634-3640.

87. Lerner, M.G.; Bowman, A.L.; Carlson, H.A., Incorporating dynamics in E. coli dihydrofolate reductase enhances structure-based drug discovery. *J Chem Inf Model* **2007**, *47*, 2358-2365.

88. Meagher, K.L.; Carlson, H.A., Incorporating protein flexibility in structure-based drug design: using HIV-1 protease as a test case. *J Am Chem Soc* **2004**, *126*, 13276-13281.

89. Damm, K.L.; Ung, P.M.U.; Quintero, J.J.; Gestwicki, J.E.; Carlson, H.A., A poke in the eye: inhibition HIV-1 proteinase through its flap-recognition pocket. *Biopolymers* **2008**, *89*, 643-652.

90. De Clercq, E., Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. *Int J Antimicrob Agents* **2009**, *33*, 307-320.

91. Tavassoli, A., Targeting the protein-protein interactions of the HIV lifecycle. *Chem Soc Rev* **2011**, *40*, 1337-1346.

92. (UNAIDS), United Nations Programme on HIV/AIDS, 2008 Report on the global AIDS epidemic, http://www.unaids.org/en/dataanalysis/epidemiology/2008reportontheglobalaidsepidemic/

93. Freed, E.O, HIV-1 Gag Proteins: diverse functions in the virus life cycle. *Virology* **251**, *251*, 1-15.

94. Bryant, M.; Ratner, L., Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. *Proc Natl Acad Sci* **1990**, *87*, 523-527.

95. Provitera, P.; El-Maghrabi, R.; Scarlata, S., The effect of HIV-1 Gag myristoylation on membrane binding. *Biophys Chem* **2006**, *119*, 23-32.

96. Murakami, T., Roles of the interactions between Env and Gag proteins in the HIV-1 replication cycle. *Microbiol Immunol* **2008**, *52*, 287-295.

97. Frankel, A.D.; Young, J.A.T., HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem* **1998**, *67*, 1-25.

98. Kohl, N.E.; Emini, E.A.; Schleif, W.A.; Davis, L.J.; Heimbach, J.C.; Dixon, R.A.; Scolnick, E.M.; Sigal, I.S., Active human immunodeficiency virus protease is required for viral infectivity. *Proc Natl Acad Sci* **1988**, *85*, 4686-4690.

99. Krausslich, H.G.; Schneider, H. Zybrath, G.; Carter, C.A.; Wimmer, E., Processing of in vitro-synthesized gag precursor proteins of human immunodeficiency virus (HIV) type 1 by HIV proteinase generated in Escherichia coli. *J Virol* **1988**, *62*, 4393-4397.

100. Rose, J.R.; Babe, L.M; Craik, C.S., Defining the level of human immunodeficiency virus type 1 (HIV-1) protease activity required for HIV-1 particle maturation and infectivity. *J Virol* **1995**, *69*, 2751-2758.

101. Navia, M.A.; Fitzgerald, P.M.; McKeever, B.M.; Leu, C.T.; Heimbach, J.C.; Herber, W.K.; Sigal, I.S.; Darke, P.L.; Springer, J.P., Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* **1989**, *337*, 615-620.

102. Rao, J.K.M.; Erickson, J.W.; Wlodawer, Structural and evolutionary relationships between retroviral and eucaryotic aspartic proteinases. *Biochemistry* **1991**, *30*, 4663-4671.

103. Pillai, B.; Kannan, K.K.; Hosur, M.V., 1.9 A x-ray study shows closed flap conformation in crystals of tethered HIV-1 PR. *Proteins* **2001**, *43*, 57-64.

104. De Clercq, E., The history of antiretrovirals: key discoveries over the past 25 years. *Rev Med Virol* **2009**, *19*, 287-299.

105. Bihani, S.; Das, A.; Prashar, V.; Ferrer, J.-L.; Hosur, M.V., X-ray structure of HIV-1 protease in situ product complex. *Proteins* **2009**, *74*, 594-602.

106. Das, A.; Mahale, S.; Prashar, V.; Bihani, S.; Ferrer, J.-L.; Hosur, M.V., X-ray snapshot of HIV-1 protease in action: observation of tetrahedral intermediate and short ionic hydrogen bond SIHB with catalytic aspartate. *J Am Chem Soc* **2010**, *132*, 6366-6373.

107. Lee, H.; Darden, T.A.; Pederson, L.G., An ab initio quantum mechanical model for the catalytic mechanism of HIV-1 protease. *J Am Chem Soc* **1996**, *118*, 3946-3950.

108. Trylska, J.; Grochowski, P.; McCammon, J.A., The role of hydrogen bonding in the enzymatic reaction catalyzed by HIV-1 protease. *Protein Sci* **2004**, *13*, 513-528.

109. Abell, Andrew, Advances in Amino Acid Mimetics and Peptidomimetics; Jai Press Ltd.; 1997

110. Turner, S.R., HIV protease inhibitors - the next generation. *Curr Med Chem Anti-Infective Agents* **2002**, *1*, 141-162.

111. Dresser, G.K.; Spence, J.D.; Bailey, D.G., Pharmacokinetic-pharmacodynamic consequences and clinical relevance of cytochrome P450 3A4 inhibition. *Clin Pharmacokinet* **2000**, *38*, 41-57.

112. Marinec, P.S.; Chen, L.; Barr, K.J.; Mutz, M.W.; Crabtree, G.R.; Gestwicki, J.E., FK506-binding protein (FKBP) partitions a modified HIV protease inhibitor into blood cells and prolongs its lifetime in vivo. *Proc Natl Acad Sci* **2009**, *106*, 1336-1341.

113. Foulkes-Murzycki, J.E.; Scott, W.R.P; Schiffer, C.A., Hydrophobic sliding: a possible mechanism for drug resistance in human immunodeficiency virus type 1 protease. *Structure* **2007**, *15*, 225-233.

114. Perryman, A.L.; Lin, J.-H.; McCammon, J.A., HIV-1 proteinase molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. *Protein Sci* **2004**, *13*, 1108-1123.

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

116. Layten, M.; Hornak, V.; Simmerling, C., The open structure of a multi-drug-resistant HIV-1 protease is stabilized by crystal packing contacts. *J Am Chem Soc* **2006**, *128*, 13360-13361.

117. Lexa, K.W.; Damm, K.L; Quintero, J.J.; Gestwicki, J.E.; Carlson, H.A., Clarifying allosteric control of flap conformations in the 1TW7 crystal structure of HIV-1 protease. *Proteins* **2009**, *74*, 872-880.

118. Bottcher, J.; Blum, A.; Dorr, S., Heine, A.; Diederich, W.E.; Klebe, G., Targeting the open-flap conformation of HIV-1 protease with pyrrolidine-based inhibitors. *ChemMedChem 3*, 1337-1344.

119. Lexa, K.W.; Carlson, H.A., Binding to the open conformation of HIV-1 protease, *Proteins - Accepted,* DOI: 10.1002/prot.23054

120. Torbeev, V.Y.; Raghuraman, H.; Mandal, K.; Senapati, S.; Perozo, E.; Kent, S.B., Dynamics of "flap" structures in three HIV-1 protease/inhibitor complexes probed by total chemical synthesis and pulse-EPR spectroscopy. *J Am Chem Soc* **2009**, *131*, 884-885.

121. Perryman, A.L.; Zhang, Q.; Soutter, H.H.; Rosenfeld, R.; McRee, D.E.; Olson, A.J.; Elder, J.E.; Stout, C.D., Fragment-based screen against HIV protease. *Chem Biol Drug Des* **2010**, *75*, 257-268.

122. Jones, S.; Thornton, J.M., Principles of protein-protein interaction. *Proc Natl Acad Sci* **1996**, *93*, 13-20.

123. Todd, M.J.; Semo, N.; Freire, E., The structural stability of the HIV-1 protease. *JMB* **1998**, *283*, 475-488.

124. Weber, I.T, Comparison of the crystal structures and intersubunit interactions of Human Immunodeficiency and Rous Sarcoma Virus proteases. *J Biol Chem* **1990**, *265*, 10492-10496.

125. Xie, D.; Gulnik, S.; Gustchina, E.; Yu, B.; Wei, Shao; Qoronfleh, W.; Nathan, A.; Erickson, J.W., Drug resistance mutations can affect dimer stability of HIV-1 protease at neutral pH. *Protein Sci* **1999**, *8*, 1702-1707.

126. Zhang, Z.-Y.; Poorman, R.A.; Maggiora, L.L.; Heinrikson, R.L.; Kezdy, F.J., Dissociative inhibition of dimeric enzymes. *J Biol Chem* **1991**, *266*, 15591-15594.

127. Davis, D.A.; Tebbs, I.R.; Daniels, S.I.; Stahl, S.J.; Kaufman, J.D.; Wingfield, P.; Bowman, M.J.; Chmielewski, J.; Yarchoan, R., Analysis and characterization of dimerization inhibition of a multi-drug-resistant Human Immunodeficiency Virus Type 1 protease using a novel size-exclusion chromatographic approach. *Biochem J* **2009**, *419*, 497-506.

128. Schramm, H.J.; Billich, A.; Jaeger, J.; Rucknagel, K.-P.; Arnold, G.; Schramm, W., The inhibition of HIV-1 protease by interface peptides. *Biochem Biophys Res Commun* **1993**, *194*, 595-600.

129. Schramm, H.J.; Boetzel, J.; Buttner, J.; Fritsche, E.; Gohring, W.; Jaeger, E.; Konig, S.; Thumfart, O.; Wenger, T.; Nagel, N.E.; Schramm, W., The inhibition of human immunodeficiency virus protease by 'interface peptides'. *Antiviral Res* **1996**, *30*, 155-170.

130. Schramm, H.J.; de Rosny, E.; Reboud-Ravaux, M.; Buttner, J.; Dick, A.; Schramm, W., Lipopeptides as dimerization inhibitors of HIV-1 protease. *Biol Chem* **1999**, *380*, 593-596.

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

132. Caflisch, A.; Schramm, H.J.; Karplus, M., Design of dimerization inhbitors of HIV-1 aspartic proteinase: A computer-based combinatorial approach. *J Comput Aided Mol Des* **2000**, *14*, 161-179.

133. Dumond, J.; Boggetto, N.; Schramm, H.J.; Schramm, W.; Takahashi, M.; Reboud-Ravaux, M., Thyroxine-derivatives of lipopeptides: bifunctional dimerization inhibitors of human immunodeficiency virus-1 protease. *Biochem Pharmacol* **2003**, *65*, 1097-1102.

134. Breccia, P.; Boggetto, N.; Perez-Fernandez, R.; Van Gool, M.; Takahashi, M.; Rene, L.; Prados, P.; Badet, B.; Reboud-Ravaux, M.; de Mendoza, J., Dimerization inhibitors of HIV-1 protease based on bicyclic guanidinium subunit. *J Med Chem* **2003**, *46*, 5196-5207.

135. Zutshi, R.; Franciskovich, J.; Shultz, M.; Schweitzer, B.; Bishop, P.; Wilson, M.; Chmielewski, J., Targeting the dimerization interface of HIV-1 protease: inhibition with cross-linked interfacial peptides. *J Am Chem Soc* **1997**, *119*, 4841-4845.

136. Shultz, M.D.; Bowman, M.J.; Ham, Y.-W.; Zhao, X.; Tora, G.; Chmielewski, J., Small-molecule inhibitors of HIV-1 protease dimerization derived from cross-linked interfacial peptides. *Agnew Chem Int Ed* **2000**, *36*, 2710-2713.

137. Bowman, M.J.; Byrne, S.; Chmielewski, J., Switching between allosteric and dimerization inhibition of HIV-1 protease. *Chem Biol* **2005**, *12*, 439-444.

138. Hwang, Y.S.; Chmielewski, Development of low molecular weight HIV-1 protease dimerization inhbitors. *J Med Chem* **2005**, *48*, 2239-2242.

139. Lee, S.-G.; Chmielewski, J., Rapid synthesis and in situ screening of potent HIV-1 protease dimerization inhibitors. *Chem Biol* **2006**, *13*, 421-426.

140. Bouras, A.; Boggetto, N.; Benatalah, Z.; de Rosny, E.; Sicsic, S.; Reboud-Ravaux, M., Design, synthesis, and evaluation of conformationally constrained tongs, new inhibitors of HIV-1 protease dimerization. *J Med Chem* **1999**, *42*, 957-962.

141. Merabet, N.; Dumond, J.; Collinet, B.; Van Baelinghem, L.; Boggetto, N.; Ongeri, S.; Ressad, F.; Reboud-Ravaux, M.; Sicsic, S., New constrained "molecular tongs" designed to dissociate HIV-1 protease dimer. *J Med Chem* **2004**, *47*, 6392-6400.

142. Bannwarth, L.; Kessler, A.; Pethe, S.; Collinet, B.; Merabet, N.; Boggetto, N.; Sicsic, S.; Reboud-Ravaux, M.; Ongeri, S., Molecular tongs containing amino acid mimetic fragments: new inhibitors of wild-type and mutated HIV-1 protease dimerization. *J Med Chem* **2006**, *49*, 4657-4664.

143. Frutos, S.; Rodrigues-Mias, R.A.; Madurga, S.; Collinet, B.; Reboud-Ravaux, M.; Ludevid, D.; Giralt, E., Disruption of the HIV-1 protease dimer with interface peptides: structural studies using NMR spectroscopy combined with [2-^{13}C]-Trp selective labeling. *Biopolymers* **2007**, *88*, 164-173.

Chapter 2

Development of MPS pharmacophore models of the HIV-1 protease dimer interface

Introduction

HIV-1 protease (HIVp) is a 99 amino acid protein that forms a C_2 symmetric homodimer.[1] Each protease chain forms half of the active catalytic pocket, and an active protease is necessary for viral maturation and propagation.[2] HIVp dimer interface is approximately 1800 Å^2 in solvent accessible surface area (SASA).[3] HIVp has a smaller contact surface than other typical protein-protein interactions (PPIs) but larger than the average of ~1700 Å^2 for most homodimers.[4]

Upon dimerization, 8.9-14.5 kcal/mol of Gibbs free energy stabilizes the homodimer under varying temperature, pH, and ionic conditions.[5] Approximately 75% of the free energy lies in the interdigitating N- and C-termini, creating a β-sheet that composes half of the dimer interface contacts.[6] The two termini of the HIVp monomer are highly dynamic, with fluctuations ranging from 5 to 9 Å.[7, 8] Xie *et al.*[9] reported a dissociation rate (K_D) equilibrium of 5.8 μM between HIVp monomer and dimer at neutral pH, approximately one order of magnitude greater than the K_D range reported by Todd *et al.*[5]

The HIVp dimer interface is considered a potential therapeutic target for two

beneficial reasons.[10] Failure to form a homodimer results in an inactive HIVp. The symmetric active site contains two aspartic acids, one from each monomer, that are required for a general acid/general base reaction with water and the substrate.[11] The second reason is the highly conserved nature of the dimer interface.[12] Mutations in the interface region are deleterious to protease activity.[9] Selective pressure from active-site, competitive inhibitors allows HIVp to mutate and escape drug efficacy.[13] Targeting the dimer interface with small molecules has high therapeutic value and deserves attention as a simple, challenging model for studying other PPIs.

In 1991, the first proof of dissociative inhibition was given by Zhang *et al.*,[14] who used kinetic analysis and sedimentation equilibration to show that a C-terminal peptidomimetic could disrupt the HIVp dimer. The kinetic analysis was based on a fluorescent-based assay, and the Zhang-Poorman assay has been used widely to discriminate the mechanism of HIVp inhibition. Dimerization inhibitors fall into two classes: monovalent peptidomimetic inhibitors derived from the C-terminus and bivalent inhibitors mimicking the N- and C-termini with a linking region joining the two segments. Both strategies have led to submicromolar inhibition rates of dimerization, $K_{i,D}$.[15]

Schramm and colleagues pioneered the monovalent class of inhibitors constructed from mimicking the C-terminus.[16] The most effective inhibitors possess the common amino-acid motif of YEL, meant to displace residues 97-99 of the C-terminus. The use of a palmytic acid extended from the N-terminus or a thryoxine moiety in place of the leucine improved dissociative inhibition rates to the low-nM range.[17, 18] Frutos *et al.* performed an NMR structural study following the chemical shift perturbations with a

labeled [2-^{13}C]-Trp residue.[19] Labeled protein was followed upon addition of the dimer inhibitor Ac-SYEL-OH, titrated and monitored by 2D [^{1}H-^{13}C]-HSQC. The inhibitor created chemical shifts at the Trp6 residue, indicating a change in the chemical environment near Trp6 upon ligand binding. A model of the bound state was created using the 1HWR crystal structure, by removing the competitive inhibitor from the active site and truncating chain B to the C-terminal sequence of TLNF. The backbone of the peptide inhibitor was taken from the C-terminal sequence, and Autodock 3 was used to optimize the orientations of the side chains.

Caflisch *et al.* used the Multiple Copy Simultaneous Search method to computationally map chemical functionalities to the dimer interface.[20] The study discovered a large, accommodating area normally occupied by the complimentary Phe99' of the other monomer. Their functional map suggested the incorporation of a thyroxine moiety into the peptidomimetic inhibitor synthesized by Dumond *et al.*, which was subsequently found to increased the potency of the inhibitor.[17] Their application of the Multiple Copy Simultaneous Search did not allow for the significant degree of flexibility inherent in the termini because restraints were placed on the protease.

Bowman and Chmielewski developed bivalent, peptidomimetic inhibitors to mimic the dimer interface region occupied by both the N- and C-termini of the HIVp.[21] The two chains were optimized and linked together by a long, flexible alkyl chain.[22] SAR studies revealed that the optimal length to join the two was a 14-carbon, alkyl chain. Further refinement of the side chains in these bivalent inhibitors resulted in $K_{i,D}$ ~30nM.[23] Reboud-Ravaux and coworkers recognized the entropic penalty associated with a long alkyl chain and constructed a bivalent inhibitor with a rigid linker, termed "molecular

tongs."[24] Geometries of the rigid linker were explored to create favorable interactions with the C-terminus, and derivatives were created to produce a non-peptidomimetic compound.[25, 26]

To date, the mechanism of dissociative inhibition is confirmed through Zhang-Poorman analysis. Most of the compounds are peptidomimetic inhibitors derived from the termini, and it is assumed that they bind to a monomer of HIVp in the same fashion as the dimer contacts. However, there is little evidence to confirm the structural complementarity between the dimer interface and the ligands targeting the region. This makes any rational design of new inhibitors very difficult. Improved inhibitors are needed because the existing inhibitors have unfavorable pharmacokinetic and toxic properties.[27] They are very large peptides (>1000 MW) that are not appropriate for drug development. If this mode of action is to ever become viable as a drug target, novel small molecules scaffolds are required.

To the best of our knowledge, only two smaller inhibitors of dimerization have been found. Simplified derivatives of didemnaketal A (an isolate from the sea squirt *Ascidian didemnum*) have a MW of 542 and inhibit HIVp in a dissociative mechanism with $K_{i,D}$ ranging 2.1-29.9 μM.[28] Schisanlactone A is a 464-MW natural product from the Vietmanese mushroom *Ganoderma colossum* that inhibits dimerization ($K_{i,D}$=17.5 μM).[29] (Some bile acids have weak activity but were incorporated into the peptidic inhibitors simply as rigid linkers.[30])

In this work, we have identified organic molecules under 500 MW with unique scaffolds that target the dimer interface, providing an important step toward more drug-

like chemical space. Our best inhibitors have affinities comparable to the natural products, but weaker than the large peptides. We used computational methods to derive pharmacophore models of the dimer interface. Molecular dynamics (MD) simulations were used to generate an ensemble of protein conformations for our multiple protein structure (MPS) method. Each conformation of the ensemble was mapped with small molecule probes, and locations conserved across the majority of structures were used to identify consensus sites. These sites were translated into pharmacophore models which were then used to identify small molecule inhibitors able to induce dissociative inhibition at the HIVp dimer interface.

Using small molecules to disrupt PPIs is a challenging new pursuit with many medically relevant applications.[31] Bcl-XL, gp120-CD4 and p53-MDM2 are a few examples of dissociative inhibition by small molecules at a PPI that are changing medicine.[32-34] We hope that our pursuits may eventually lead to another example for the field.

Methods

Molecular Dynamics Simulations

Several years ago, we published 3-ns MD simulations of the dimeric form of apo HIVp, based on very standard procedures.[35] We initially used snapshots from that simulation to provide a proof of principle for our approach to identify inhibitors of dimerization. We then extended the previous simulation to 30 ns to provide greater conformational sampling (using a combination of AMBER8 or 10 and the FF99SB force

field[36-38]). We also calculated a 35-ns trajectory of the monomeric form of HIVp specifically for this study. Protocols are detailed here.

MD simulations of the monomer were performed using AMBER10 and the FF99SB force field.[36-38] Initial atomic coordinates for the HIVp monomer simulation were obtained from the Protein Data Bank (PDB) repository (PDB ID: 1HHP).[39] This was the same structure used to initiate the dimer simulations, except that the symmetry partner was not used for the monomeric state. Hydrogen atoms were added via the AMBER tLEaP module.[40] The +2*e* charge of the HIVp monomer was neutralized with the addition of two chloride ions. The counter ions were placed 10 Å from the protein surface, guided by an electrostatic potential map calculated using APBS and its PyMOL plugin developed in house.[41-43] The system was solvated with an octahedral box that extended 16 Å out from the protein surface in all directions; a closeness parameter of 0.5 was used to place 10,836 TIP3P water molecules.[44]

Minimization and equilibration of the solvated protein system occurred over a series of steps. Hydrogen atoms, water molecules, and ions were first minimized while all non-hydrogen atoms of the protein were restrained. Next restraints were removed from the side chains and the system further minimized, and finally, all restraints were removed and all atoms were minimized.

