APPLICATION OF FRAGMENT SCREENING BY X-RAY CRYSTALLOGRAPHY FOR THE IDENTIFICATION OF NOVEL CHEMICAL SCAFFOLDS TARGETING VIRAL PROTEINS OF HIV AND INFLUENZA

By

DISHA PATEL

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Medicinal Chemistry

written under the direction of

Dr. Eddy Arnold

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October 2013

ABSTRACT OF THE DISSERTATION

Application of Fragment Screening by X-ray crystallography For The Identification of Novel Chemical Scaffolds Targeting Viral Proteins of HIV and Influenza

By: Disha Patel

Dissertation Director:

Dr. Eddy Arnold

Fragment-based drug discovery (FBDD) is a relatively new approach for discovery of small molecule ligands for subsequent lead development. Here, libraries of relatively small molecules or fragments are screened against a target protein to identify hits with relatively weak binding affinity. Several approaches, which include fragment linking, merging, and expansion, can then be utilized to develop these hits into lead candidates.

The intrinsically weak binding affinity of fragments requires screening techniques capable of detecting very weak non-covalent interactions with the protein. Biophysical methods, such as nuclear magnetic resonance (NMR), X-ray crystallography, and surface plasmon resonance (SPR), have been traditionally used for FBDD. X-ray crystallography, in particular, has been underappreciated due to major drawbacks associated with low throughput and labor associated. However, advancements in rapid, automated data

collection has improved the throughput and reduced the labor associated with this approach.

This thesis reports the development of a high throughput screening method using X-ray crystallography and its successful application as a primary screening tool to identify new compounds for lead development against viral drug targets. Specifically, screening was conducted against HIV-1 reverse transcriptase (RT), the catalytic core domain of HIV-1 integrase (IN CCD), and the cap-snatching endonuclease domain of influenza A. Screening revealed fragment binding to promising, druggable allosteric sites within HIV-1 RT and IN CDD. The availability of a three dimensional structure proved to be highly advantageous as it allowed for structure-based drug design (SBDD) to be pursued. Using this approach, a fragment hit identified to bind to the endonuclease domain of influenza was developed into a new class of endonuclease inhibitors with antiviral activity. This thesis not only details the results from each of the screening campaign but also highlights many key advantages of crystallographic fragment screening.

Acknowledgments

The knowledge that I possess and the person that I am today are due to the contribution of countless of individuals and I am extremely grateful to them for their encouragement and support. First and foremost, I would like to thank my advisor, Eddy Arnold, for his constant support, generosity, encouragement, and mentorship. The depth of your knowledge and your unique analytical perspective are traits that I have always admired and strive to emulate.

I am also extremely thankful my committee members: Dr. Edmond LaVoie, Dr. Longquin Hu, and Dr. Joseph Marcotrigiano for being extremely kind and generous with their time and knowledge. A special thank you to the beamline staff at CHESS F1, Dave Schuller, Scott Smith, Bill Miller, and Kathy Dedrick, and at Brookhaven X25, Annie Heroux. I am extremely grateful to our many collaborators for their invaluable contributions and insight to the many projects I have had an opportunity to work.

I would be lost without the support of the current and previous members of the Arnold lab. I would like to thank Chhaya Dharia, my "lab mom", for being kind enough to teach me the tricks of the trade with regards to protein chemistry. But, I am even more grateful for your constant support and friendship. I am grateful to Arthur Clark (Dudley) for always answering my endless amount of questions.

I would like to thank Joe Bauman for being my go-to for pretty much everything. You are an amazing scientist and mentor but even a better friend. Thank you for pushing me to think "smarter" and, more importantly, keeping me sane.

I am grateful to Gail Ferstandig Arnold for her friendship and guidance. I am extremely grateful to Mary Fitzgerald Ho for being my walking manual for

crystallography and being a wonderful person to work with. I would like to thank Steve Tuske and William Ho for their amazing conversations and “unique sense” of humor. Many thanks to Vijayan for the enjoyable scientific and political discussions. I am also grateful to Tom Mariano, Rachel Bradley Rosen, Tiffany Tsay, Sergio Martinez, Kalyan Das, Brian Hudson, and Angela Fortner McKoy for their support and guidance. I would like to thank Barbara Shaver, Mary Haffey, and Minoti Sharma for the early morning conversations.

On a more personal note, I would like to thank my parents—Illa and Virendra—for making many sacrifices to support the dreams of their children. I am very fortunate to have my sisters, Sapna and Kesha, who always kept me line and supported my dreams. To my brothers, Nitesh and Shivam, thank you for believing in me and putting up with my early AM phone calls. To my niece, Hansini—you light up my world and nothing beats your hugs and kisses. I know you are a “big girl” now, but you will always be my poppet and I love you. I would also like to thank the rest of my family that is too big to thank individually. I will always be grateful for the never-ending support you all have provided me and I would not be the person that I am without you all.

I would also like to thank Gena Russo and Christdianne Villamor for your friendship for the past 13 years despite my nerdish ways. A big thank you to my twisted sister, Ankita Basant, and Priti Tiwari for always being there. I really don’t have the words to express what your friendship means to me and fortunately I don’t have to since our mental wavelengths are always aligned.

For my parents, Virendra and Illa Patel,

for the unconditional love

Table of Contents

List of Abbreviations

AIDS	Acquired immunodeficiency syndrome
CPMG	Carr-Purcell-Meiboom-Gill
D4T	Stavudine
DMSO	Dimethylsulfoxide
FBDD	Fragment-based drug discovery
FDA	The Food and Drug Administration
HAART	High active antiretroviral therapy
HIV-1	Human immunodeficiency virus type 1
HIV-2	Human immunodeficiency virus type 2
HTS	High-throughput screening
JDAM	Dual activity monitor assay
IN	HIV-1 integrase
ITC	Isothermal calorimetry
LEDGF	Lens epithelium-derived growth factor
LE	Ligand efficiency
MS	Mass spectrometry
MSCS	Multiple solvent crystal structure
NABS	NNRTI adjacent binding site
NHA	Non-hydrogen atom
NMR	Nuclear magnetic resonance spectroscopy
NNIBP	NNRTI-binding pocket
NNRTI	Non-nucleoside reverse transcriptase inhibitors

NRTI	Nucleoside reverse transcriptase inhibitors
NVP	Nevirapine
PA	Polymerase acidic protein
PA_N	N-terminal domain of the PA subunit
PAINS	Pan-assay interference compounds
pH1N1	Pandemic H1N1
PIC	