In equilibration, SHAKE and 1-fs time step were used unless otherwise noted. A 10-Å vdw cutoff was used and PME boundary conditions were applied.[45, 46] Equilibration occurred in four steps: 1) Water and counter ions are heated from 10 to 310 K under constant volume for 50 ps (time step = 2 fs). During the temperature increase,

the remainder of the system was restrained (50 kcal/mol). 2) Water and counter ions were then allowed to equilibrate at 310 K for another 10 ps while the restraints were maintained. 3) The protein was then allowed to heat from 10 K to 310 K in three, 10-ps steps under constant volume. In the first three equilibration steps, the temperature increased from 10 K to 110 K, then 110 K to 210 K, and then 210 K to 310 K. Meanwhile, the restraints on the hydrogen atoms, water molecules, and counter ions were relaxed from 2 kcal/mol to 1 kcal/mol and, finally, to 0.5 kcal/mol. 4) The system was then allowed to equilibrate at a constant pressure in a series of steps. The first step was a 10-ps equilibration with a restraint (0.5 kcal/mol) on all non-hydrogen atoms of the protein. Next, the system was equilibrated for 10 ps with a weaker restraint (0.1 kcal/mol) on the same atoms. Last, the system was equilibrated for 2 ns while fully unrestrained. The final production phase of the MD was run for 35 ns.

The simulation was monitored and evaluated based on the RMSD of the protein fluctuations, calculated using AMBER's ptraj program.[40] The RMSD was based on Cα within the flap region (45-55), the N-terminus (1-4), the C-terminus (95-99), and the core (5-45 and 55-94). The equilibrated structure was used as the reference state. Secondary structure of the protease monomer was measured using the Dictionary of Protein Secondary Structure (DSSP) program.[47]

Pharmacophore Design

The MPS method for deriving pharmacophore models was previously applied to the active site of HIVp.[48, 49] In this work, the same basic procedure was used, except that the focus was on surface of the dimer interface. Structures for the initial pharmacophore

model were produced from snapshots of the previous 3-ns MD simulation.[35] The initial 0-ps frame, the 100-ps frame and every 300-ps snapshot afterward (until the 2800-ps structure) were selected for MPS. Each chain in the dimer was separated, and the two monomers were both used independently for a total of 22 conformations representing one face of the symmetric dimer contacts. Each structure was flooded with 500 small molecule probes (methanol, benzene, and ethane) in 16-Å radius sphere that encompassed the dimer interface. The probes were minimized to the protein surface using a low-temperature, random-walk procedure called the Multi Unit Search for Interacting Conformers (MUSIC) which is implemented within the BOSS program.[43] The OPLS force field[50, 51] was used, and the dielectric was set to 20 to reduce the attractive force of charge residues. With a dielectric of 1, the methanol probes are strongly attracted to Asp 25 and important local, hydrogen-bonding minima can be missed. The protein structures were held rigid during the minimization of the probe molecules.

The minimized probes form clusters along the protein surface in each snapshot. If the cluster contained ≥8 probes, a "parent" probe (the probe with the lowest energy in the cluster) was selected as a representative. Each snapshot and its parent probes were overlaid to the equilibrated 1HHP structure, using a wRMSD alignment program.[52] Alignment of the structures resulted in clusters of parent probes, and represented "clusters of clusters". A consensus cluster was required to have parent probes from ≥50% of the MPS. Each consensus cluster was translated to a spherical pharmacophore element, centered at the average position of the parent probes (benzene centroid, ethane carbon-carbon bond midpoint, and methanol's oxygen atom). Each element's radius was

based on the RMSD of those same features. Benzene pharmacophore elements were labeled as aromatic. Clusters with overlapping benzene and ethane probes were considered hydrophobic or aromatic in nature. Methanol pharmacophore elements were classified a hydrogen-bond donor or acceptor based on the directionality of their interaction with the protein surface.

Two pharmacophore models were derived from the longer simulations of the dimer and monomer of HIVp. They were developed following the procedure above, except that the MPS were not taken from evenly spaced snapshots of the simulation. For the longer simulations, the MPS snapshots were selected by clustering the MD trajectory into 10 representative conformations. The means-clustering algorithm was employed and the dimer-interface residues were selected (all atoms of residues 1-6, 23-29, and 87-99). The dimer conformations were separated, and each chain was clustered into 10 representative clusters to give 20 individual snapshots. Snapshots from every 5 ps were used in the clustering.

For the initial probe placement, the Cα of residue Thr26 was selected as the flooding center and the cutoff was a 15-Å radius. Minimization was performed as previously described. An all-atom wRMSD was performed to incorporate all of the significant flexible motion observed at the N- and C-termini. A Jarvis-Patrick clustering program[53] was used to group the probes. Jarvis-Patrick parameters, j and k respectively, for each probe are as follows: 17 and 1 for benzene, 15 and 1 for ethane, and 17 and 1 for methanol. The consensus clusters and pharmacophore elements were determined as described for the 3-ns model.

Lastly, a "core" pharmacophore model was developed to represent pharmacophore elements common to both the dimer and monomer data. The core model was built by removing any elements from the dimer model that were not observed in the monomer model.

Virtual Screening

A computational database of ~34K compounds was generated for the chemical library of the University of Michigan's Center for Chemical Genomics (CCG). Conformations for the compounds were pre-generated using OpenEye's OMEGA program.[54] A maximum of 300 conformations was allowed, with energy and RMS thresholds of 14 kcal/mol and 1 Å, respectively.[55] The pharmacophores were compared to the database using the search query in the Molecular Operating Environment (MOE) program.[56] Two means were used to relax the matching criteria and identify more compounds. First, the radii of the elements were sequentially increased (1×RMSD, 1.33×RMSD, 1.67×RMSD, 2×RMSD…), and second mismatches were allowed (fitting all elements, missing one element, missing 2 elements, etc). Additional criteria (≤500 MW, chemical diversity, commercial availability, etc) were used to prioritize the hits.

Compound Screening

Before characterizing the inhibition mechanism, purchased compounds were screened *in vitro* for HIVp inhibition. A fluorescence-based assay was used.[57] The substrate was an oligopeptide, RE(EDANS)SQNYPIVQK(dabcyl)R, purchased from Molecular Probes (Cat. No. H-2930). HIVp was originally purchased from Bachem Biosciences (Product H-9040) with larger quantities donated by Prof. Cecilia Schiffer of

the University of Massachusetts (both wild-type and multi-drug resistant strains of HIVp). Compounds were purchased from Chembridge or ChemDiv and initially dissolved in DMSO at 5 mM or 10 mM concentrations. Due to the low solubility of the compounds, PEG-400 was added to Buffer A (20mM phosphate, 1 mM DTT, 1 mM EDTA, 20% glycerol, 0.1% CHAPS and 0.1 % PEG-400 at pH 5.1) and Buffer B (100mM Sodium Acetate, 1M NaCl, 1mM EDTA, 1mM DTT 20% v/v Glycerol, 0.1% w/v CHAPS and 0.2% PEG-400 at pH 4.7). Screened compounds were tested at a final concentration of 50 μM against 30 nM HIVp (final concentration) in either Buffer A or Buffer B. The compounds were screened in duplicate after 40 minutes incubation. The enzymatic reaction was initiated with 2 μM of substrate (final concentration), and the initial velocities of the reaction with putative dimer inhibitors were compared to uninhibited HIVp. Compounds reported to inhibit HIVp by 25% or more were further characterized with the Zhang-Poorman assay. The competitive inhibitor pepstatin A was used as a control.

Zhang-Poorman Analysis

The FRET-release assay of Zhang *et al.*[14] was used to determine the inhibition rates and mechanism of action for the putative dimer disrupters. Multiple fluorometric assays were performed, in triplicate, in 384-well, low-volume plates (Corning No. 3676) and read using a SpectraMax M5 from Molecular Devices. The excitation/emission wavelengths of the substrate are 340/490 nm, and we employed a cutoff filter at 475 nm. The final concentration of compound was 40 μM (2 μL of compound diluted in water), and 8 μL of HIVp was diluted into Buffer A or Buffer B for a final concentration range of 1-15nM or 5-50nM, respectively. After 40 minutes of incubation at room temperature,

10 μL of substrate (6.1 μM final concentration) initiated the assay. Fluorescence data was recorded in 1-minute intervals for 20 minutes at 37 ^{o}C. A fourth-order polynomial, non-linear regression was used to fit the kinetic data.[16] Initial velocities were calculated from the regression fit at the sixth minute and data was plotted according to Zhang-Poorman. Least-squares linear regression was used to fit the rearranged data, and dissociative inhibition rates, $K_{i,D}$, were determined from the y-intercepts. The equation used was $K_{i,D} = b_0([I]/(b-b_0))$, where [I] is inhibitor concentration and b and b_0 are the y-intercepts in the presence and absence of inhibitor, respectively. The competitive inhibitor pepstatin A was used as a control to distinguish competitive inhibition in the Zhang-Poorman plots (data not shown).

Discussion and Results

Initial MPS Model from the first 3-ns MD of the HIVp Dimer

The 3-ns pharmacophore model was developed from our previous MD simulation of the dimer, based on the 1HHP crystal structure.[48, 49] Snapshots of the protein were taken at 11 points in the simulation. The dimer was split into its monomers, and the monomers were overlaid. The consensus sites mapped by benzene, ethane, and methanol probes are shown in Figure 2.1.

Within the MOE program,[56] the pharmacophore model was used to screen Michigan's CCG database. The criteria were sequentially relaxed by increasing the radii of the elements and decreasing the number of elements required for a hit. The number of hits for each combination of criteria is given in Table 2.1. A total of 184 hits were

selected and clustered into 75% chemical similarity using MOE, which resulted in several sets that fell into three general shape categories: “spoked”, “rigid linear”, and “flexible linear”. Representative compounds were chosen based on chemical diversity and availability for purchase. A total of 12 compounds, four from each shape category, were purchased and screen *in vitro* in Buffer A. Zhang-Poorman kinetic analysis was performed in Buffer A with a HIVp concentration range from 1-15 nM.

Three of the 12 compounds from the 3-ns model were found to inhibit HIVp. Indeed, Zhang-Poorman kinetic analysis showed the compounds to inhibit through a dissociation of the dimer. The $K_{i,D}$ inhibition rates were 13.4 (± 2.24) µM, 52.4 (± 2.23) µM, and 99.2 (± 4.05) µM for the compounds **1-3**, respectively. The assay results are in Figure 2.2.

We were intrigued by the good affinity of compound **1**, so we purchased four more compounds from the vendor that were most chemically similar to **1**. Of those, two were active with $K_{i,D}$ of 38.6 and 42.9 µM , Figure 2.3. Compounds **1**, **4**, **5**, and **8** (identified by the 30-ns model) constitute the largest family of molecules identified in this work.

A)

Pharmacophore Model Based on the 3-ns MD of the Dimer					
Element Type		x	y	z	RMSD (Å)
1	Aro/phobic	53.5953	45.1589	-2.3560	1.7277
2	Aro/phobic	39.9640	51.0975	2.2804	2.2804
3	Aro/phobic	49.7914	50.5249	6.5096	1.4436
4	Aro	44.9667	54.9316	7.7915	1.7782
5	Aro/phobic	50.4455	54.0263	-7.6683	1.8469
6	Aro/phobic	48.3853	50.9182	-1.6079	1.2943
7	Acc	54.0651	41.5045	-2.5040	1.8872
8	Don	39.4232	52.7636	2.8752	1.2407
9	Don	46.1961	49.5191	-1.1167	1.5183
10	Don/Acc	49.4266	52.9248	-8.0318	1.2849

B)

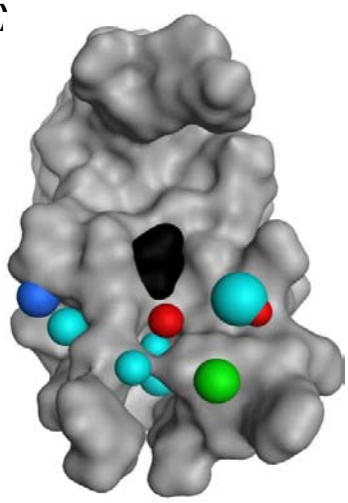

Figure 2.1: Coordinates and spatial representation of the MPS pharmacophore model based on the 3-ns MD simulation of the dimer. A) The elements and coordinates are given relative to the 1HHP crystal structure. B) The pharmacophore sites are shown with 1×RMSD radii, colored green for aromatic, cyan for aromatic/hydrophobic, blue for hydrogen-bond acceptor, and red for hydrogen-bond donor interactions. The purple donor-acceptor site is buried behind the recessed aromatic/hydrophobic site at the center-bottom of the figure and is not visible from this orientation. The protein surface is gray with the surface of the catalytic Asp25 colored black for reference. The flap region is at the top, oriented toward the reader. The active site lies between the flaps and Asp25. The pharmacophore model maps the dimer interface below the active site. The red element in the center of the figure, below Asp25, represents a hydrogen bond to Thr26 which is characteristic of the "fireman's grip" motif of a hydrogen-bonding network shown to contribute to the stability of the dimer.[58]

Table 2.1: Number of hits obtained for *in silico* screening with the 3-ns model as the radii of the elements are increased and number of required features is reduced

Radii Multiplier	**Number of Elements Required for a Matching Hit**					
(#×RMSD)	10 of 10	9 of 10	8 of 10	7 of 10	6 of 10	5 of 10
1	0	0	0	0	***78***	3488
1.33	0	0	0	***1***	–	–
1.67	0	0	2	188	–	–
2	0	0	***33***	–	–	–
2.33	0	5	226	–	–	–
2.67	0	***65***	–	–	–	–
3	***7***	–	–	–	–	–

Total number of compounds considered (in bold italics above): 7+65+33+1+78=184 with duplicates

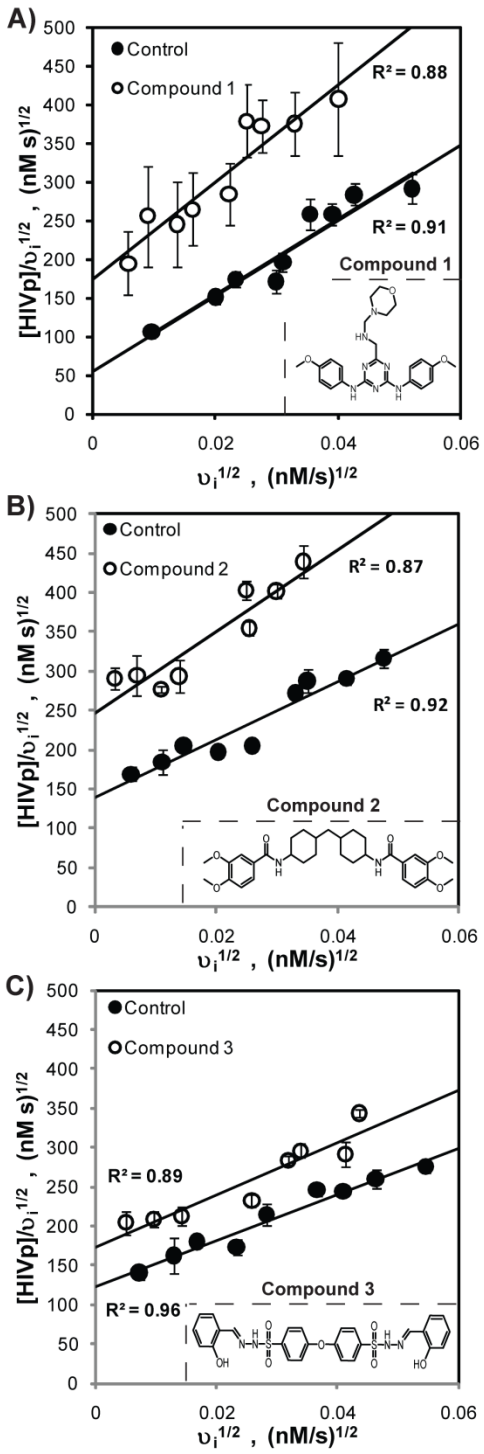

Figure 2.2: Kinetic analysis of inhibitors identified with the 3-ns model of the dimer. A screen of 12 compounds resulted in 3 inhibitors of HIVp. The mechanism of dimerization inhibition was verified by Zhang-Poorman assay showing non-intersecting, near-parallel lines with the no-inhibitor control. A) Compound **1**, $K_{i,D}$ = 13.4 μM. B) Compound **2**, $K_{i,D}$ = 52.4 μM. C) Compound **3**, $K_{i,D}$ = 99.2 μM.

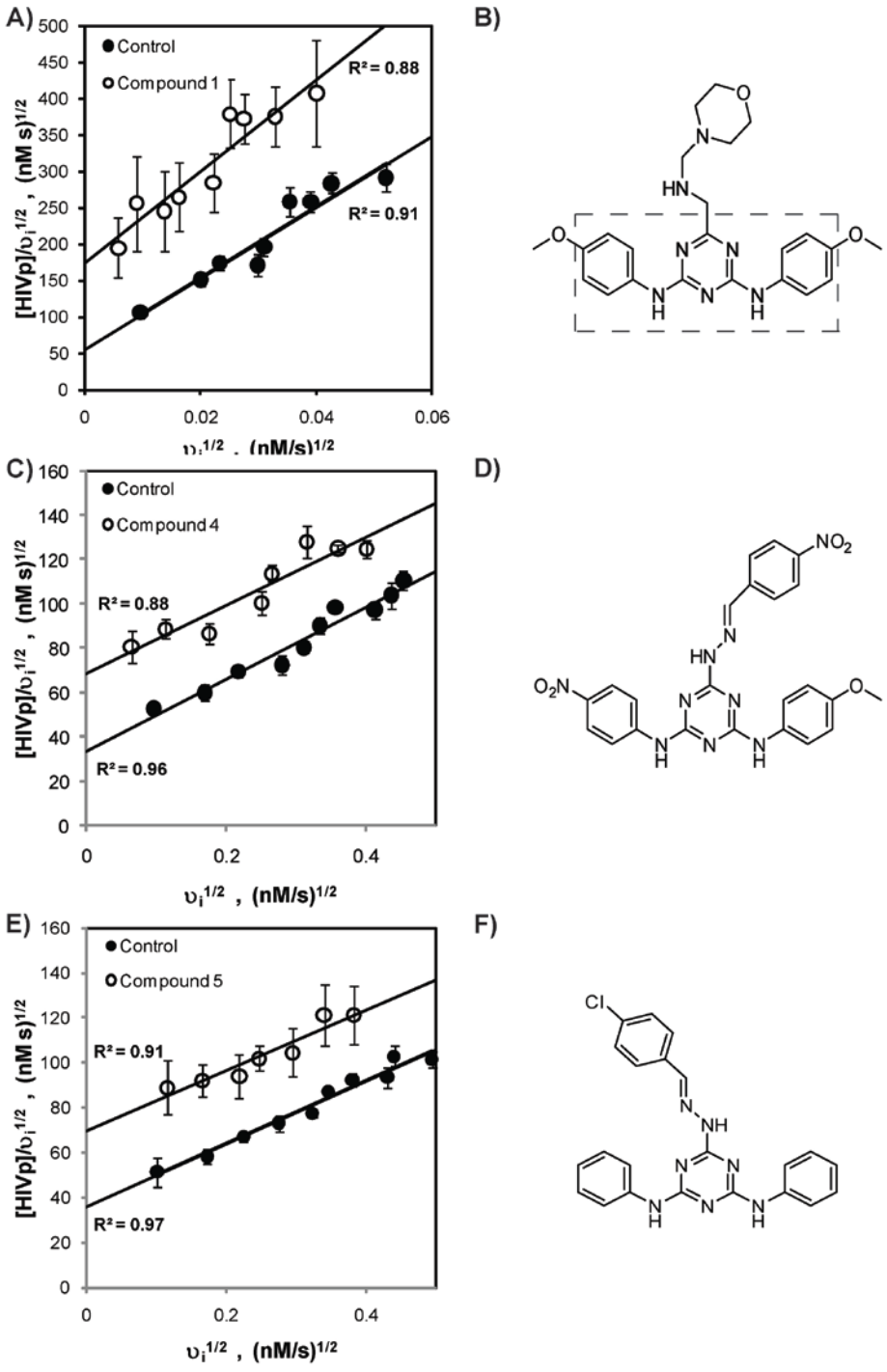

Figure 2.3: Dimerization inhibitors with similar scaffolds. A and B) The Zhang Poorman analysis and structure of compound **1**. D and F) Compounds **4** and **5** are chemically similar to compound **1**. C and E) Kinetic analysis showed compounds **4** and **5** to be dissociative inhibitors with $K_{i,D}$ values of 38.6 and 42.9 μM, respectively.

MPS Model from the 30-ns MD of the HIVp Dimer

The initial model was based on a very short MD simulation, pursued simply as a proof-of-concept. That simulation is much too short by current standards. Furthermore, extending the length of MD simulations can improve the results for MPS pharmacophore models, as reported by Meagher *et al.*[48, 49] and Lerner *et al.*[59] A 30-ns model, shown in Figure 2.4, was created by extending the previous MD simulation and using conformations representative of much greater conformational sampling.

A)

Pharmacophore Model Bases on the 30-ns MD of the Dimer					
Element Type		x	y	z	RMSD (Å)
1	Aro	55.3167	39.6753	-3.2984	0.9308
2	Aro/phobic	53.6327	45.5824	-2.6711	1.0854
3	Aro/phobic	52.6310	48.0631	-1.1350	1.2316
4	Aro/phobic	49.5117	50.0135	6.9862	1.0119
5	Aro/phobic	41.2878	50.7591	2.8420	1.3449
6	Don	56.4128	56.4714	10.0085	1.4511
7	Don	57.0321	49.3395	1.6845	0.9405
8	Don	54.2489	44.4524	-3.9576	1.2007
9	Acc	54.7309	40.3557	-3.3089	1.2785
10	Acc	53.4296	37.9627	-3.9918	0.9895
11	Don	47.9795	36.7125	-1.0512	0.9032

B)

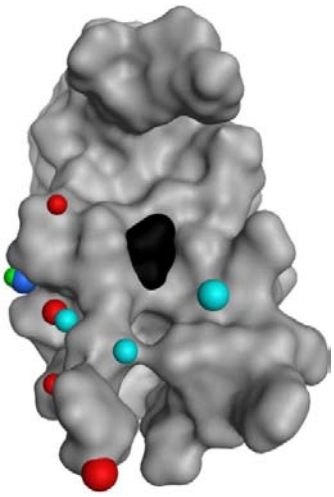

Figure 2.4: Coordinates and spatial representation of the MPS pharmacophore model based on the 30-ns MD simulation of the dimer. A) The coordinates are given relative to the 1HHP crystal structure. B) The pharmacophore sites are shown with 1×RMSD radii, colored green for aromatic, cyan for aromatic/hydrophobic, blue for hydrogen-bond acceptor, and red for hydrogen-bond donor interactions. The protein surface is gray with the surface of the catalytic Asp25 colored black for reference. The orientation is the same as in Figure 2.1.

The 30-ns model was used to filter the CCG library with *in silico* screening in MOE. The number of hits under different selection criteria is given in Table 2.2. Compounds with MW >500 were eliminated, and a total of 108 compounds were considered. The small molecules were clustered by 75% chemical similarity and 65% overlap in MOE. A total of 55 compounds were purchased and screened *in vitro* in Buffer B. The 30-ns pharmacophore model produced 7 compounds characterized as dimerization inhibitors by Zhang-Poorman in Buffer B, with a HIVp concentration range of 5-50 nM. The kinetic analyses of compounds **6**-**12** are reported in Figure 2.5. The $K_{i,D}$ range for these seven compounds was 18.4-86.0 μM and the median $K_{i,D}$ was 58.0 μM. All dissociative inhibitors from this study are summarized in Table 2.3.

Table 2.2: Number of hits obtained for *in silico* screening with the 30-ns model as the radii of the elements are increased and number of required features is reduced

Radii Multiplier	**Number of Elements Required for a Matching Hit**					
(#×RMSD)	11 of 11	10 of 11	9 of 11	8 of 11	7 of 11	6 of 11
1	0	0	0	0	0	490
1.33	0	0	0	0	***20***	2291
1.67	0	0	0	0	157	7152
2	0	0	0	7	603	–
2.33	0	0	0	30	1397	–
2.67	0	0	0	***119***	2450	–
3	0	0	***1***	316	3993	–

Total number of compounds considered (in bold italics above): 1+119+20=140 with duplicates and no MW limit

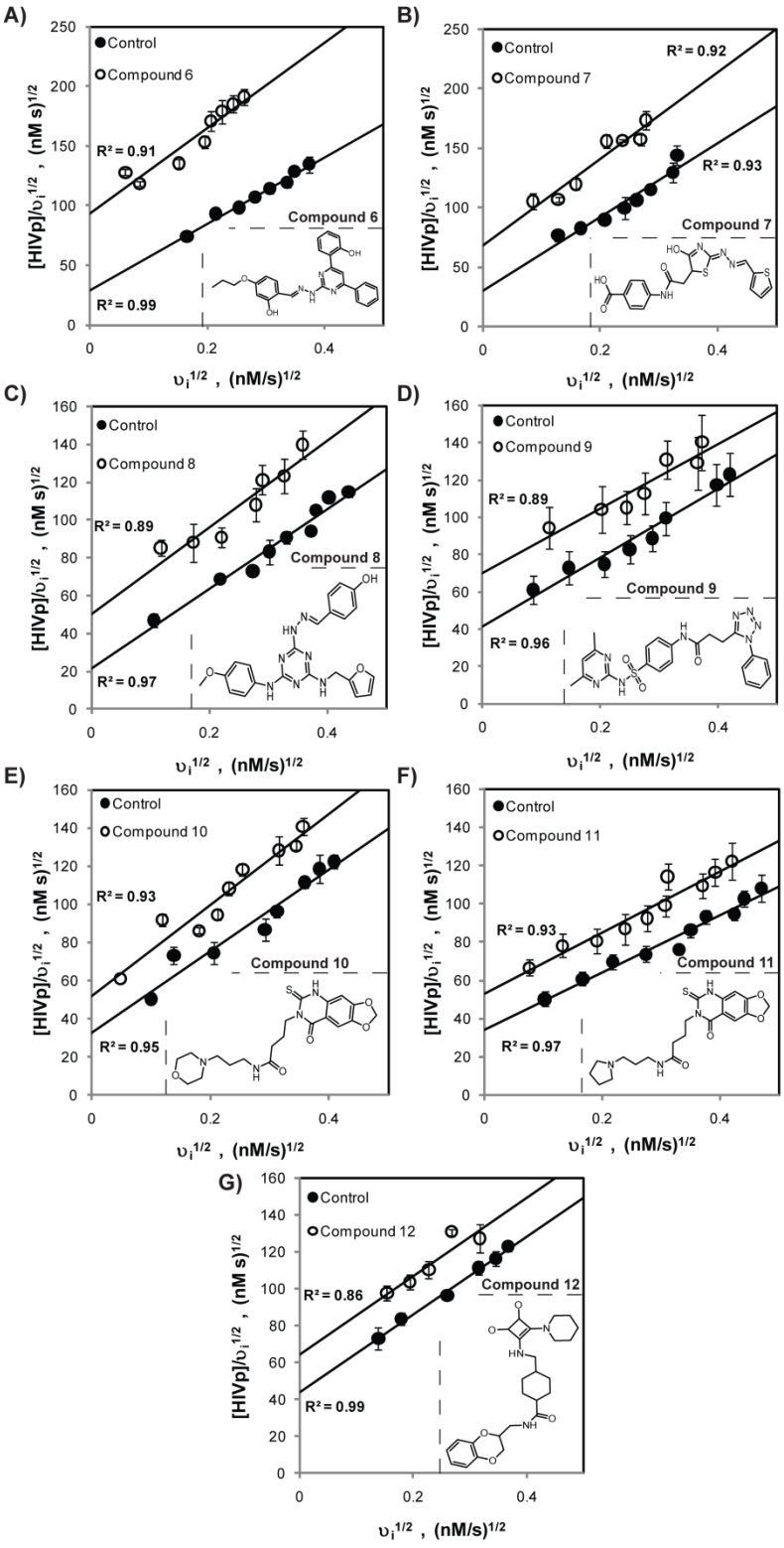

Figure 2.5: Kinetic analysis of inhibitors identified by 30-ns model. The Zhang-Poorman analysis for compounds **6-12** confirm a dissociative mechanism of inhibition. A-G) with $K_{i,D}$ ranging 18.4-86.0 μM, see Table 2.3.

Table 2.3: Summary of inhibitors identified by the various MPS models.

Compound	ChemBridge Catalog Number	ChemDiv Catalog Number	Inhibition $K_{i,D}$ (μM)	MW	Ligand Efficiency (-log $K_{i,D}$ / MW)
From the 3-ns model					
1		5594-3017	13.4	451.2	0.0108
2	5549891		52.4	538.3	0.0080
3	5472583		99.2	566.1	0.0071
4		1071-0033	38.6	501	0.0088
5		8002-9140	42.9	415	0.0105
From the 30-ns model					
6	5341525		18.4	440.2	0.0108
7	6138348		31.8	402.1	0.0112
8	5533247		36.6	431.5	0.0103
9	6249591		58.0	478.2	0.0089
10		K284-5225	69.1	434.2	0.0096
11		K284-5234	72.4	418.2	0.0099
12		C906-0998	86.0	469.3	0.0087
From the Xtal model (presented further below)					
none					
From the consensus model (presented further below)					
13		C597-0141	100.7	438.2	0.0083
14		5056-0019	109.3	384.1	0.0103
15		3334-5637	123.6	422.1	0.0093
16	5565207		165.7	483.1	0.0078

For comparison, in Lee and Chmielewski the ligand efficiencies for their best and smallest inhibitor are 0.0048 and 0.0076, respectively.[23] The natural products have ligand efficiencies of 0.0103 for schisanlactone A and a range of 0.0083-0.0105 for the simplified derivatives of didemnaketal A.

Pharmacophore Model Generated from the 1HHP Crystal Structure

Another pharmacophore model was created based solely on the complementary chain inherent to the 1HHP crystal structure (our so-called Xtal model). Only a single monomer is present in the crystal structure, and its symmetry partner for the dimer (chain B) must be created by C_2 rotation. The Xtal pharmacophore was created in a way that represents traditional approaches to structure-based drug design. As outlined in the introduction, many groups have created dimer inhibitors by building mimics of the termini of Chain B. Here, we focused on the contacts between the cores of the

monomers, rather than the β-sheet region. The most prominent characteristics of the interface were identified through inspection of the hydrogen-bonding patterns and the curvature of the hydrophobic vdw surface. Highly curved surfaces indicated regions where deeper sub-pockets in Chain A were complemented by Chain B. The atoms selected for the centers of the hydrogen-bonding elements were: the backbone oxygen of Leu5, the backbone oxygen of Trp6, the backbone oxygen of Leu24, the backbone nitrogen of Thr26, and the side-chain hydroxyl oxygen of Thr26. These hydrogen-bonding elements encode the "fireman's grip" motif which has been shown to contribute to the stability of the dimer.[58] The chemical properties of the elements were assigned accordingly (i.e., the backbone oxygens were defined as hydrogen-bond acceptors, the nitrogens were donors, and the hydroxyl group was either/both). A 1.4-Å radius was used for all hydrogen-bonding elements of the model. The Cγ of Leu5 and Cγ of Leu97 were chosen for hydrophobic/aromatic centers, and a 2.0-Å radius was used to roughly parallel the radii used for hydrophobic carbons in molecular mechanics. The Xtal model is given in Figure 2.6.

The Xtal pharmacophore model was used to computationally screen the CCG library in the same manner as the MD-based models, Table 2.4. Of the 35 compounds considered, 24 were purchased based on availability and MW. All were screened *in vitro* in Buffer B. None of compounds from the Xtal pharmacophore model inhibited HIVp protease by either a competitive or dissociative mechanism. This underscores the importance of incorporating protein flexibility in structure-based inhibitor discovery.

A)

Pharmacophore Model Based on the Dimer Interface in the 1HHP Crystal Structure (Xtal)					
Element Type		x	y	z	RMSD (Å)
1	Aro/phobic	49.121	52.222	-1.577	2.0
2	Aro/phobic	53.176	45.655	-2.295	2.0
3	Acc	55.105	39.283	-3.738	1.4
4	Acc	57.041	39.283	-3.786	1.4
5	Acc	48.859	45.270	3.092	1.4
6	Don	45.873	47.351	1.794	1.4
7	Don/Acc	46.467	48.515	-0.579	1.4

B)

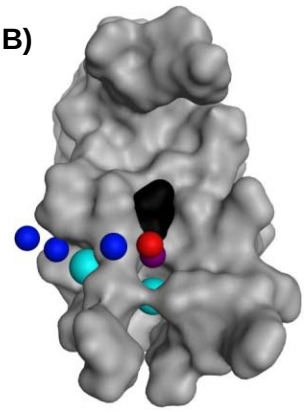

Figure 2.6: The coordinates and spatial representation based on the symmetry partner in PDB 1HHP. A) The coordinates are given relative to 1HHP. B) The pharmacophore sites are shown with 1×RMSD radii, colored cyan for aromatic/hydrophobic, blue for hydrogen-bond acceptor, red for hydrogen-bond donor interactions and purple for hydrogen-bond acceptor/donor. The protein surface is gray with the surface of the catalytic Asp25 colored black for reference. The orientation is the same as in Figures 2.1 and 2.3.

Table 2.4: Number of hits obtained for *in silico* screening against the Xtal model as the radii of the elements are increased and number of required features is reduced

Radii Multiplier (#×RMSD)	**Number of Elements Required for a Matching Hit** 7 of 7	6 of 7	5 of 7
1	2	309	3592+
1.33	***35***	1978	
1.67	507	5085	
2	880	2091+	
2.33	1808		
2.67	3066 +		

Total number of compounds considered (in bold italics above): 35 with duplicates and no MW limit. A plus sign (+) indicates the pharmacophore search was terminated early.

Multi-Drug Resistance HIVp

Isolated from clinical HIVp infected patients, the multi-drug resistant (MDR) form of HIVp evades the efficacy of all inhibitors in clinical use.[60] Bannwarth *et al.* tested the inhibition of competitive and dissociative inhibitors against wild-type and MDR forms of HIVp.[25, 61] When compared to the competitive inhibitors, the dissociative inhibitors inhibited comparatively both forms of HIVp. Since Bannwarth et al. demonstrated the comparative success in MDR and wild-type HIVp using peptidomimetic HIVp inhibitors, we wanted to demonstrate similar results for our PH4-derived small molecule dissociative inhibitor.

The four best inhibitors from the 30-ns model, compounds **6-9**, were selected for Zhang-Poorman characterization against MDR-HIVp. Table 2.5 compares the Zhang-Poorman results between the wild-type and MDR forms of HIVp. Although the inhibition rates for the MDR form was approximately 2 to 4-folds greater than the wild-type HIVp, that figure compares favorably to competitive inhibitors. As reported in Banwarth *et al.*, competitive inhibitors can lose several orders of magnitude in inhibition rates.[25, 61]

Table 2.5: Comparison of Zhang-Poorman characterized inhibition rates between wild-type and MDR forms of HIVp.

Compound	**Zhang-Poorman assessed inhibition rates ($K_{i,d}$)**		
Number	Wild-type (μM)	MDR (μM)	MDR:Wilt-type $K_{i,d}$ Ratio
6	18.4	41.5	2.26
7	31.8	116.8	3.67
8	36.6	77.3	2.11
9	58.0	155.0	2.67

MPS Model Based on the MD Simulation of Monomeric HIVp

Our simulation of the HIVp dimer did not allow for any conformational rearrangement inherent to the monomeric state. To incorporate this important physical information into our study, a 35-ns MD simulation of the monomer was initiated from the same crystallographic coordinates (1HHP). The motion of the N- and C-termini was measured using RMSD from the equilibrated structure, shown in Figure 2.7. The high degree of sampling for the termini was also observed by Levy and Caflisch[7] and by Yan *et al.*[8] in their simulations of the HIVp monomer. Yan *et al.* described the N- and C-termini forming a stable β-sheet in their simulations at 300 K. The β-sheet is questionable, as Caflisch *et al.* did not report any stable secondary structure in their simulations.[62] However, the two studies did agree that the termini were highly dynamic.

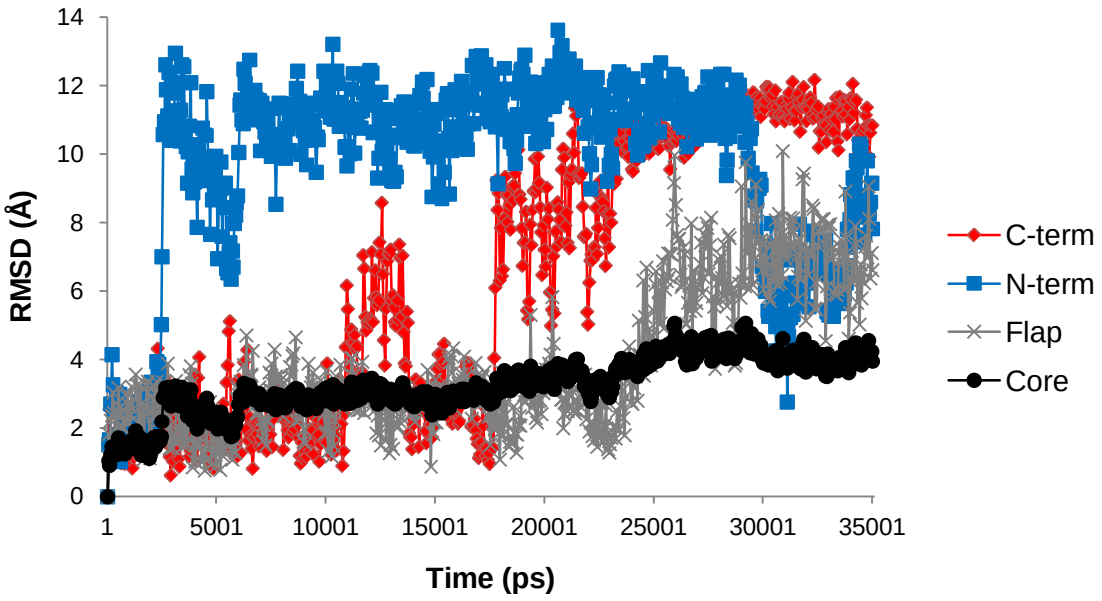

Figure 2.7: RMSD of the monomer of HIVp. The RMSD of the Cα over the course of the MD simulation is shown in black for the core of the monomer (residues 5-44 and 56-94), in blue for the N-terminus (residues 1-4), in red for the C-terminus (residue 95-99) and in gray for the flap (45-55). The average RMSD was 1.2, 9.6, 6.7, and 3.9 Å for the four regions, respectively.

Our own simulation shows the termini of the monomer significantly reorder and are much more variable than the relatively-stable core region. The amplitude of motion at the termini ranges from 6 to 10 Å and appears to be more disordered than the mobile flap region. The degree of flexibility is comparable to that observed by Yan *et al.* and to Levy *et al.* in their simulated monomeric protease.[7, 8]

We calculated the secondary structure of over the course of our simulation using the DSSP program, and the results are presented in Figure 2.8. The simulation of the monomer showed no stable secondary structure in the termini, agreeing with the findings of Levy *et al.*[62]

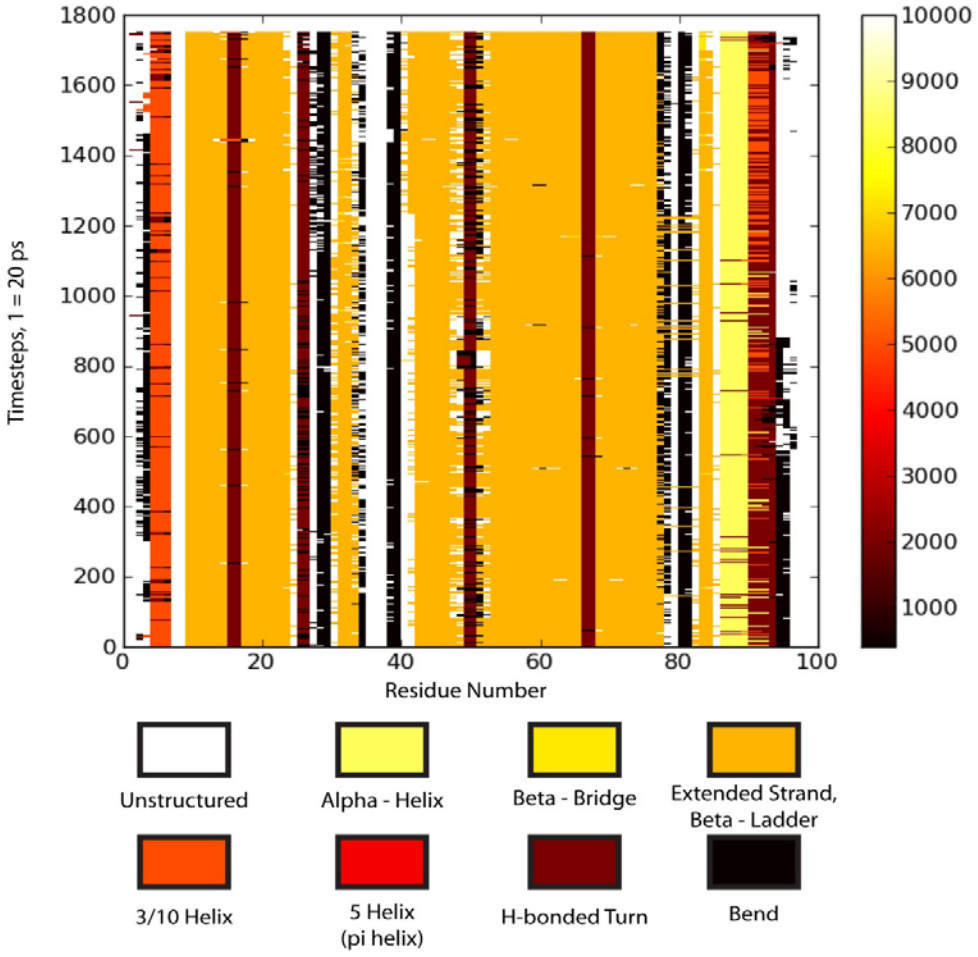

Figure 2.8: Secondary structure of the monomer of HIVp over the course of the MD simulation. The secondary structure was characterized using the DSSP program. The termini of the 1HHP monomer remained largely unstructured.

To design a pharmacophore model from the MD simulation of the monomer, the trajectory was clustered into ten representative conformations using the ptraj module in the AMBER program. The ten representative structures were used for our MPS procedure. The procedure was the same as used for the other pharmacophore models, except that the Cα of Thr26 was selected as the center of the 15-Å radius sphere used for initial placement of the small chemical probes. The model resulting from the 35-ns monomer simulation is described in Figure 2.9. The CCG library was screened against

the pharmacophore model, see Table 2.6. At this time, 125 compounds are being considered, and another lab member will test the purchased compounds.

A)

Pharmacophore Model Based on the 35-ns MD of the HIVp monomer					
Element Type		x	y	z	RMSD (Å)
1	Aro	37.6358	48.8882	0.3348	0.7108
2	Aro	48.8924	35.4444	-0.5960	0.5956
3	Aro	50.3660	39.2816	1.9372	0.9924
4	Aro/phobic	54.8234	43.3583	-1.8584	0.9539
5	Acc	50.8056	54.9348	0.3424	1.308
6	Acc	53.5527	38.1876	-4.6026	1.0672
7	Acc	54.6076	43.3746	-2.6758	1.2281
8	Acc	38.3942	48.5002	0.1560	1.1402
9	Don	48.4316	36.5660	-0.6022	1.0618
10	Don/Acc	56.1486	46.5532	0.5000	1.3553

B)

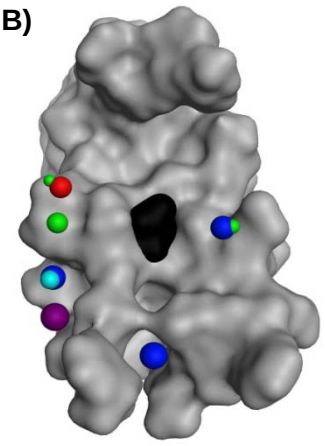

Figure 2.9: Coordinates and spatial representation of the MPS pharmacophore model based on the 35-ns MD simulation of the monomer. A) The coordinates are given relative to 1HHP. B) The pharmacophore sites are colored green for aromatic, cyan for aromatic/hydrophobic, blue for hydrogen-bond acceptor, red for hydrogen-bond donor, and purple for hydrogen-bond acceptor/donor interactions. The protein surface is gray with the surface of the catalytic Asp25 colored black for reference. The orientation is the same as in the figures of the other pharmacophore models above.

Table 2.6: Number of hits obtained for *in silico* screening with the 35-ns Monomer model as the radii of the elements are increased and number of required features is reduced

Radii Multiplier	**Number of Elements Required for a Matching Hit**				
(#×RMSD)	10 of 10	9 of 10	8 of 10	7 of 10	6 of 10
1	0	0	0	0	7
1.33	0	0	0	0	***77***
1.67	0	0	0	4	60
2	0	0	0	***47***	205
2.33	0	0	0	235	4719
2.67	0	0	0	820	8387
3	0	0	***1***	1830	12024

Total number of compounds considered (in bold italics above): 1+47+77=125 with duplicates and no MW limit

A *"Common Element" Pharmacophore Model*

The "common" pharmacophore was developed to represent a core number of elements seen in both MPS models from MD simulations of the dimer and the monomer. It is possible that molecules which compliment both conformational states of HIVp would have an entropic advantage. The model from the 30-ns MD of the dimer was used as the basis for the common pharmacophore. The two models were compared, and elements were eliminated from the 30-ns model of the dimer if the element was not observed in both the dimer and monomer pharmacophores. A hydrogen-bonding element near Asp29 was eliminated due to the distance from the other pharmacophore elements. The common pharmacophore model is shown in Figure 2.10.

A)

Pharmacophore Model of the Elements Common to both the Dimer and Monomer MPS Models					
Element Type		x	y	z	RMSD (Å)
1	Aro/phobic	53.6327	45.5824	-2.6711	1.0854
2	Aro/phobic	52.6310	48.0631	-1.1135	1.2316
3	Aro/phobic	49.5117	50.0135	6.9862	1.0119
4	Aro/phobic	41.2878	50.7591	2.8420	1.3449
5	Don	57.0321	49.3395	1.6845	0.9405
6	Don	54.2489	44.4524	-3.9576	1.2007

B)

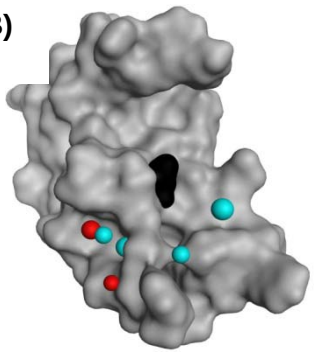

Figure 2.10: The "common" model based on 30-ns model of the dimer with elements eliminated when not observed in the monomer model. A) The coordinates are given relative to 1HHP. B) The consensus pharmacophore model is shown. The view is rotated slightly so that all four aromatic/hydrophobic sites are visible.

The common pharmacophore was then used to filter the CCG compound library (Table 2.7), for a total of 208 total putative dissociative inhibitors (duplicates and high-MW ligands removed). After clustering the hits based on chemical similarity, a set of

115 compounds from this library were purchased and screened *in vitro* for inhibition in Buffer B. Compounds found to inhibit HIVp were further characterized by the Zhang-Poorman kinetic assay in Buffer B, with a HIVp concentration range of 5-50 nM. Four compounds were characterized to be dissociative inhibitors. The $K_{i,D}$ for compounds **13-16** are 100.7 (± 4.3) µM, 109.3 (± 8.1) µM, 123.6 (± 6.8) µM and 165.7 (± 10.4) µM, respectively. The Zhang-Poorman analysis is shown in Figure 2.11.

None of the compounds from the resulting common model had low micromolar $K_{i,D}$. The common model also resulted in fewer dissociative inhibitors compared to the 30-ns model. Despite being developed from the 30-ns model and filtering the same three-dimensional small molecule library, the common model performed worse than the 30-ns model. However, both models were able to selectively discriminate against competitive inhibitors.

Table 2.7. Number of hits obtained for *in silico* screening with the common model as the radii of the elements are increased and number of required features is reduced

Radii Multiplier (#×RMSD)	**Number of Elements** 6 of 6	5 of 6
1	0	41
1.33	0	***161***
1.67	0	492
2	6	1458
2.33	20	2911
2.67	***72***	5139
3	254	7920

Total number of compounds considered (in bold italics above): 72+161=233 with duplicates and no MW limit

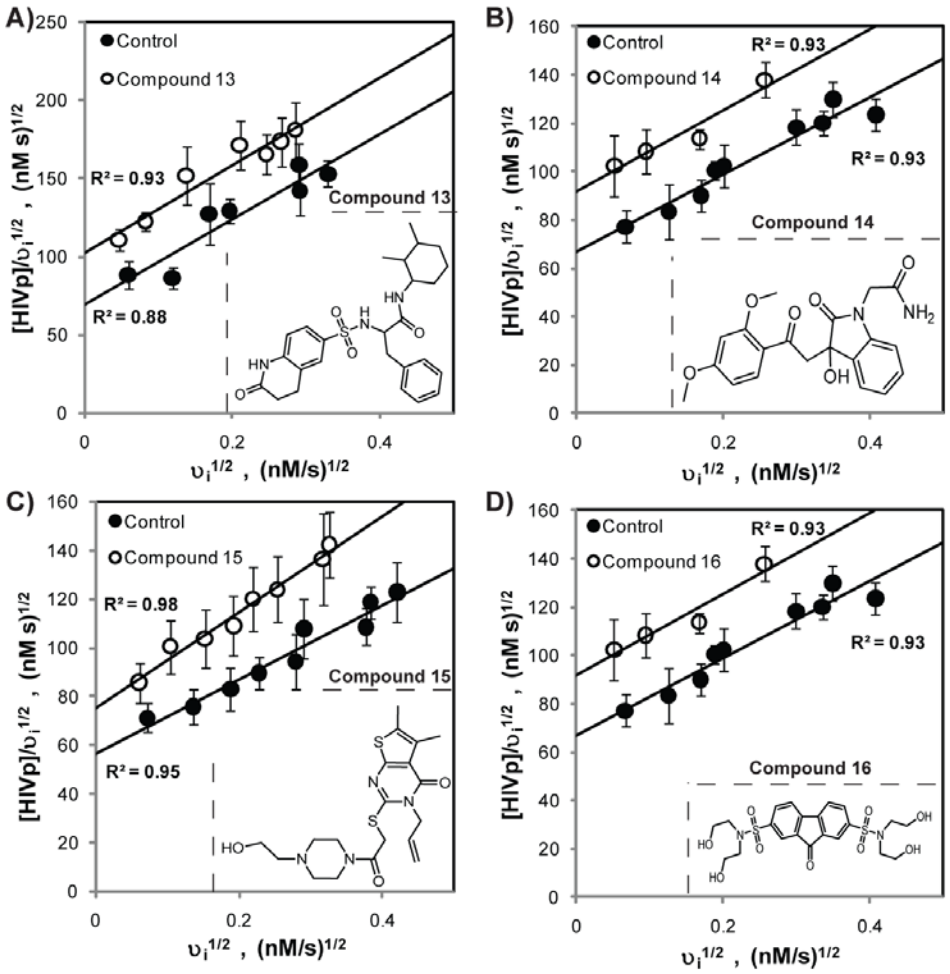

Figure 2.11: Consensus pharmacophore results. A screen of 115 compounds resulted in four compounds that inhibited the dimerization of HIVp. The mechanism of inhibition was characterized by the Zhang-Poorman assay. A) Compound **13**, $K_{i,D}$ = 100.7 μM. B) Compound **14**, $K_{i,D}$ =109.3 μM. C) Compound **15**, $K_{i,D}$ =123.6 μM. D) Compound **16**, $K_{i,D}$ =165.7 μM.

Conclusion

From simulations of the 1HHP dimer and monomer, we developed several dynamic receptor-based pharmacophores capable of identifying dimerization inhibitors from a virtual screen of a small molecule library. Each pharmacophore was able to discriminate dissociative inhibitors over competitive inhibitors. Although none of the

small molecules had sub-micromolar inhibition rates, they are approximately half the molecular weight of current bivalent compounds. The small molecules are also non-peptidomimetic. Moreover, we showed the small molecules to be effective against the wild-type and MDR strains of HIVp. Small molecules targeting the dimer interface are desirable to circumvent existing resistance seen in clinical therapy.

The number of higher affinity inhibitors can be limited the small molecule library, the MPS-derived pharmacophore description of the dimer interface, or the inherent difficulty in attacking the HIVp dimer interface with a small molecule. The MPS pharmacophore models can be used to filter other small molecule libraries to search for better inhibitors. Whether the dimer interface is well suited for inhibition by small molecules has not been determined. The compounds in this study have undergone no optimization to better target the dimer interface. The models can possibly be improved to incorporate more information from PPI "hot spots". Hot spots are patches along the protein surface that contribute disproportionally to oligomerization.[63]

This study demonstrates the MPS procedure can be used to target PPIs in a case where any "pocket" of the PPI is ill-defined. We have previously targeted the p53-HDM2 PPI and identified several novel chemical scaffolds.[55] However, the p53-MDM2 PPI is smaller (660 Å and 809 Å for $MDM2_{25\text{-}109}$ and $p53_{17\text{-}29}$, respectively) compared to the HIVp dimer interface.[64] The p53-MDM2 PPI also has a deeper binding cleft. This study demonstrates the robustness of the MPS procedure against a more difficult PPI. With an estimated 650,000 PPIs in the human interactome, there is a significant probability that a number of druggable PPIs exist.[65] It is exciting that the MPS procedure can be used to facilitate the pursuit of inhibitors to target PPIs.

Acknowledgements

I would like to thank Drs. Anna Bowman, Kelly Damm and Michael Lerner for their assistance in the MPS procedure. Thanks to Dr. Kristin Meagher for initiating the HIVp dimer simulation and Peter M. Ung for help developing assay conditions. The reported work was funded by the NIH (Grant GM065372) and by the Center for Computational Medicine and Bioinformatics. I thank the University of Michigan's Molecular Biophysics Training Program for generous support (GM008270). Thanks to Martha Larsen and the University of Michigan's Center for Chemical Genomics for donated small molecules for initial testing. Computational support was provided by the University of Michigan's Center for Advanced Computing. A special thanks to Allen Bailey for maintaining the computers used in this work.

Supporting Information

A "simplified" model of the 3-ns model is given in the appendix.

References

1. Navia, M.A.; Fitzgerald, P.M.; McKeever, B.M.; Leu, C.T.; Heimbach, J.C.; Herber, W.K.; Sigal, I.S.; Darke, P.L.; Springer, J.P., Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* **1989**, *337*, 615-620.

2. Wlodawer, A.; Erickson, J.W., Structure-based inhibitors of HIV-1 protease. *Annu Rev Biochem* **1993**, *62*, 543-585.

3. Tong, L.; Pav, S.; Pargellis, C.; Do, F.; Lamarre, D.; Anderson, P.C., Cystal structure of human immunodeficiency virus (HIV) type 2 protease in complex with a reduced amide inhibitor and comparison with HIV-1 protease structures. *Proc Natl Acad Sci* **1993**, *90*, 8387-8391.

4. Jones, S.; Thornton, J.M., Principles of protein-protein interaction. *Proc Natl Acad Sci* **1996**, *93*, 13-20.

5. Todd, M.J.; Semo, N.; Freire, E., The structural stability of the HIV-1 protease. *JMB* **1998**, *283*, 475-488.

6. Weber, I.T, Comparison of the crystal structures and intersubunit interactions of Human Immunodeficiency and Rous Sarcoma Virus proteases. *J Biol Chem* **1990**, *265*, 10492-10496.

7. Levy, Y. and Caflisch, A., Flexibility of monomeric and dimeric HIV-1 protease. *J Phys Chem B* **2003**, *107*, 3068-3079.

8. Yan, M.-C.; Sha, Y.; Wang, J.; Xiong, X.-Q.; Ren, J.-H.; Cheng, M.-S., Molecular dynamics simulations of HIV-1 protease monomer: assembly of N-terminus and C-terminus into b-sheet in water solution. *Proteins* **2008**, *70*, 731-738.

9. Xie, D.; Gulnik, S.; Gustchina, E.; Yu, B.; Wei, Shao; Qoronfleh, W.; Nathan, A.; Erickson, J.W., Drug resistance mutations can affect dimer stability of HIV-1 protease at neutral pH. *Protein Sci* **1999**, *8*, 1702-1707.

10. Bannwarth, L.; Reboud-Ravaux, M., An alternative strategy for inhibiting multidrug-resistant mutants of the dimeric HIV-1 protease by targeting the subunit interface. *Biochem Soc Trans* **2007**, *35*, 551-554.

11. Das, A.; Mahale, S.; Prashar, V.; Bihani, S.; Ferrer, J.-L.; Hosur, M.V., X-ray snapshot of HIV-1 protease in action: observation of tetrahedral intermediate and short ionic hydrogen bond SIHB with catalytic aspartate. *J Am Chem Soc* **2010**, *132*, 6366-6373.

12. Gustchina, A. and Weber, I.T., Comparative analysis of the sequences and structures of HIV-1 and HIV-2 proteases. *Proteins* **1991**, *10*, 325-339.

13. Wensing, A.M.; van de Vijver, D.A.; Angarano, G.; Asjö, B.; Balotta, C.; Boeri, E.; Camacho, R.; Chaix, M.L.; Costagliola, D.; De Luca, A.; Derdelinckx, I.; Grossman, Z.; Hamouda, O.; Hatzakis, A.; Hemmer, R.; Hoepelman, A.; Horban, A.; Korn, K.; Kücherer, C.; Leitner, T.; Loveday, C.; MacRae, E.; Maljkovic, I.; de Mendoza, C.; Meyer, L.; Nielsen, C.; Op de Coul, E.L.; Ormaasen, V.; Paraskevis, D.; Perrin, L.; Puchhammer-Stöckl, E.; Ruiz, L.; Salminen, M.; Schmit, J.C.; Schneider, F.; Schuurman, R.; Soriano, V.; Stanczak, G.; Stanojevic, M.; Vandamme, A.M.; Van Laethem, K.; Violin.,M.; Wilbe, K.; Yerly, S.; Zazzi, M.; Boucher, C.A.; SPREAD Programme, Prevalence of drug-resistant HIV-1 variants in untreated individuals in Europe: implications for clinical management. *J Infect Dis* **2005**, *192*, 958-966.

14. Zhang, Z.-Y.; Poorman, R.A.; Maggiora, L.L.; Heinrikson, R.L.; Kezdy, F.J., Dissociative inhibition of dimeric enzymes. *J Biol Chem* **1991**, *266*, 15591-15594.

15. Boggetto, N.; Reboud-Ravaux, M., Dimerization inhibitors of HIV-1 protease. *Biol Chem* **2002**, *383*,

16. Schramm, H.J.; Boetzel, J.; Buttner, J.; Fritsche, E.; Gohring, W.; Jaeger, E.; Konig, S.; Thumfart, O.; Wenger, T.; Nagel, N.E.; Schramm, W., The inhibition of human immunodeficiency virus protease by 'interface peptides'. *Antiviral Res* **1996**, *30*, 155-170.

17. Dumond, J.; Boggetto, N.; Schramm, H.J.; Schramm, W.; Takahashi, M.; Reboud-Ravaux, M., Thyroxine-derivatives of lipopeptides: bifunctional dimerization inhibitors of human immunodeficiency virus-1 protease. *Biochem Pharmacol* **2003**, *65*, 1097-1102.

18. Schramm, H.J.; de Rosny, E.; Reboud-Ravaux, M.; Buttner, J.; Dick, A.; Schramm, W., Lipopeptides as dimerization inhibitors of HIV-1 protease. *Biol Chem* **1999**, *380*, 593-596.

19. Frutos, S.; Rodrigues-Mias, R.A.; Madurga, S.; Collinet, B.; Reboud-Ravaux, M.; Ludevid, D.; Giralt, E., Disruption of the HIV-1 protease dimer with interface peptides: structural studies using NMR spectroscopy combined with [2-^{13}C]-Trp selective labeling. *Biopolymers* **2007**, *88*, 164-173.

20. Caflisch, A.; Schramm, H.J.; Karplus, M., Design of dimerization inhbitors of HIV-1 aspartic proteinase: A computer-based combinatorial approach. *J Comput Aided Mol Des* **2000**, *14*, 161-179.

21. Bowman, M.J.; Chmielewski, J., Novel strategies for targeting the dimerization interface of HIV protease with cross-linked interfacial peptides. *Biopolymers* **2002**, *66*, 126-133.

22. Shultz, M.D.; Bowman, M.J.; Ham, Y.-W.; Zhao, X.; Tora, G.; Chmielewski, J., Small-molecule inhibitors of HIV-1 protease dimerization derived from cross-linked interfacial peptides. *Agnew Chem Int Ed* **2000**, *36*, 2710-2713.

23. Lee, S.-G.; Chmielewski, J., Rapid synthesis and in situ screening of potent HIV-1 protease dimerization inhibitors. *Chem Biol* **2006**, *13*, 421-426.

24. Bouras, A.; Boggetto, N.; Benatalah, Z.; de Rosny, E.; Sicsic, S.; Reboud-Ravaux, M., Design, synthesis, and evaluation of conformationally constrained tongs, new inhibitors of HIV-1 protease dimerization. *J Med Chem* **1999**, *42*, 957-962.

25. Bannwarth, L.; Kessler, A.; Pethe, S.; Collinet, B.; Merabet, N.; Boggetto, N.; Sicsic, S.; Reboud-Ravaux, M.; Ongeri, S., Molecular tongs containing amino acid mimetic fragments: new inhibitors of wild-type and mutated HIV-1 protease dimerization. *J Med Chem* **2006**, *49*, 4657-4664.

26. Merabet, N.; Dumond, J.; Collinet, B.; Van Baelinghem, L.; Boggetto, N.; Ongeri, S.; Ressad, F.; Reboud-Ravaux, M.; Sicsic, S., New constrained "molecular tongs" designed to dissociate HIV-1 protease dimer. *J Med Chem* **2004**, *47*, 6392-6400.

27. Abell, Andrew, Advances in Amino Acid Mimetics and Peptidomimetics; Jai Press Ltd.; 1997

28. Fan, X.; Flentke, G.R.; Rich, D.H., Inhibition of HIV-1 protease by a subunit of Didemnaketal A. *J Am Chem Soc* **1998**, *120*, 8893-8894.

29. Dine, R.S.E.; Halawany, A.M.E.; Chao-Mei, M.; Hattori, M., Inhibition of the dimerization and active site of HIV-1 protease by secondary metabolites from the Vietnamese mushroom *Ganoderma colossum*. *J. Nat. Prod.* **2009**, *72*, 2019-2023.

30. Quere, L.; Wenger, T.; Schramm, H.J., Triterpenes as potential dimerization inhibitors of HIV-1 protease. *Biochem Biophys Res Commun* **1996**, *2*,

31. Arkin, M.R.; Wells, J.A., Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat Rev Drug Discov* **2004**, *3*, 301-317.

32. Oltersdorf, T; Elmore, S.W.; Shoemaker, A.R.; Armstrong, R.C.; Augeri, D.J.; Belli, B.A.; Bruncko, M.; Deckwerth, T.L.; Dinges, J.; Hajduk, P.J.; Joseph, M.K.; Kitada, S.; Korsmeyer, S.J.; Kunzer, A.R.; Letai, A.; Li, C.; Mitten, M.J.; Nettesheim, D.G.; Ng, S.; Nimmer, P.M.; O'Conner, J.M.; Oleksijew, A.; Petros, A.M.; Reed, J.C.; Shen, W.; Tahir, S.K.; Thompson, C.B.; Tomaselli, K.J.; Wang, B.; Wendt, M.D.; Zhang, H.; Fesik, S.W.; Rosenberg, S.H., An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* **2005**, *435*, 677-681.

33. Tran, T.-D.; Adam, F.M.; Calo, F.; Fenwick, D.R.; Fok-Seang, J.; Gardner, I.; Hay, D.A.; Perros, M.; Rawal, J.; Middleton, D.S.; Parkinson, T.; Pickford, C.; Platts, M.; Randall, A.; Stephenson, P.T.; Vuong, H.; Williams, D.H., Design and optimisation of potent gp120-CD4 inhibitors. *Bioorg Med Chem Lett* **2009**, *19*, 5250-5255.

34. Shangary, S.; Wang, S., Small-molecule inhibitors of the MDM2-p53 protein-protein interaction to reactivate p53 function: a novel approach for cancer therapy. *Annu Rev Pharmacol Toxicol* **2009**, *49*, 223-241.

35. Meagher, K.L.; Carlson, H.A., Solvation influences flap collapse in HIV-1 protease. *Proteins* **2005**, *58*, 119-125.

36. Case, D.A.; Darden, T.A.; Cheatham III, T.E.; Simmerling, C.L.; Wang, J.; Duke, R.E.; Luo, R.; Crowley, M.; Walker, R.C.; Zhang, W.; Merz, K.M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K.F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S.R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D.H.; Seetin, M.G.; Sagui, C.; Babin, V.; Kollman, P.A. (2008), AMBER 10, University of California, San Francisco

37. Case, D.A.; Darden, T.A.; Cheatham III, T.E.; Simmerling, C.L.; Wang, J.; Duke, R.E.; Luo, R.; Merz, K.M.; Wang, B.; Pearlman, D.A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J.W.; Ross, W.S.; Kollman, P.A. (2004), AMBER 8, University of California, San Francisco

38. Wickstrom, L.; Okur, A.; Simmerling, C., Evaluating the performance of the ff99SB force field based on NMR scalar coupling data. *Biophys J* **2009**, *97*, 853-856.

39. Spinelli, S.; Liu, Q.Z.; Alzari, P.M.; Hirel, P.H.; Poljak, R.J., The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. *Biochimie* **1991**, *73*, 1391-1396.

40. Pearlman, D.A.; Case, D.A.; Caldwell, J.W.; Ross, W.S.; Cheatham III, T.E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P., AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp Phys Commun* **1995**, *91*, 1-41.

41. Baker, N.A.; Sept, D.; Joseph, S.; Holst, M.J.; McCammon, J.A., Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci* **2001**, *98*, 10037-10041.

42. Lerner, M.G.; Carlson, H.A. APBS plugin. University of Michigan: Ann Arbor, MI

43. DeLano, W.L., *The PyMOL Molecular Graphics System, v0.99*, DeLano Scientific: Palo Alto, CA 2002

44. Jorgensen, W.L., Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L., Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **1983**, *79*, 926-935.

45. Darden, T.; York, D.; Pedersen, L., Particle Mesh Ewald - an N.Log(N) method for Ewald Sums in large systems. *J. Chem. Phys* **1993**, 10089-10092.

46. 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.

47. Kabsch, W. and Sander, C., Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577-2637.

48. Meagher, K.L.; Lerner, M.G.; Carlson, H.A., Refining the MPS pharmacophore method: consistency across three independent HIV-1 protease models. *J Med Chem* **2006**, *49*, 3478-3484.

49. Meagher, K.L.; Carlson, H.A., Incorporating protein flexibility in structure-based drug design: using HIV-1 protease as a test case. *J Am Chem Soc* **2004**, *126*, 13276-13281.

50. Jorgensen, W.L., *BOSS, Version 4.2*; Yale University: New Haven, CT, 2000

51. Jorgensen, W.L.; Maxwell, D.S.; Tirado-Rives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* **1996**, *118*, 11225-11236.

52. Damm, K.L.; Carlson, H.A., Gaussian-weighted RMSD superposition of proteins: a structural comparison for flexible proteins and predicted protein structures. *Biophys J* **2006**, *90*, 4558-4573.

53. Lerner, M.G.; Meagher, K.L., Carlson, H.A., Automated clustering of probe molecules from solvent mapping of protein surfaces: new algorithms applied to hot-spot mapping and structure-based drug design. *J Comput Aided Mol Des* **2008**, *22*, 727-736.

54. *OMEGA, Version 1.8.b1*; OpenEye Scientific Software, Inc.: Santa Fe, NM, 2004

55. Bowman, A.L.; Nikolovska-Coleska, Z.; Zhong, H.; Wang, S.; Carlson, H.A., Small molecule inhibitors of the MDM2-p53 interaction discovered by ensemble-based receptor models. *J Am Chem Soc* **2007**, *129*, 12809-12814.

56. *MOE, v2009.10;* Chemical Computing Group: Montreal, Canada, 2009

57. Matayoshi, E.D.; Wang, G.T.; Krafft, G.A.; Erickson, J., Novel fluorogenic substrates for assaying retroviral proteases by resonance energy transfer. *Science* **1990**, *247*, 954-958.

58. Strisovsky, K.; Tessmer, U.; Langer, J.; Konvalinka, J.; Krausslich, H.G., Systematic mutational analysis of the active-site threonine of HIV-1 protease: rethinking the "fireman's grip" hypothesis. *Protein Sci* **2000**, *9*, 1631-1641.

59. Lerner, M.G.; Bowman, A.L.; Carlson, H.A., Incorporating dynamics in E. coli dihydrofolate reductase enhances structure-based drug discovery. *J Chem Inf Model* **2007**, *47*, 2358-2365.

60. Ohtaka, H.; Schon, A.; Freire, E., Multidrug resistance to HIV-1 protease inhibition requires cooperative coupling between distal mutation. *Biochemistry* **2003**, *42*, 13659-13666.

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

62. Levy, Y.; Caflisch, A.; Onuchic, J.N.; Wolynes, P.G., The folding and dimerization of HIV-1 protease: evidence for a stable monomer from simulations. *JMB* **2004**, *340*, 67-79.

63. Bogan, A.A. and Thorn, K.S., Anatomy of hot spots in protein interfaces. *JMB* **1998**, *280*, 1-9.

64. Chene, P., Inhibiting the p53-MDM2 interaction: an important target for cancer therapy. *Nat Rev Cancer* **2003**, *3*, 102-109.

65. Stumpf, M.P.H.; Thorne, T.; de Silva, E.; Stewart, R.; An, H.J.; Lappe, M.; Wiuf, C., Estimating the size of the human interactome. *Proc Natl Acad Sci* **2008**, *105*, 6959-6964.

Chapter 3

Computational studies of the HIV-1 protease dimer interface complexed with dissociative inhibitors

Introduction

HIV-1 protease (HIVp) is an obligate dimeric homodimer protein with C_2 symmetry.[1, 2] Zhang *et al.* created the first dissociative inhibitor targeting HIVp by creating a short, four residue peptide mimic, a peptidomimetic, of the C-terminus at the dimer interface.[3] Based on the kinetic model of HIVp, Zhang also developed a fluorescent release, kinetic-based assay capable of discriminating between the two modes of inhibition: competitive and dissociative inhibition. Sedimentation equilibration experiments were performed to confirm the dissociative mechanism and resulted in the observation of dimeric or monomeric forms of HIVp in the absence or presence of the dissociative tetrapeptide.

Schramm *et al.* tested a series of compounds based on the N- and C-termini to determine the peptidimimetic core residues to efficiently inhibit HIV-1.[4] Schramm *et al.* continued to increase HIVp dimer inhibition, refining the monovalent compounds based on the C-terminus, conjugating a palmitic acid lipid moiety to the N-terminus of his monovalent dissociative inhibitor.[5, 6] In 2000, Calfisch *et al.* mapped the HIVp dimer interface using their Multiple Copy Simultaneous Search (MCSS) method.[7] The MCSS method mapped a large and accommodating region normally occupied in the area

occupying symmetric residue Phe99'. The MCSS map of the HIVp monomer allowed Schramm to incorporate a thyroxine moiety to the dissociative lipopeptide corresponding to the position Phe99' would occupy, increasing ligand potency.[8]

Investigators continued to optimize the monovalent dissociative inhibitor. Breccia *et al.* included a bicyclic guanidinium group to the N-terminus of the monovalent ligand, between the peptide and lipid moieties of the compound.[9] The bicyclic guanidinium group is a positively charged molecule and its incorporation was designed to add a favorable interaction to the negatively charged C-terminus. Bannwarth *et al.* varied the length of the palmitic acid in the dissociative lipopeptide and showed inhibition rates for dimer inhibitors were comparable for the wild-type HIVp and multidrug resistant protease MDR-HM (L10I-M46I-I54V-V82A-I84V-L90M) and ANAM-11 (L10I-M36I-S37D-M46I-R57K-L63PA71V-G73S-I84V-L90M-I93L), unlike competitive inhibitors.[10] The ANAM-11 and MDR-HM HIVp variants are analogous to proteases found in multi-drug resistant patients.[11, 12]

In 1997, Zutshi *et al.* pioneered a novel set of bivalent compounds targeting the HIVp dimer interface.[13] Zutshi cross-linked N- and C-termini derived peptidomimetics with an alkyl linker. The cross-linking strategy generated the focus of Chmielewski's group to produce bivalent compounds targeting the dimer interface. Bowman *et al.* varied the linker length, changed linker moieties, and restricted the degrees of freedom within the alkyl linker; however, a 14 carbon alkyl linker produced the best dissociative inhibition.[14] To reduce the size of the bivalent, Hwang and Chmielewski reduced each peptidomimetic chain to the core residues required to induce dissociation.[15] Derivatives

of the bivalent dimer inhibitor increased potency of the compound, primarily through side-chain modification of the bivalent peptidomimetic.[16, 17]

Recognizing the large entropic penalty associated with an alkyl linker, Bouras *et al.* created a set of 'molecular tongs'.[18] Molecular tongs are bivalent compounds with an aromatic group in the linking region. Linker moieties included resorcinol, pyridinediol, or naphthalenediol as the restricting group.[19] Dissociative inhibition rates were determined in the micromolar to submicromolar range. A positively charged quinoline based linker was designed to restrict and create a favorable complement to the negatively charged C-terminus, and non-peptidomimetic chemical fragments mimicking amino acid side chains were created to increase affinity to the dimer interface.[10]

To design stronger, more efficient inhibitors, more structural information regarding the relationship between the protein-ligand complex is needed. Frutos *et al.* attempted to describe the relationship using a [2-^{13}C]-labeled Trp residue and 2D HSQC NMR.[20] The Trp residue was chosen because only two Trp residues exist in each chain of the HIVp: Trp6 and Trp42. When HIVp was exposed to the SYEL tetrapeptide, a dissociative inhibitor, NMR chemical shifts were observed with the labeled Trp6, within the dimer interface. These chemical shifts indicated that the monovalent compound was bound near the Trp6 residue; however, the detailed mechanism of binding at an atomic resolution remains unclear.

We feel computational modeling techniques can explain the binding mechanism and guide future medicinal chemistry efforts. We decided to model dimer compounds derived from the Chmielewski group, due to the extensive and diverse library of compounds tested against HIVp. Due to the peptide mimic nature of the bivalent

compounds, the initial coordinates for the protein-ligand computational model can be developed from the symmetry inherent in HIVp.

Methods

Ligand Construction

The protein-ligand computational models were developed as described in Bowman *et al.* inhibition mechanism study of their dimer inhibitors.[21] Ligands of interest are found in Figure 1, compounds (1) and (2). We termed compound (1) as "Longest" in our study, due to the length of the compound and termed compound (2) as "Common" due to their choice to use the compound as a scaffold for derivatives. The term "Northern" chain will refer to the N-terminal peptidomimetic portion and "Southern" chain will refer to the C-terminal peptidomimetic portion.

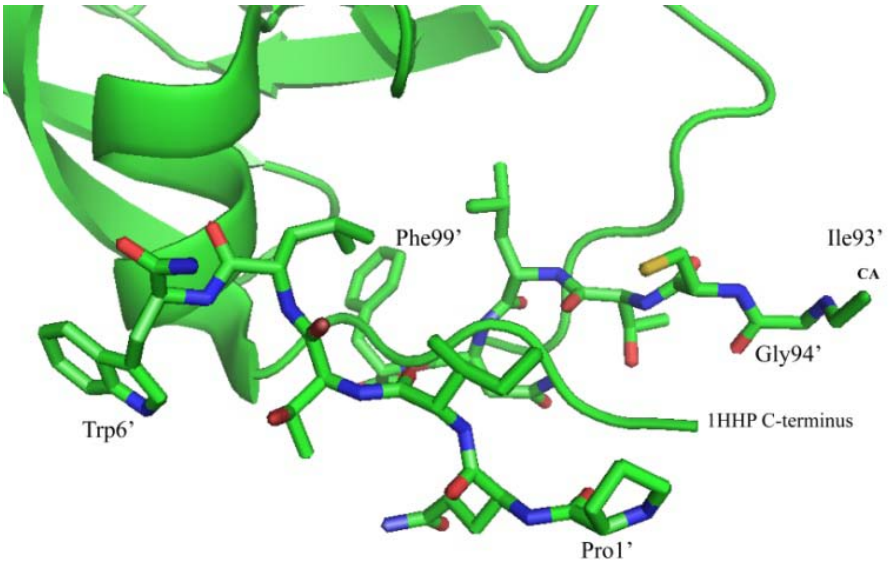

Figure 3.1: 'Longest' compound construction. A PyMOL generated image of the putative position of the N- and C-terminal peptidomimetics. PDB: 1HHP is shown in cartoon format and the atoms (and coordinates) used to build the peptidomimetics are displayed in sticks. The symmetry partner of 1HHP was first generated and then atoms were trimmed to build the Longest compound. One acetamide (ACE) residue was generated using the carboxyl of Gly94' and the actyl chain continued through Ile93'. The other ACE residue carboxyl group was added to the Pro1' N-terminus.

Design of Chmielewski's Longest dimer inhibitor required preservation as much of the original PDB 1HHP dimer information, shown in Figure 2.1. An acetamide (ACE) residue was added to the fixed Pro1 amide and the Gly94 carbonyl was replaced with the carbonyl for another ACE residue. The carbon linker was built using the coordinates for the Gly94 backbone following to the alpha-carbon of Ile93. The rest of the carbon linker was built around the C-terminus end of the 1HHP monomer and minimized in MOE, using the MMFF94x force field. All other atomic coordinates were fixed within MOE during minimization.

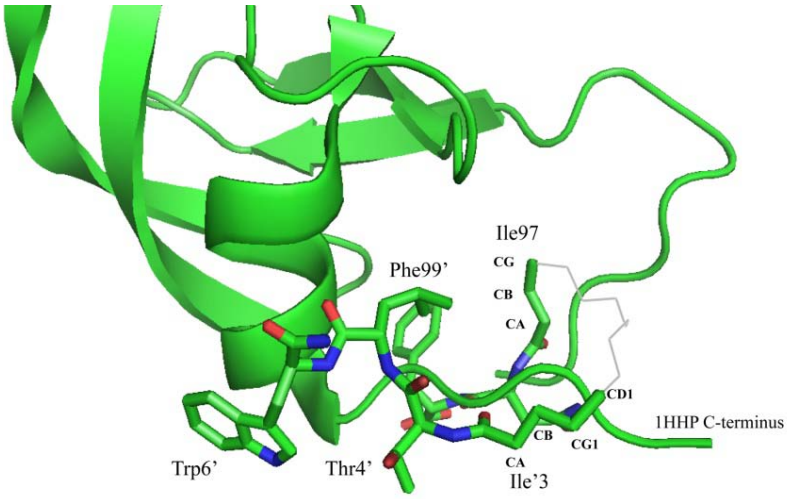

Figure 3.2: 'Common' compound construction. A PyMOL generated image of the putative position of the N- and C-terminal peptidomimetics. PDB: 1HHP is shown in cartoon format and the atoms (and coordinates) used to build the peptidomimetics are displayed in sticks. The symmetry partner of 1HHP was first generated and then atoms were trimmed to build the Common compound. The two Ace residues were built using atoms from Ile3' and Ile97', following the backbone through the side chain. The acyl linker is shown in gray, lines representation.

The same methodology for the Longest compound was applied to construct Chmielewski's common compound. The carbonyl groups for the two ACE residues were constructed using the coordinates for the carbonyl groups in Ile3 and Leu 97. Ile3 and Leu97 have long side chains, helping to guide the linker chain. Coordinates for Ile3 atoms CA, CB, CG1, and CD1 were all used for the linker carbons. Leu97 atomic coordinates for CA, CB, and CG were used to build the carbon linker over the C-terminus. The added linker carbon atoms were minimized within MOE, using the MMFF94x force field.

Parameter Modification

Modifications to Carr *et al.* AMBER94 force field parameters were needed to create the linker region in Chmielewski's dimer inhibitors.[22] The Ace residue was adapted due to similarity to the linker used by Chmielewski, shown in Figure 2.3. The

ACE residue was modified by replacing a HC atom with a CT atom. The previous charge on the HC atom was distributed to the original CT atom (-0.3662 to -0.2539 charge). The linker contained several CT (-0.2246 charge) atoms and 2 HC (0.1123 charge) for every CT atom. The linker was built over the C-terminus (buried side) and connected to another ACE residue conjugated to the peptidomimetic region. The chain was allowed to minimize in MOE, using the MMFF94x force field while the remainder atoms were restrained from movement.[23]

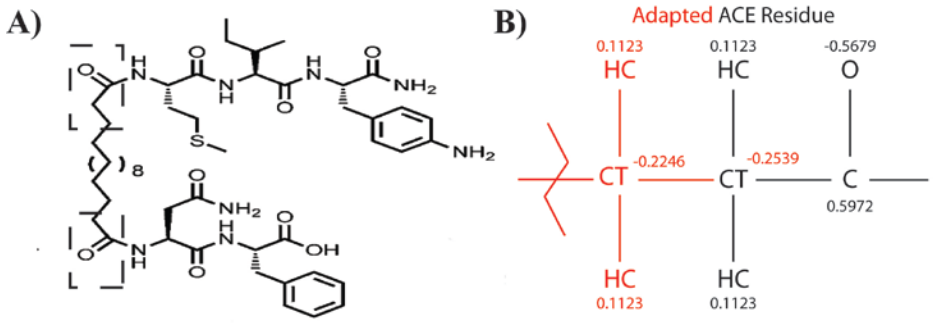

Figure 3.3: ACE parameter modification. A) The 'Common' inhibitor from Bowman, *et al.* The hash boxes indicate the modified Ace residues. B) Carr *et al.* AMBER94 force field ACE residue parameter was adapted to the construct the alkyl linker for Chmielewski's two compounds. A hydrocarbon was replaced with a cabon tetrahedral and the charge distributed to the original carbon tetrahedral. A chain of carbon tetrahedrals and hydrocarbons, with the charge of -0.2246 and 0.1123 respectively, completed the alkyl linker.

Protein and Ligand Preparation

All-atom molecular dynamics (MD) simulations were performed using the AMBER8 or AMBER10 programs and the FF99SB force field.[24-26] Initial atomic coordinates for the HIVp monomer simulation were obtained from the Protein Data Bank (PDB) repository (PDB ID: 1HHP). Hydrogen atoms were added *via* the AMBER tLEaP module.[27] The +2 electron charge of the HIVp monomer was neutralized internally with

the -2 charge of the longest ligand termini. The system was solvated with a 16 Å octahedron solvation box, with a closeness parameter of 0.5 and TIP3P water.[28]

Minimization and equilibration of the solvated protein system occurs over a series of steps. Hydrogen atoms, water molecules, and ions are first minimized while everything else is restrained. All-atoms, minus the restrained protein backbone atoms, are then minimized and, finally, all-atoms are minimized unrestrained.

In equilibration, SHAKE and 1 fs is the time-step unless noted. A 10 Å nonbonded cutoff is used and PME boundary conditions are applied. Equilibration occurs in four steps: 1) Water and counter ions are heated from 10 to 310 K under constant volume for 50 ps (time-step = 2 fs). During the temperature increase, the remainder of the system is restrained (50 kcal/mol). 2) Water and counter ions are then allowed to equilibrate for another 10 ps while the remainder of the system is restrained (50 kcal/mol). 3) The protein is now allowed to heat from 10 K to 310 K in three, 10 ps steps under constant volume. In the first three equilibration steps, the temperature increases 10 K to 110 K, then 110 K to 210 K, and last, 210 K to 310 K. Meanwhile, the restraints on the hydrogen atoms, water molecules, and counter ions are relaxed from 2 kcal/mol, then 1 kcal/mol and, finally, 0.5 kcal/mol. 4) The system is now allowed to equilibrate at a constant pressure in a series of steps. The first step is a 10 ps equilibration with a restraint (0.5 kcal/mol) on all atoms, except for hydrogen atoms, water molecules, and counter ions. Next, the system is equilibrated for 10 ps with a weaker restraint (0.1 kcal/mol) on the previous atoms. Last, the system is equilibrated for 2 ns unrestrained. The solvated protein system is now ready for the production phase of the trajectory.

Langevin Dynamics Preparation.

The initial steps of LD setup are similar to MD setup. HIVp atomic coordinates were obtained from the Protein Data Bank (PDB) repository (PDB ID: 1HHP). AMBER8 and FF99SB parameters were used to perform LD simulations.[26] Hydrogen atoms were added during import into AMBER's tleap program. The aGB^{OBC}(II) parameters were used for the Generalized Born approach to model the solvent implicitly.[29] A 999 Å cutoff was used for non-bonded interactions, $\gamma = 1\ ps^{-1}$, 1 fs time-step and no boundary conditions were used. Default dielectrics were used: 78.5 for the exterior and 1 for the interior.

Two minimization steps were required prior to equilibration. Hydrogen atoms were minimized first, followed by an all-atom minimization. The equilibration procedure is a series of steps to unrestrain the protein atoms and to ramp the temperature from 100 to 300K. The first equilibration step increased the temperature from 100 K to 200 K with restrain (2 kcal/mol) on all heavy atoms for 10 ps. The second equilibration step increased the temperature from 200 K to 300 K with a reduced restrain (1 kcal/mol) on heavy atoms. The following two equilibration steps further reduced the previous restraint, 0.5 kcal/mol followed by 0.1 kcal/mol, for 10 ps each. Side chain atoms were released from the restraint, while the backbone atoms remained restrained (0.1 kcal/mol), for 10 ps. Carbons in the linker region of Chmielewski's dimer inhibitors were restrained in addition to the backbone atoms. The last equilibration step included 300 ps of unrestrained production. Ten separate 10 ns LD simulations were performed with unique seed numbers, for a total of 100 ns of LD production.

Trajectory Analysis

A root mean square deviation (RMSD) measures the degree of fluctuation when compared to a reference structure or an average position; however, RMSD is not a good measurement for dimer inhibitors. The N- and C-termini in the HIVp monomer are dynamic, measuring an RMSD for the N- and C-termini of 6 and 8 Å, respectively, complicating an accurate measure of stability at the dimer interface. The Multiscale Modeling Tools for Structural Biology (MMTSB) tool set and the MMTSB contact script determined ligand stability by measuring the ligand's contact to the protein surface.[30] Contact, as defined by the MMTSB script, is a pair-wise inter-residue distance calculation of all pairs of heavy atoms less than 4.2 Å. The contact script ascertained the percentage of contacts maintained over the course of a trajectory. The initial coordinates of the built protein-ligand complex was defined as the reference structure. The peptidomimetics were designed to mimic the peptides native to the original structure and, constitute the reference contacts.

Results and Discussion

Molecular Dynamics Simulation

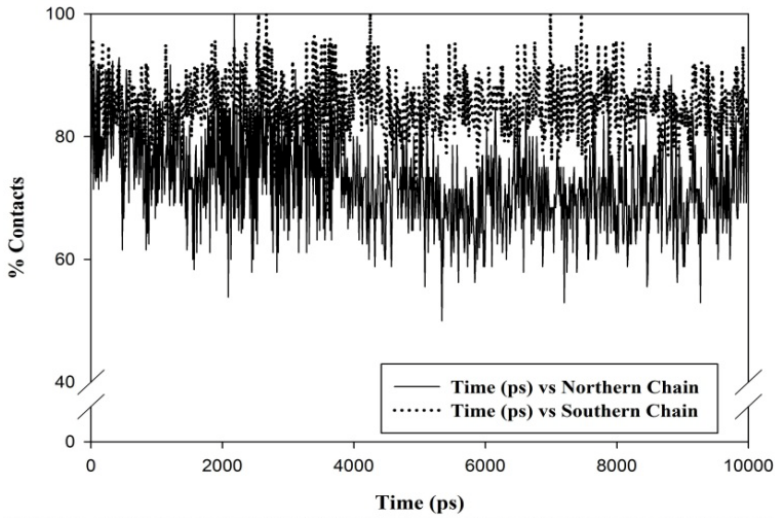

Figure 3.4: Northern and Southern Chains contact percentage from an MD simulation of the Longest compound. Solid line indicates the Northern chain and the dotted line indicates the Southern chain. The average percentage contact through the course of the trajectory for the Northern and Southern chains are 72.9 % and 85.4 %, respectively.

The MD simulation with the Longest compound complexed to the HIVp monomer in the dimer was ambiguous. The dissociative inhibition rate, $K_{i,D}$, of the Longest compound was reported as 220 nM. The stability of the ligand's contact, plotted in Figure 2.4, of the Longest compound in simulation will reflect the inhibition rate. The 10 ns simulation showed the compound maintained contacts over the course of the trajectory, when compared to the putative binding mechanism; however, there is a caveat to the protein-ligand simulated complex. It was undetermined if the stability seen in the MD simulation was a result of binding mechanism, or the low amount of sampling produced from a MD simulation. A higher sampling method, such as a LD dynamic simulation, was decided as a more suitable method to determine if the stability was a

result of favorable complementary between the HIVp dimer interface and the Longest compound.

Langevin Dynamic simulations

Ten separate LD simulations of the Longest and Common ligands complexed to the HIVp monomer were produced to parallel the MD simulation. The Longest and Common compound were simulated to compare the stability of the two compounds. The dissociative inhibition rate, $K_{i,D}$, of the Common compound was reported as 3.0 µM, higher than the Longest compound $K_{i,D}$ of 220 nM. We expected our ligand stability in our simulations to reflect the inhibition rates. The perseverance of ligand contacts to the protein surface in the LD simulation will provide a metric to indicate inhibition rates.

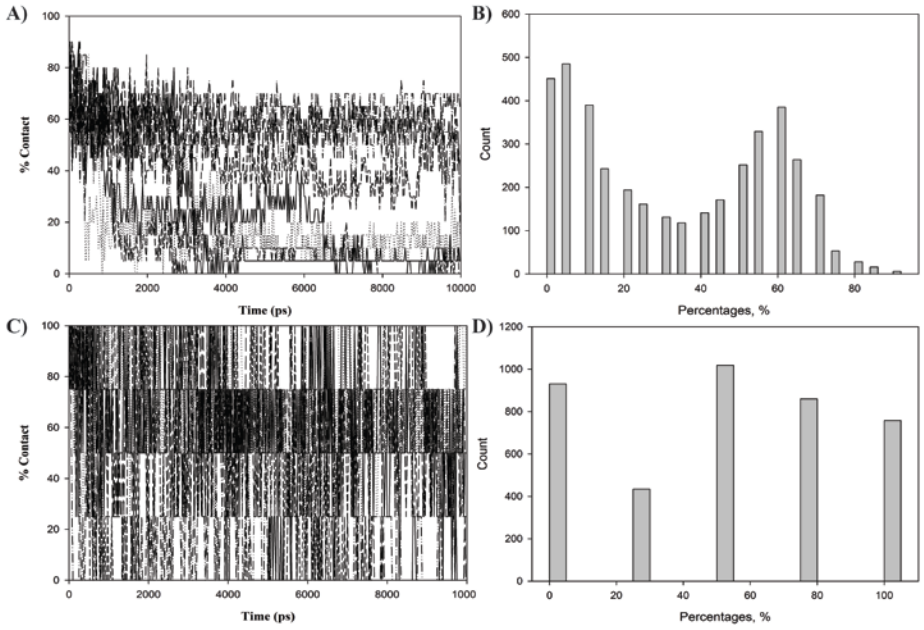

Figure 3.5: Northern Chain percentage contact information for Longest and Common LD simulations. (A) is a line graph of the Longest, Northern chain contacts over the course of the 10 independent LD trajectories. B) is a histogram of the percentages in (A). C) is a line graph of the Common, Northern chain contacts over the course of the 10 independent LD trajectories. D) is a histogram of the percentages in (C). The average percentage for Northern Longest and Common chain contacts for all 10 trajectories were 32.8 % and 50.5%, respectively.

In Figure 3.5 (A) and (B), the contact percentage is detailed for the Longest, Southern compound's ten LD trajectories. The histograms displayed in (B) and (D) is a graphical representation of the combined ten LD trajectory contact percentage distribution for all ten LD trajectories. The Longest, Northern chain shows a bimodal distribution centered near 5 and 60%. The average percentage for the combined ten LD trajectories contact percentages was approximately 33%. The Common, Northern chain, Figure 3.5 (C) and (D), performed comparatively better compared to the Longest, Southern chain. The Common, Southern chain displayed a unimodal distribution near 50%. Despite the tail in Figure 3.5 (D) towards the left, the average contact percentage for the Common, Nothern chain was 50%, better than the Longest, Northern chain.

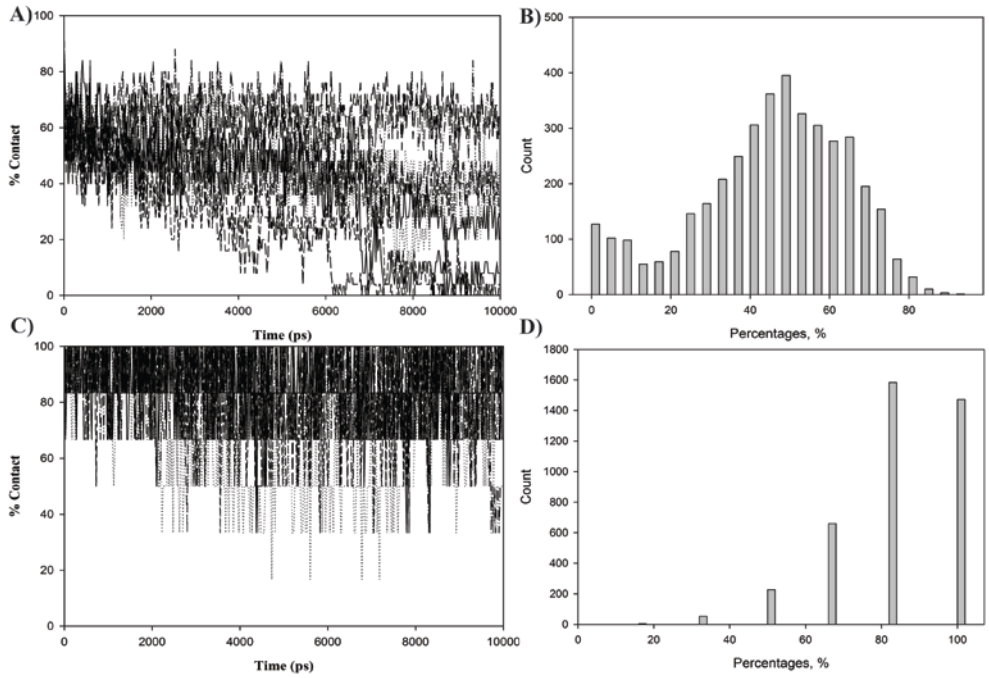

Figure 3.6: Southern Chain percentage contact information for the Longest and Common inhibitor. (A) is a line graph of the the Longest, Southern chain contacts over the course of the 10 independent LD trajectories. B) histogram of the percentages in (A). (C) is a line graph of the Common, Southern chain contacts over the course of the 10 independent LD trajectories. (D) histogram of the percentages in (C). The average percentage for the Longest and Common Southern chain contacts for all 10 trajectories are 44.3% and 84.1%, respectively.

In Figure 3.6 (A) and (B), the contact percentage is detailed for the Longest, Southern compound's ten LD trajectories. The histograms in (B) and (D) plotted a graphical representation of the combined ten LD trajectories in (A) and (C), respectively. The Longest, Southern chain shows a unimodal distribution centered near 50%. The average percentage for the combined ten LD trajectories contact percentages was approximately 44%. The Common, Southern chain, Figure 3.5 (C) and (D), performed significantly better compared to the Longest, Southern chain. The Common, Southern chain displayed a unimodal distribution near 80%.

Alternative Binding Mechanism

The putative binding mode is not likely the binding mechanism measured *in vitro*. As stated earlier, the N- and C-termini are dynamic and there is a large entropic penalty for binding in the proposed mechanism. The prospect of an alternative binding mechanism is a possibility. Our simulations could discover an alternative binding mechanism, a rearrangement by either of the chains resulting in a more stable binding mechanism. We can use the ligand contacts to search for a common HIVp surface region amongst the ten LD trajectories for the Longest and Common compounds. If the Northern or Southern chain did uniformly rearrange, the ligands will share high agreement in the surface area contacts made on the HIVp dimer interface region.

Table 3.1: A pairwise comparison of Longest, Northern chain contacts. To check for agreement across LD trajectories of Chmielewski's longest compound, a pairwise comparison of the Northern chain in the 9976 ps frame was performed. Areas of agreement will share similar contacts and have high contact agreement. Generally, there is little to no agreement amongst the frames when the contacts are compared across productions (Prod). Note: LD production 7 exited the dimer region entering the active site of HIVp.

	LD Prod	1	2	3	4	5	6	7	8	9	10
Reference Structure	1	100.00	0.00	0.00	10.00	0.00	5.88	0.00	0.00	0.00	0.00
	2	0.00	100.00	16.67	10.00	15.38	11.76	0.00	8.33	28.57	0.00
	3	0.00	10.00	100.00	30.00	15.38	23.53	0.00	8.33	28.57	16.67
	4	16.67	10.00	66.67	100.00	23.08	23.53	0.00	16.67	42.86	16.67
	5	0.00	20.00	33.33	30.00	100.00	41.18	0.00	0.00	57.14	0.00
	6	16.67	20.00	66.67	40.00	53.85	100.00	0.00	8.33	28.57	0.00
	7	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00
	8	0.00	10.00	16.67	20.00	0.00	5.88	0.00	100.00	14.29	16.67
	9	0.00	20.00	33.33	30.00	30.77	11.76	0.00	8.33	100.00	16.67
	10	0.00	0.00	16.67	10.00	0.00	0.00	0.00	8.33	14.29	100.00

Table 3.1, above, is a pair-wise comparison the contacts made by the Northern chain compound in the Longest compound. The left side of the table is the reference structure being compared to by the top LD production frame. The 9976 ps frame is being compared in all LD productions. Across the table, there is low agreement in the contacts to the protein surface area. LD production seven dissociated from the dimer interface and entered the catalytic pocket of the protease monomer.

Table 3.2: A pairwise comparison of Longest, Southern chain contacts. To check for agreement across LD trajectories of Chmielewski's Longest compound, a pairwise comparison of the Southern chain in the 9976 ps frame was performed. Areas of agreement will share similar contacts and have high contact agreement. Generally, there is moderate agreement amongst the frames when the contacts are compared across productions (Prod). Note: LD production 7 exited the dimer region entering the active site of HIVp.

Reference Structure \ LD Prod	1	2	3	4	5	6	7	8	9	10
1	100.00	10.00	0.00	11.76	8.70	0.00	50.00	28.57	7.69	0.00
2	40.00	100.00	37.50	58.82	39.13	70.00	0.00	85.71	84.62	9.09
3	0.00	30.00	100.00	52.94	34.78	40.00	0.00	14.29	38.46	0.00
4	40.00	50.00	56.25	100.00	56.52	60.00	0.00	57.14	61.54	9.09
5	40.00	45.00	50.00	76.47	100.00	40.00	0.00	57.14	46.15	9.09
6	0.00	35.00	25.00	35.29	17.39	100.00	0.00	28.57	38.46	9.09
7	20.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00
8	40.00	30.00	6.25	23.53	17.39	20.00	0.00	100.00	23.08	0.00
9	20.00	55.00	31.25	47.06	26.09	50.00	0.00	42.86	100.00	9.09
10	0.00	5.00	0.00	5.88	4.35	10.00	0.00	0.00	7.69	100.00

In LD trajectories of the Longest inhibitor, the Southern peptidomimetic chain shares more of the HIVp dimer interface contacts than the Northern chain, shown in Table 3.2. Besides LD production nine to LD production two and four, there is not a significant amount of uniformity across the trajectories to consider a consensus binding mechanism. The Longest inhibitor modeled in the ten trajectories contact diverse areas along the protein surface.

Table 3.3: A pairwise comparison of 'Common', Northern chain contacts. To check for agreement across LD trajectories of Chmielewski's Common compound, a pairwise comparison of the Northern chain in the 9976 ps frame was performed. Areas of agreement will share similar contacts and have high contact percentage. There is high agreement in LD production (Prod) 6 with several of the reference structures.

Reference Structure \ LD Prod	1	2	3	4	5	6	7	8	9	10
1	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	100.00	0.00	0.00	0.00	66.67	50.00	45.45	54.55	0.00
3	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	100.00	33.33	0.00	0.00	0.00	0.00	80.00
5	0.00	0.00	0.00	16.67	100.00	0.00	0.00	0.00	0.00	20.00
6	0.00	14.29	0.00	0.00	0.00	100.00	75.00	27.27	27.27	0.00
7	0.00	14.29	0.00	0.00	0.00	100.00	100.00	36.36	36.36	0.00
8	0.00	35.71	0.00	0.00	0.00	100.00	100.00	100.00	81.82	0.00
9	0.00	42.86	0.00	0.00	0.00	100.00	100.00	81.82	100.00	0.00
10	0.00	0.00	0.00	66.67	33.33	0.00	0.00	0.00	0.00	100.00

In Table 3.3 above, the Common, Northern chain in the modeled ligand did not show uniformity across all frames; however, there was is a reasonable conformity amongst a few of the frames amongst the ten trajectories. The frame for LD production six shares 100% of the surface contacts with the reference frames seven through nine, and reasonable agreement with LD production two; nonetheless, there would need to be more trajectories in agreement to declare the result represented the alternative binding mechanism seen in LD productions six through nine.

Table 3.4: A pairwise comparison of 'Common', Southern chain contacts. To check for agreement across LD trajectories of Chmielewski's Common compound, a pairwise comparison of the Southern chain in the 9976 ps frame was performed. Areas of agreement will share similar contacts and have high contact percentage. Generally, there is high agreement across all the production (Prod) runs.

	LD Prod	1	2	3	4	5	6	7	8	9	10
Reference Structure	1	100.00	77.78	70.00	75.00	77.78	83.33	81.82	41.18	100.00	83.33
	2	70.00	100.00	80.00	50.00	66.67	83.33	63.64	41.18	66.67	66.67
	3	70.00	88.89	100.00	75.00	77.78	100.00	72.73	35.29	83.33	66.67
	4	30.00	22.22	30.00	100.00	44.44	50.00	27.27	17.65	50.00	50.00
	5	70.00	66.67	70.00	100.00	100.00	100.00	72.73	41.18	100.00	100.00
	6	50.00	55.56	60.00	75.00	66.67	100.00	54.55	29.41	66.67	66.67
	7	90.00	77.78	80.00	75.00	88.89	100.00	100.00	47.06	100.00	83.33
	8	70.00	77.78	60.00	75.00	77.78	83.33	72.73	100.00	83.33	83.33
	9	60.00	44.44	50.00	75.00	66.67	66.67	54.55	23.53	100.00	66.67
	10	50.00	44.44	40.00	75.00	66.67	66.67	45.45	29.41	66.67	100.00

The Common, Southern chain reported a high agreement across the frames, shown in Table 3.4. The Common Northern and Southern chains in our ten LD simulations described a greater consensus in shared protease dimer interface contacts than the Longest compound. There was not an apparent alternative binding mechanism in our LD simulations and our results contrasted with the empirical data. There could still be an alternative binding mechanism, as the hypothetical binding mode is unstable at the termini, but our LD simulations of the Longest and Common inhibitor failed to find an alternative receptor pocket along the dimer interface.

Conclusions

Despite the initial MD results showing a high degree of stability by the Longest dissociative inhibitor, our LD results do not corroborate with the MD simulation. The LD simulation allows a greater amount of sampling without the explicit water molecules present in the simulation. If the ligand were to dissociate from the compound, as happened in the Longest simulation, the dissociation event would likely occur in LD simulation and not in MD.

The LD results were not as we predicted. The $K_{i,D}$ for the Longest inhibitor is 220 nM and the Common $K_{i,D}$ is 3 μM, at least one order of magnitude greater than the Longest inhibitor. We expected the Longest inhibitor to show greater stability in LD simulations than the Common inhibitor. Our results showed the opposite. Not only did the Longest inhibitor failed to maintain contact on the HIVp dimer interface, they did not rearrange into a common binding mode. The seventh trajectory of the Longest inhibitor had another unique distinction, it was the only trajectory out of the twenty total Longest and Common simulations in which the ligand dissociated from the dimer interface and entered the active site.

Although our simulations failed to reveal the protein-ligand relationship, we do know the inhibitors are not likely bound in the hypothesized, peptidomimetic mechanism as designed. More structural information is needed to properly setup future simulations. Better structural information will help construct the compound's initial atomic coordinate position for computational modeling. The modeled dissociative inhibitors should then

find a favorable low energy well and relate the protein-ligand relationship at the dimer interface.

References

1. Wlodawer, A.; Erickson, J.W., Structure-based inhibitors of HIV-1 protease. *Annu Rev Biochem* **1993**, *62*, 543-585.

2. Oroszlan, S.; Luftig, R.B., Retroviral Proteinases. *Curr Top Microbiol Immunol* **1990**, *157*, 153-185.

3. Zhang, Z.-Y.; Poorman, R.A.; Maggiora, L.L.; Heinrikson, R.L.; Kezdy, F.J., Dissociative inhibition of dimeric enzymes. *J Biol Chem* **1991**, *266*, 15591-15594.

4. Schramm, H.J.; Billich, A.; Jaeger, J.; Rucknagel, K.-P.; Arnold, G.; Schramm, W., The inhibition of HIV-1 protease by interface peptides. *Biochem Biophys Res Commun* **1993**, *194*, 595-600.

5. Schramm, H.J.; Boetzel, J.; Buttner, J.; Fritsche, E.; Gohring, W.; Jaeger, E.; Konig, S.; Thumfart, O.; Wenger, T.; Nagel, N.E.; Schramm, W., The inhibition of human immunodeficiency virus protease by 'interface peptides'. *Antiviral Res* **1996**, *30*, 155-170.

6. Schramm, H.J.; de Rosny, E.; Reboud-Ravaux, M.; Buttner, J.; Dick, A.; Schramm, W., Lipopeptides as dimerization inhibitors of HIV-1 protease. *Biol Chem* **1999**, *380*, 593-596.

7. Caflisch, A.; Schramm, H.J.; Karplus, M., Design of dimerization inhbitors of HIV-1 aspartic proteinase: A computer-based combinatorial approach. *J Comput Aided Mol Des* **2000**, *14*, 161-179.

8. Dumond, J.; Boggetto, N.; Schramm, H.J.; Schramm, W.; Takahashi, M.; Reboud-Ravaux, M., Thyroxine-derivatives of lipopeptides: bifunctional dimerization inhibitors of human immunodeficiency virus-1 protease. *Biochem Pharmacol* **2003**, *65*, 1097-1102.

9. Breccia, P.; Boggetto, N.; Perez-Fernandez, R.; Van Gool, M.; Takahashi, M.; Rene, L.; Prados, P.; Badet, B.; Reboud-Ravaux, M.; de Mendoza, J., Dimerization inhibitors of HIV-1 protease based on bicyclic guanidinium subunit. *J Med Chem* **2003**, *46*, 5196-5207.

10. Bannwarth, L.; Kessler, A.; Pethe, S.; Collinet, B.; Merabet, N.; Boggetto, N.; Sicsic, S.; Reboud-Ravaux, M.; Ongeri, S., Molecular tongs containing amino acid mimetic fragments: new inhibitors of wild-type and mutated HIV-1 protease dimerization. *J Med Chem* **2006**, *49*, 4657-4664.

11. Muzammil, S.; Ross, P.; Freire, E., A major role for a set of non-active site mutations in the development of HIV-1 protease drug resistance. *Biochemistry* **2003**, *42*, 631-638.

12. Ohtaka, H.; Schon, A.; Freire, E., Multidrug resistance to HIV-1 protease inhibition requires cooperative coupling between distal mutation. *Biochemistry* **2003**, *42*, 13659-13666.

13. Zutshi, R.; Franciskovich, J.; Shultz, M.; Schweitzer, B.; Bishop, P.; Wilson, M.; Chmielewski, J., Targeting the dimerization interface of HIV-1 protease: inhibition with cross-linked interfacial peptides. *J Am Chem Soc* **1997**, *119*, 4841-4845.

14. Bowman, M.J.; Chmielewski, J., Novel strategies for targeting the dimerization interface of HIV protease with cross-linked interfacial peptides. *Biopolymers* **2002**, *66*, 126-133.

15. Hwang, Y.S.; Chmielewski, Development of low molecular weight HIV-1 protease dimerization inhbitors. *J Med Chem* **2005**, *48*, 2239-2242.

16. Lee, S.-G.; Chmielewski, J., Rapid synthesis and in situ screening of potent HIV-1 protease dimerization inhibitors. *Chem Biol* **2006**, *13*, 421-426.

17. Shultz, M.D.; Ham, Y.-W.; Lee, S.-G.; Davis, D.A.; Brown, C.; Chmielewski, J., Small-molecule dimerization inhibitors of wild-type and mutant HIV protease: a focused library approach. *J Am Chem Soc* **2004**, *126*, 9886-9887.

18. Bouras, A.; Boggetto, N.; Benatalah, Z.; de Rosny, E.; Sicsic, S.; Reboud-Ravaux, M., Design, synthesis, and evaluation of conformationally constrained tongs, new inhibitors of HIV-1 protease dimerization. *J Med Chem* **1999**, *42*, 957-962.

19. Merabet, N.; Dumond, J.; Collinet, B.; Van Baelinghem, L.; Boggetto, N.; Ongeri, S.; Ressad, F.; Reboud-Ravaux, M.; Sicsic, S., New constrained "molecular tongs" designed to dissociate HIV-1 protease dimer. *J Med Chem* **2004**, *47*, 6392-6400.

20. Frutos, S.; Rodrigues-Mias, R.A.; Madurga, S.; Collinet, B.; Reboud-Ravaux, M.; Ludevid, D.; Giralt, E., Disruption of the HIV-1 protease dimer with interface peptides: structural studies using NMR spectroscopy combined with [2-^{13}C]-Trp selective labeling. *Biopolymers* **2007**, *88*, 164-173.

21. Bowman, M.J.; Byrne, S.; Chmielewski, J., Switching between allosteric and dimerization inhibition of HIV-1 protease. *Chem Biol* **2005**, *12*, 439-444.

22. Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz Jr., K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* **1995**, *117*, 5179-5197.

23. Halgren, T.A., MMFF VI. MMFF94s option for energy minimization studies. *J Comp Chem* **1999**, *20*, 720-729.

24. Case, D.A.; Darden, T.A.; Cheatham III, T.E.; Simmerling, C.L.; Wang, J.; Duke, R.E.; Luo, R.; Merz, K.M.; Wang, B.; Pearlman, D.A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J.W.; Ross, W.S.; Kollman, P.A. (2004), AMBER 8, University of California, San Francisco

25. Case, D.A.; Darden, T.A.; Cheatham III, T.E.; Simmerling, C.L.; Wang, J.; Duke, R.E.; Luo, R.; Crowley, M.; Walker, R.C.; Zhang, W.; Merz, K.M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K.F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S.R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D.H.; Seetin, M.G.; Sagui, C.; Babin, V.; Kollman, P.A. (2008), AMBER 10, University of California, San Francisco

26. Wickstrom, L.; Okur, A.; Simmerling, C., Evaluating the performance of the ff99SB force field based on NMR scalar coupling data. *Biophys J* **2009**, *97*, 853-856.

27. Pearlman, D.A.; Case, D.A.; Caldwell, J.W.; Ross, W.S.; Cheatham III, T.E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P., AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp Phys Commun* **1995**, *91*, 1-41.

28. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L., Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **1983**, *79*, 926-935.

29. Feig, M.; Onufriev, A.; Lee, M.S.; Im, W.; Case, D.A.; Brooks III, C.L., Performance comparison of generalized born and poisson methods in the calculation of electrostatic solvation energies for protein structures. *J Comp Chem* **2003**, *25*, 265-284.

30. Feig, M.; Karanicolas, J.; Brooks III, C.L. (2001), MMTSB Tool Set, MMTSB NIH Research Resource, The Scripps Research Institute

Chapter 4

HIV-1 protease dimer stability: the role of critical residues to anchor the active protease dimer

Introduction

A high replication error rate allows HIV-1 protease (HIVp) to easily escape drug efficacy and coupled with selective pressure produced from active site competitive inhibitors.[1] The multi-drug resistant variants of HIVp necessitates alternative mechanisms to inhibit HIVp to fight the propagation of the HIV virion. HIVp must dimerize to form an active catalytic site and the HIVp dimer interface is highly conserved.[2, 3] These two features of the HIVp dimer interface create an attractive target to block the protein-protein interaction stabilizing the dimeric HIVp. Dissociative inhibition of the HIVp originates from the creation of peptidomimetic compounds, derived from the N- and C-termini at the dimer interface.[4, 5]

The structural stability of the HIVp dimer interface was studied by Todd, *et al.*[6] Using differential scanning calorimetry, Todd characterized the Gibbs free energy resulting from dimerization. At pH 5 and 298 K, the free energy gained from dimerization is 14.5 kcal/mol. The interdigitating β-sheet at the dimer interface accounted for approximately 75% of the free energy gained upon dimerization. Several of the residues in the C-terminus (Phe99, Asn98 and Leu97) were determined to be highly responsible for stabilizing the dimer interface. Xie *et al.* studied HIVp

dimerization at a more physiological, neutral pH. Xie reported the dissociative equilibrium constant of HIVp as 5.8 μM.[7] Substrate or competitive inhibitor bound to HIVp shifted the equilibrium toward a catalytically active dimeric state.

Deletion of the last four residues in the C-terminus dissociates HIVp dimer to a monomeric state, as reported by Ishima, et. al, using NMR.[8] This monomeric state of HIVp ($HIVp^{1-95}$) was deposited in the Protein Data Bank repository (PDB: 1Q9P). A stable monomer unit was predicted prior to Ishima's work by Levy *et al.*, after performing several Gō-model simulations to fold monomeric HIVp.[9]

Bogan and Thorn collected a database of alanine-substituted mutations involved in protein-protein interactions and the free energy differences on the stabilization of the protein-protein interaction.[10] Their study found no correlation in buried residue surface area and free energy contribution. Instead, Bogan and Thorn discovered what is now known as "hot spots" in protein-protein interactions. Hot spots are unevenly distributed areas in the protein-protein interfaces, and contribute disproportionally to the free energy of protein-protein oligomerization. These hot spots can be incorporated in the design of ligands to disrupt protein-protein interactions.

We know the free energy contributions to stabilize HIVp dimerization on a per residue basis, due to the work of Todd and colleagues; however, there is little information describing the interactions on the protein surfaces these residues are in contact with. We will characterize the hot spot residue's role within the dimer interface of HIVp and incorporate their contributions to create a pharmacophore based on mapped hotspots. Pharmacophores can be created by determining the density map of the residue's position

through the course of a trajectory. The pharmacophore chemical properties will be assigned according to the residue's chemical property (i.e. Phe 99 is hydrophobic and aromatic). A new pharmacophore based on the hot spot map of HIVp can filter a small molecule library complementing the properties of the hot spots or incorporated into previous pharmacophores (see Chapter 2).

Using implicit solvent molecular dynamics simulations with a sufficient Langevin viscosity term (γ), we can sample the $HIVp^{1-95}$ monomer from a dimeric form.[11] The amount of sampling needed to dissociate HIVp will be determined by recreating the $HIVp^{1-95}$ monomer in implicit solvent simulation. The $HIVp^{1-95}$ should provide a baseline trajectory length for further HIVp truncations or mutation we investigate.

Methods

Protein preparation and setup

The initial steps of LD setup are similar to MD setup. HIVp atomic coordinates are obtained from the Protein Data Bank (PDB) repository (PDB ID: 1HHP). AMBER10 and FF99SB parameters were used to perform LD simulations.[12, 13] The aGB^{OBC}(II) parameters were used for the Generalized Born approach to model the solvent implicitly.[14, 15] A 999 Å cutoff was used for the non-bonded interactions, $\gamma = 1\ ps^{-1}$ for the Langevin equation, 1 fs time-step and no boundary conditions were used. Default dielectrics were used: 78.5 for the exterior and 1 for the interior.

The HIVp symmetry partner was generated to create a homodimer by the C_2 symmetry operation in PyMOL.[16] The last four residues in the dimeric 1HHP C-terminus were truncated to resemble a $HIVp^{1-95}$ dimer. The residues Cys95 and Cys95' were

transformed with carboxyl termini and coordinates were then exported from PyMOL. The dimeric HIVp$^{1-95}$ was imported into the AMBER8 tLEaP program and hydrogen atoms were added upon import.[17, 18]

Two minimization steps were required prior to equilibration. Hydrogen atoms were minimized first, followed by an all-atom minimization. The equilibration procedure is a series of steps to slowly unrestrain the protein atoms and to ramp the temperature from 100 to 300K. The first equilibration step increased the temperature from 100 K to 200 K and had a restraint (2 kcal/mol) on all heavy atoms for 10 ps. The second equilibration step increased the temperature from 200 K to 300 and a reduced restraint (1 kcal/mol) on the heavy atoms. The next two equilibration steps further reduce the restraint, 0.5 kcal/mol followed by 0.1 kcal/mol, for 10 ps each. Side chain atoms were released from the restraint, while the backbone atoms remained restrained (0.1 kcal/mol), for 10 ps. The last equilibration step included 300 ps of unrestrained production. Ten separate 20 ns LD simulations were performed with unique seed numbers, for a total of 200 ns of LD production.

Results and Discussion

HIVp$^{1-95}$ Simulations

A monomeric state of HIVp$^{1-95}$ was previously characterized, PDB 1Q9P.[8] We expected our simulations to result in a similar monomeric state after initially beginning in dimeric form, reproducing the HIVp$^{1-95}$ monomer. The C-terminal truncated HIVp$^{1-95}$ would serve as a baseline for other mutational HIVp variants and to explore regions of the dimer interface to map the interface hot spots. Unfortunately, we did not reproduce

the monomer HIVp$^{1-95}$ in all but one of our simulations. Figure 4.1 below describes the measured distance between the two alpha carbons in the two catalytic Asp residues.

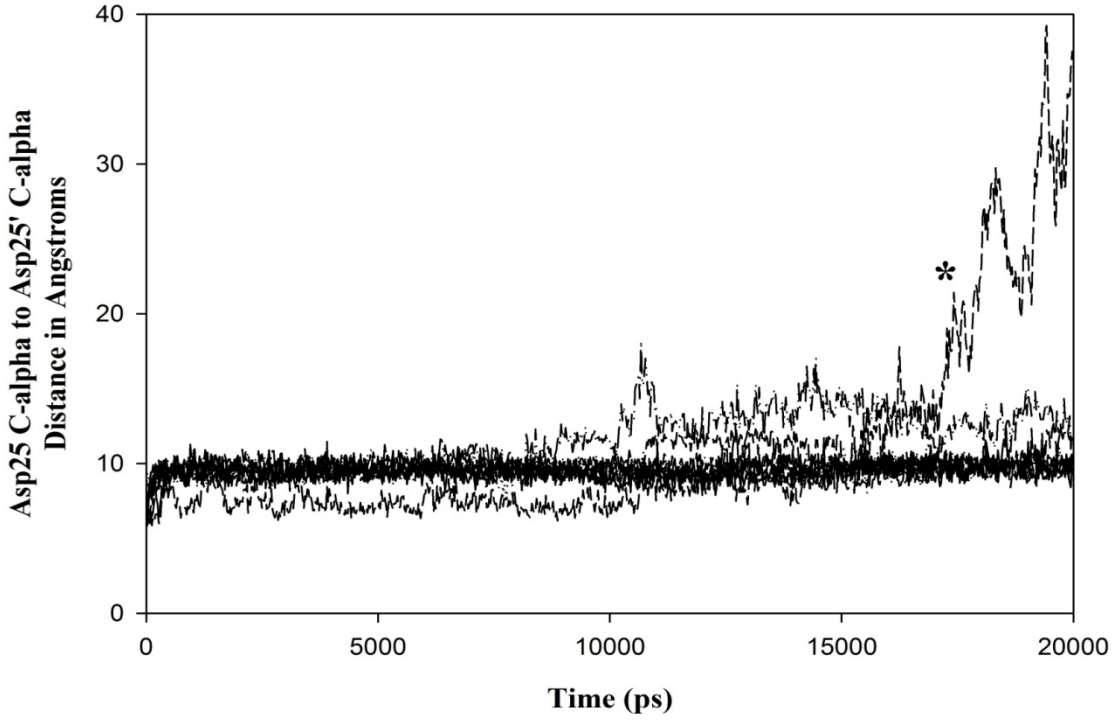

Figure 4.1: Distance between catalytic Asp Cα atoms in HIVp$^{1-95}$ LD simulation. Ten LD dynamic simulations were produced modeling the C-terminal truncated HIVp$^{1-95}$. LD production 3, indicated with the asterick (*), dissociated while the remainder LD trajectories continued as dimers. For reference, the equivalent atom distance in PDB 1HHP is 6.4 Angstroms.

Besides LD production 3, none of the remainder trajectories resulted in a monomeric HIVp$^{1-95}$. The third trajectory did demonstrate that our simulations of the dimeric HIVp$^{1-95}$ can dissociate, but the amount of time needed for the remainder to dissociate is unknown. The amount of computational hours needed for HIVp$^{1-95}$ and any variant of HIVp to map protein hot spots can rapidly increase. Following studies by Todd and colleagues on the stability of HIVp, there are several uncharacterized candidate residue interactions we could investigate. Some previously characterized interactions, such as the inter- and intra-hydrogen bond network of Arg87-Asp29-Arg8' described by

Louis, *et al.*, do not need to be defined computationally; however, there are a number of important residues to consider. The top six free energy contributors are Phe99, Asn98, Leu97, Leu5, Pro1 and Cys95, in that order. We halted the investigation due to the large demand on computational time and the resources needed to accomplish the investigation to completion.

Conclusion

A map of the protein hot spots at the HIVp dimer interface is still a worthy goal; although, a different strategy would be necessary. A computational method, such as UMBRELLA sampling, could provide the detailed information needed and would result in a monomeric form of $HIVp^{1\text{-}95}$ in silico.[19] A distance restraint between the centers of each HIVp core, with the properties of a spring force, will produce free energy differences between the normal and truncated HIVp dimer. The computational simulations will result in a monomeric protease because the restraint distance is user defined and will increase in length to force dissociation. UMBRELLA sampling will map the energy landscape as the HIVp variants are dissociated and the free energy landscapes of the truncated protease variants can be compared to the wild-type HIVp.

References

1. Preston B.D.; Poiesz, B.J.; Loeb, L.A., Fidelity of HIV-1 reverse transcriptase. *Science* **1988**, *242*, 1168-1171.

2. Babe L.M.; Rose, J.; Craik, C.S., Synthetic "interface" peptides alter dimeric assembly of the HIV 1 and 2 proteases. *Protein Sci* **1992**, *1*, 1244-1253.

3. Gustchina A. and Weber, I.T., Comparative analysis of the sequences and structures of HIV-1 and HIV-2 proteases. *Proteins* **1991**, *10*, 325-339.

4. Franciskovich J.; Houseman, K.; Mueller, R.; Chmielewski, J., A systematic evaluation of the inhibition of HIV-1 protease by its C- and N-terminal peptides. *Bioorg Med Chem Lett* **1993**, *3*, 765-768.

5. Schramm H.J.; Billich, A.; Jaeger, J.; Rucknagel, K.-P.; Arnold, G.; Schramm, W., The inhibition of HIV-1 protease by interface peptides. *Biochem Biophys Res Commun* **1993**, *194*, 595-600.

6. Todd M.J.; Semo, N.; Freire, E., The structural stability of the HIV-1 protease. *JMB* **1998**, *283*, 475-488.

7. Xie D.; Gulnik, S.; Gustchina, E.; Yu, B.; Wei, Shao; Qoronfleh, W.; Nathan, A.; Erickson, J.W., Drug resistance mutations can affect dimer stability of HIV-1 protease at neutral pH. *Protein Sci* **1999**, *8*, 1702-1707.

8. Ishima R.; Torchia, D.A.; Lynch, S.M.; Gronenborn, A.M.; Louis, J.M., solution structure of the mature HIV-1 protease monomer. *J Biol Chem* **2003**, *278*, 43311-43319.

9. Levy Y. and Caflisch, A., Flexibility of monomeric and dimeric HIV-1 protease. *J Phys Chem B* **2003**, *107*, 3068-3079.

10. Bogan A.A. and Thorn, K.S., Anatomy of hot spots in protein interfaces. *JMB* **1998**, *280*, 1-9.

11. Zagrovic B.; Pande, V., Solvent viscosity dependence of the folding rate of a small protein: distributed computing study. *J Comp Chem* **2003**, *24*, 1432-1436.

12. Case D.A.; Darden, T.A.; Cheatham III, T.E.; Simmerling, C.L.; Wang, J.; Duke, R.E.; Luo, R.; Crowley, M.; Walker, R.C.; Zhang, W.; Merz, K.M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K.F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S.R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D.H.; Seetin, M.G.; Sagui, C.; Babin, V.; Kollman, P.A. (2008), AMBER 10, University of California, San Francisco

13. Wickstrom L.; Okur, A.; Simmerling, C., Evaluating the performance of the ff99SB force field based on NMR scalar coupling data. *Biophys J* **2009**, *97*, 853-856.

14. Onufriev A.; Bashford, D.; Case, D.A., Modification of the generalized Born model suitable for macromolecules. *J Phys Chem B* **2000**, *104*, 3712-3720.

15. Feig M.; Onufriev, A.; Lee, M.S.; Im, W.; Case, D.A.; Brooks III, C.L., Performance comparison of generalized born and poisson methods in the calculation of electrostatic solvation energies for protein structures. *J Comp Chem* **2003**, *25*, 265-284.

16. The PyMOL Molecular Graphics System, Version 1.2, Schrödinger, LLC

17. Case D.A.; Darden, T.A.; Cheatham III, T.E.; Simmerling, C.L.; Wang, J.; Duke, R.E.; Luo, R.; Merz, K.M.; Wang, B.; Pearlman, D.A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J.W.; Ross, W.S.; Kollman, P.A. (2004), AMBER 8, University of California, San Francisco

18. Pearlman D.A.; Case, D.A.; Caldwell, J.W.; Ross, W.S.; Cheatham III, T.E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P., AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp Phys Commun* **1995**, *91*, 1-41.

19. Kumar S.; Rosenberg, J.M.; Bouzida, D.; Swendsen, R.H.; Kollman, P.A., The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *J Comp Chem* **1992**, *13*, 1011-1021.

Chapter 5

Future Directions: Incorporating Hydrogen-Deuterium Exchange/ Mass Spectrometry Techniques to Determine Ligand Induced Dissociative Mechanism

Introduction

As of this dissertation's publication, there is no detailed structural information regarding the relationship between the HIVp dimer interface and small molecules causing dissociative inhibition. All HIVp dimer inhibitors are characterized by the Zhang-Poorman kinetic assay and published investigations were operating under the theoretical presumption based on the HIVp kinetic model.[1] An NMR study by Frutos and co-workers has shown a monovalent, C-terminal peptidomimetic inhibitor causes chemical shifts at the Trp6 residue; however, their docked simulation portraying the protein-ligand relationship is not conclusive evidence of the binding mechanism.[2] We know that SYEL, C-terminal derived ligand is likely bound near the Trp6 residue, but greater atomic resolution is needed to thoroughly understand the protein-ligand relationship. The relationship is necessary for building better, non-peptidomimetic inhibitors.

Our previous study modeled the protein-ligand interaction between the dimer interface and Chmielewski's peptidomimetic derivatives, linked with an alkyl chain, were indeterminate regarding ligand binding. Despite our computational models being inconclusive to illustrate the binding mechanism, the models did show the putative mechanism of binding to be unfavorable. There is a large amount of disorder present at

the protease monomer interface. The N- and C-termini accounted for approximately 6 and 9 Å for the N- and C-termini, respectively, in our study (Chapter 2). For the bivalent ligand to bind in an anti-parallel β-sheet with the protease monomer termini, there would be a large entropic penalty due to the degree of termini motions. More empirical, structural information should result in more determinate computational models by providing better initial coordinates for future MD/ LD simulations. Greater information regarding the ligand's initial coordinates will position the ligand closer to its energetically favorable binding mechanism and permit better computational models to understand the protein-ligand relationship.

Hydrogen-Deuterium Exchange

Hydrogen-deuterium exchange (HDX) coupled with mass spectrometry (MS) has gained popularity as a valuable analytical tool for the study of protein dynamics, protein folding, protein-protein interactions and protein-ligand binding.[3] The principles of HDX are simple; any amide hydrogen atom along the protein backbone (minus proline residues) inherently exchanges with protons in solution (known as hydrogen exchange, or HX). HX allows the incorporation of free deuterium atoms in solution ($\geq$ 99% deuterium content) to interchange for hydrogen atoms, incorporating deuterium into the amide backbone. The additional mass due to the deuterium atom can be resolved with MS and the location of the incorporated deuterium may be determined by cleaving the protein into short oligopeptides.

HX depends on the availability of the exposed amide hydrogen to exchange with the environment. Exchange can occur quickly, within 1-10 seconds, for some amide hydrogens; however, HX can require longer incubation in solution (deuterated or not) if a domain motion or partial protein unfolding events is required to expose the amide hydrogen. For a protein's amide hydrogen to exchange with deuterium atoms the protein must be in a deuterated solution, obviously, but there are two methods to introduce deuterium for amide HX. Choosing a method, dilution or rapid buffer switch, for deuterium introduction is determined by the MS and the instrument's ability to identify small quantities of protein. Dilution is the easier method of the two, involving the dilution of the protein in a deuterated solution. Diluted samples may need to be concentrated after dilution prior to MS detection, typically in an HPLC column under high pressure. In the rapid buffer switch approach, a protein solution is switched using a gel filtration spin column.

Dependent on the needs of the researcher, there are two methods (pulsed or continuous) to label the protein.[4] The first method, pulse-labeling method, is frequently used for studies involving protein folding. Since a protein folds rapidly, the best method to study protein folding is by slowly unfolding the protein.[5] In a pulse-labeling experiment, a perturbant (i.e. urea or a change in pH) is added to the solution and equilibrated. The perturbed protein, now in a partially unfolded state, may then be briefly deuterated labeling the newly exposed, less stable regions of the protein.[6] To capture the multiple kinetic intermediates involved in protein folding, the perturbation, equilibration, deuteration and analysis steps would need to be iterated over several pertubant

concentrations. Automation, such as a quench-flow scheme, of the pulse-labeling sequence of events may be required for analysis reproducibility.[7]

The second method, the continuous labeling method, is the more favorable method to label a protein because it is technically easier to perform. Continuous labeling is used for slower transitions to unfolded states. In continuous labeling experiments, a population of states exists within a mixture, allowing transitions from one state to another as a pertubant (pH change, urea, or a small molecule inhibitor) is added to the mixture and deuterated for several fixed intervals. In theory, every state including intermediate states can be captured; however, the amount of sample in the low populated states will likely be too miniscule for accurate MS detection.

One problem naturally associated with HDX is the loss of deuterium label due to back-exchange. Retention of the label can be assisted by change in solution conditions, such as freezing the sample or an acidic pH change; however, the amount of time between labeling and analysis remains a primary concern. For electrospray ionization (ESI) loss of label can occur at the HPLC step; for matrix-assisted laser desorption/ionization (MALDI) experiments, loss of label occurs primarily during the sample preparation step. Additional loss of the deuterium label can occur if a proteolysis step is included prior to analysis. Several calculations may be performed to compensate for the amount of label loss, but the calculations do not significantly compensate for the information lost when compared to the initial deuterated sample.[8]

Protein fragmentation is used to help decipher the small, local changes a deuterium ion will cause. To receive a higher spatial resolution, the labeled protein is

fragmented, often by proteases, into smaller fragments (5-10 residues) for analysis rather than using the whole labeled protein. Since buffer conditions are acidic to discourage deuterium back-exchange, fragmentation is limited to acidic proteases. Pepsin is the most common of these acidic proteases and other acidic proteases may be used; however, pepsin is still a more efficient protease.[9] Pepsin is a non-specific enzyme, cutting along hydrophobic residues, but pepsin digestion is highly reproducible given equivalent proteolytic conditions.[10, 11] Investigators tested multiple proteases to give smaller, overlapping fragments increasing spatial resolution.[12] There are methods using non-protease fragmenting techniques, but the methods are still in the development stage.[13]

For detection of the labeled protein, there are two major MS techniques that can be used to detect the labeled protein: ESI and MALDI. More deuterium back-exchange occurs during MALDI analysis and the loss of deuterium causes the spectrum resolution to be less distinct and more difficult to interpret.[14] In addition to the back-exchange problem, since there is no HPLC procedure prior to MALDI-MS, all peptide lengths are present concurrently in the MS and crowd the spectrum peaks. Due to its disadvantages when compared to the ESI technique, MALDI is less commonly applied in HDX experiments.

ESI is a technique, similar to an aerosol and used to introduce a solvated protein system into a gas-phase environment for MS detection and analysis. ESI applies a voltage through an emitter across a solution and emits the positively charged solution through a Taylor Cone as liquid droplets.[15] The positively charged droplets will shrink in size, as the droplets evaporate, becoming a positively charged nanodroplet.[16] As the nanodroplet continues to shrink in size, the positive charge of the nanodroplet will cause

the protein in solution to eject itself from the nanodroplet and into the gas-phase known as ion evaporation model (IEM). In IEM, as the radius of the droplet decreases the field strength at the droplet surface becomes strong enough to cause desorption.[17, 18] The protein, or peptide fragment, can then be detected by the MS detector.

MALDI is a two step process allowing the detection of biomolecules.[19] The first step is the absorption of UV light emitted from a laser beam into the matrix solution.[20] The matrix solution is composed of biomolecules (the analyte), purified water, organic solvent (i.e. acetonitrile) and crystallized molecules (i.e. 3,5-dimethoxy-4-hydroxycinnamic acid, 2,5-dihydroxybenzoic acid).[21, 22] Absorption of the UV light causes ablation of the top layer of matrix solution.[20] Ablation produces a plume with a mixture of matrix and analyte where the second MALDI step occurs, ionization of the solution. The process of ionization is not completely understood. Three hypothetical models exist: the photochemical ionization, the cluster ionization model, and the energy induced disproportionate model; however, each model lacks empirical validation.[23] The ionized plume then undergoes the IEM and desorption process similar to the ionized nanodroplets in ESI, introducing the analyte to the gas phase.

Several protein binding interactions were characterized using HDX: protein-protein interactions, protein-polypeptide, protein-nucleic acid, protein-lipid and protein-small molecule. Protein-ligand interactions determined by HDX include: the substrate- and inhibitor-bound Escherichia coli dihydrodipicolinate reductase, ATP-bound α-crystallin and several ligand interactions with MAP kinases.[24-26] Mapping ligand interactions with HDX is successful and can be extended to HIVp applications. The

dimer HIVp protein is small, approximately 24 kDa, and the inherent C_2 symmetry provides two monomer-dimer inhibitor interactions for every HIVp dimer.

HIV-1 protease

Currently, there are no traditional biophysical methods to characterize the protein-ligand relationship. The intrinsic entropy at the N- and C-termini in the dimer interface creates a large amount of disorder for X-ray diffraction patterns for crystallographic efforts. NMR has solubility issues associated with the ligands.[2] The computational simulations previously presented in this dissertation failed to converge on a consensus binding mode for Chmielewski's dissociative inhibitors. There is a growing amount of research which coupled low- to medium-resolution experiments with other biophysical techniques, hybridizing the results, and creating a structural model from the combined information.[27]

HDX is well suited as a technique to characterize the binding mechanism of protease dissociative inhibitors. The protease monomer is small, 11.8 kDa in size, and the inherent symmetry of HIVp homodimer will provide twice the monomer structural information per a dissociative event. MS benefits from accurate sample detection at low concentrations (nM), similar to those used *in vitro* assay conditions. The dimer interface contains several amides in the dimer interface that will be exposed for HX in solution upon dissociation. The N- and C-termini, normally hydrogen-bonded in an interdigitating β-sheet, will become more exposed upon dissociative inhibition. Active

site residues, associated in the hydrogen-bonding network described as the "Fireman's Grip", will also become unprotected and available for HX.[28]

The method may provide structural information to characterize the binding mechanism for future allosteric inhibitors. Inhibitors targeting the "cheek" and "eye" regions of HIVp currently do not have empirical structural validation.[29, 30] A recent study by Chang et. al. is an example of allosteric inhibition of HIVp without structural validation.[31] Chang and colleages performed a library screen against wild-type and a mutant strain of HIVp. The library screen resulted in one compound shown to inhibit HIVp a series of assays (Michealis-Menten, and Yonetain-Theorell assays) suggesting the compound inhibits through an allosteric mechanism. The authors eliminated a dimer inhibition as the mechanism for their reported compound, by comparing similar IC50 values between the monomeric form of HIVp and the IC50 value for the dimeric HIVp. Recognizing the cheek region as site for allosteric control, as reported by Perryman and coworkers, Chang and colleagues then performed computational docking experiments against the allosteric groove. As of publication of this dissertation, there is no structural validation of the bound protein-ligand relationship for the compound.

Protein-protein and protein-ligand interactions HDX studies are immediately related to our interests; however, HDX studies can be extended to include HIVp monomer folding. Computational simulations of HIVp monomer folding were conducted by Levy et. al. Levy found two regions, term local elementary structures (LES), responsible for stabilizing the HIVp monomer.[32] Broglia, et. al. attempted to block monomer folding by developing peptidomimetics based on the LES peptides, destabilizing the HIVp monomer.[33] A circular dichroism spectroscopy experiment and

Lineweaver-Burk kinetic analysis was performed to suggest the inhibition mechanism was due to monomer unfolding; however, there is no direct structural evidence of the protein-ligand relationship.

Protein-protein, protein-ligand and protein-folding are all areas HDX can contribute structural information. The potential for hydrogen-Deuterium Exchange/Mass Spectrometry applies towards HIVp is large due to its therapeutic and scientific value. The significance of the work can contribute to future generations of HIVp inhibitors and prevent affecting inhibitor binding without adversely influencing HIVp enzymatic activity by destabilizing the dimeric protease fold.

References

1. Zhang, Z.-Y.; Poorman, R.A.; Maggiora, L.L.; Heinrikson, R.L.; Kezdy, F.J., Dissociative inhibition of dimeric enzymes. *J Biol Chem* **1991**, *266*, 15591-15594.

2. Frutos, S.; Rodrigues-Mias, R.A.; Madurga, S.; Collinet, B.; Reboud-Ravaux, M.; Ludevid, D.; Giralt, E., Disruption of the HIV-1 protease dimer with interface peptides: structural studies using NMR spectroscopy combined with [2-^{13}C]-Trp selective labeling. *Biopolymers* **2007**, *88*, 164-173.

3. Konerman, L.; Pan, J.; Liu, Y.H., Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem Soc Rev* **2011**, *40*, 1224-1234.

4. Deng, Y.; Zhang, Z.; Smith, D.L., Comparison of continuous and pulsed labeling amide hydrogen exchange/mass spectrometry for studies of protein dynamics. *J Am Soc Mass Spectrom* **1999**, *10*, 675-684.

5. Fersht, A.R.; Dagget, V., Protein folding and unfolding at atomic resolution. *Cell* **2002**, *108*, 1-20.

6. Mazon, H.; Marcillat, O.; Forest, E.; Smith, D.L.; Vial, C., Conformational dynamics of the GdmHCl-induced molten globule state of creatine kinase monitored by hydrogen exchange and mass spectrometry. *Biochemistry* **2004**, *43*, 5045-5054.

7. Yang, H.; Smith, D.L., Kinetics of cytochrome c folding examined by hydrogen exchange and mass spectrometry. *Biochemistry* **1997**, *36*, 14992-14999.

8. Zhang, Z.; Smith, D.L., Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci* **1993**, *2*, 522-531.

9. Wang, L.; Pan, H.; Smith, D.L., Hydrogen exchange-mass spectrometry: optimization of digestion conditions. *Mol Cell Proteomics* **2002**, *1*, 132-138.

10. Dunn, B.M., Overview of pepsin-like aspartic peptidases. *Curr Protoc Protein Sci* **2001**, *21*, 3.1-3.6.

11. Busby, S.A.; Chalmers, M.J.; Griffin, P.R., Improving digestion efficiency under H/D exchange conditions with activated pepsinogen coupled columns. *Int. J. Mass Spectrom.* **2007**, *259*, 130-139.

12. Garcia, R.A.; Pantazatos, D.; Villarreal, F.J., Hydrogen/deuterium exchange mass spectrometry for investigating protein-ligand interactions. *Drug Dev Technol* **2004**, *2*, 81-91.

13. Pan, J.; Han, J.; Borchers, C.H.; Konerman, L., Hydrogen/deuterium exchange mass spectrometry with top-down electron capture dissociation for characterizing structural transitions of a 17 kDa protein. *J Am Chem Soc* **2009**, *131*, 12801-12808.

14. Hoofnagle, A.N.; Resing, K.A.; Ahn, N.G., Protein analysis by hydrogen exchange mass spectrometry. *Annu Rev Biophys Biomol Struct* **2003**, *32*, 1-25.

15. Fenn, J.B.; Mann, M.; Meng, C.K.; Wong, S.F.; Whitehouse, C.M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246*, 64-71.

16. Konerman, L, A simple model for the disintegration of highly charged solvent droplets during electrospray ionization. *J Am Soc Mass Spectrom* **2009**, *20*, 465-506.

17. Iribarne, J.V.; Thomson, B.A., On the evaporation of small ions from charged droplets. *J Chem Phys* **1976**, *64*, 2287-2294.

18. Ahadi, E.; Konermann, L., Ejection of Solvated Ions from Electrosprayed Methanol/Water Nanodroplets Studied by Molecular Dynamics Simulations. *J Am Chem Soc* **2011**, *133*,

19. El-Aneed, A.; Cohen, A.; Banoub, J., Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *Appl Spectrosc Rev* **2009**, *44*,

20. Knochenmuss, R., Ion formation mechanisms in UV-MALDI. *Analyst* **2006**, *131*, 966-986.

21. Beavis, R.C.; Chait, B.T., Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins. *Rapid Commun Mass Spectrom* **1989**, *3*, 432-435.

22. Strupat, K.; Karas, M.; Hillenkamp, F., 2,5-Dihidroxybenzoic acid: a new matrix for laser desorption-ionization mass spectrometry. *Int J Mass Spectrom Ion Processes* **1991**, *72*, 89-102.

23. Chang, W.C.; Huang, L.C.; Wang, Y.S.; Peng, W.P.; Chang, H.C.; Hsu, N.Y.; Yang, W.B.; Chen, C.H., Matrix-assisted laser desorption/ionization (MALDI) mechanism revisted. *Anal Chim Acta* **2007**, *582*, 1-9.

24. Wang, F.; Blanchard, J.S.; Tang, X.J., Hydrogen exchange/electrospray ionization mass spectrometry studies of substrate and inhibitor binding and conformational changes of Escherichia coli dihrdrodipicolinate reductase. *Biochemistry* **1997**, *36*, 3755-3759.

25. Hasan, A.; Smith, D.L.; Smith, J.B., Alpha-crystallin regions affected by adenosine 5'-triphosphate identified by hydrogen-deuterium exchange. *Biochemistry* **2002**, *41*, 15876.

26. Lee, T.; Hoofnagle, A.N.; Kabuyama, Y.; Stroud, J.; Mix, X.; Goldsmith, E.J.; Chen, L.; Resing, K.A.; Ahn, N.G., Docking motif interactions in MAP kinases revealed by hydrogen exchange mass spectrometry. *Mol Cell* **2004**, *14*, 43-55.

27. Henzler-Wildman, K.; Kern, D., Dynamic personalities of proteins. *Nature* **2007**, *450*, 964-972.

28. Strisovsky, K.; Tessmer, U.; Langer, J.; Konvalinka, J.; Krausslich, H.G., Systematic mutational analysis of the active-site threonine of HIV-1 protease: rethinking the "fireman's grip" hypothesis. *Protein Sci* **2000**, *9*, 1631-1641.

29. Perryman, A.L.; Lin, J.-H.; McCammon, J.A., HIV-1 proteinase molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. *Protein Sci* **2004**, *13*, 1108-1123.

30. Damm, K.L.; Ung, P.M.U.; Quintero, J.J.; Gestwicki, J.E.; Carlson, H.A., A poke in the eye: inhibition HIV-1 proteinase through its flap-recognition pocket. *Biopolymers* **2008**, *89*, 643-652.

31. Chang, M.W.; Giffin, M.J.; Muller, R.; Savage, J.; Lin, Y.C.; Hong, S.; Jin, W.; Whitby, L.R.; Elder, J.H.; Boger, D.L; Torbett, B.E., Identification of broad-based HIV-1 protease inhibitors from combinatorial libraries. *Biochem J* **2010**, *429*, 527-532.

32. Levy, Y.; Caflisch, A.; Onuchic, J.N.; Wolynes, P.G., The folding and dimerization of HIV-1 protease: evidence for a stable monomer from simulations. *JMB* **2004**, *340*, 67-79.

33. Broglia, R.A.; Provasi, D.; Vasile, F.; Ottolina, G.; Longhi, R.; Tiana, G., A folding inhibitor of the HIV-1 protease. *Proteins* **2006**, *62*, 928-933.

Appendix

Alternative Pharmacophore Model

Supplemental information from Chapter 2

Simplified pharmacophore model

A)

"Simplified" 3-ns MPS Pharmacophore Model					
Element Type		x	y	z	RMSD (Å)
1	Aro|Hyd	53.5953	45.1589	-2.3560	1.7277
2	Aro|Hyd	39.9640	51.0975	2.2804	2.2804
3	Aro|Hyd	49.7914	50.5249	6.5096	1.4436
4	Aro	44.9667	54.9316	7.7915	1.7782
5	Aro|Hyd	50.4455	54.0263	-7.6683	1.8469
6	Aro|Hyd	48.3853	50.9182	-1.6079	1.2943
7	Acc	54.0651	41.5045	-2.5040	1.8872
8	Don	39.4232	52.7636	2.8752	1.2407
9	Don|Acc	49.4266	52.9248	-8.0318	1.2849

B)

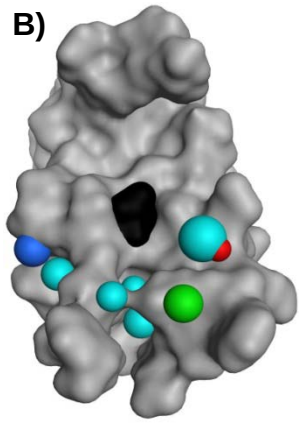

Figure A.1: Coordinates and diagram of the MPS pharmacophore model based on the "Simplified" model. A) The elements and coordinates are given relative to the 1HHP crystal structure. B) The pharmacophore sites are shown with 1×RMSD radii, colored cyan for aromatic, green for aromatic/hydrophobic, blue for hydrogen-bond acceptor, and red for hydrogen-bond donor interactions.

The simplified model, Figure 2.12, eliminated the hydrogen-bond donor to Thr26 in the 3-ns model, Figure 2.1(B). The element was eliminated due to it proximity to the binding site. The goal was to shift the focus toward the elements at the core of the binding interface. The same *in silico* screening procedure was used within the MOE program. A total of 288 hits were identified and clustered into three very general

chemical families. Twelve compounds were purchased for *in vitro* testing. Two weak inhibitors were identified, and the mechanism of dimerization inhibition was verified by Zhang-Poorman kinetic analysis, results in Figure 2.13 below.

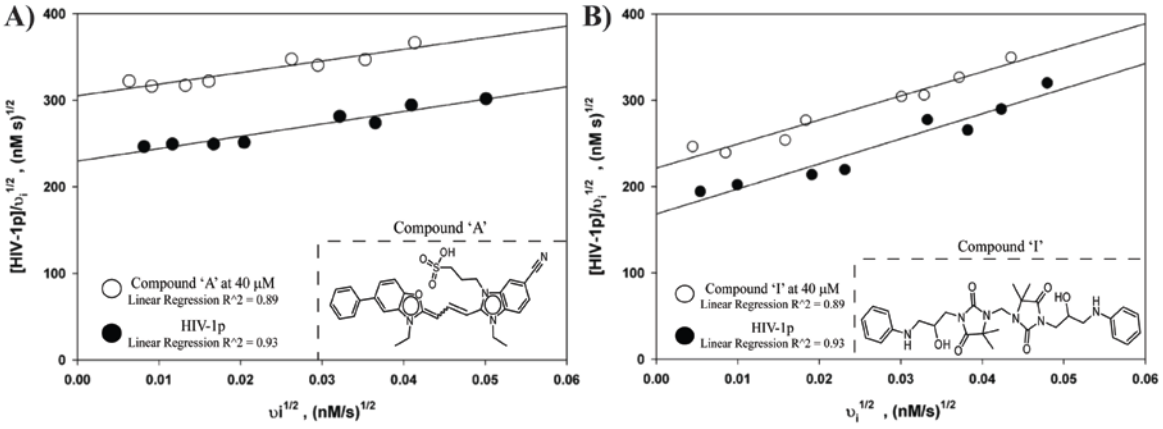

Figure A.2: 'Simplified' pharmacophore results. Testing of 12 compounds identified two inhibitors of HIVp which both targeted dimerization. A) The Zhang-Poorman analysis of compound 'A' resulted in a $K_{i,D}$ of 122 μM. B) The assay of compound 'I' resulted in a $K_{i,D}$ of 127 μM.

Center for Chemical Genomics (CCG) and Pharmacophore Screening Results. The following tables are the filtered results for each pharmacophore table, reported in Table 2.5. The table indicates the criteria used in the pharmacophore search for a small molecule to match the pharmacophore elements. Bold and italicized numbers are the conditional sets we further clustered by chemical similarity.

Table A.1: Number of hits obtained for *in silico* screening with the 3-ns model as the radii of the elements are increased and number of features is reduced

Radii Multiplier	Number of Elements				
(#×RMSD)	9 of 9	8 of 9	7 of 9	6 of 9	5 of 9
1	0	0	0	2	1328
1.33	0	0	0	***30***	–
1.67	0	0	6	674	–
2	0	0	***95***	3016	–
2.33	0	18	549	–	–
2.67	3	***151***	3830	–	–
3	***12***	714	–	–	–

Total number of compounds considered (in italics above): 12+151+95+30=288 with duplicates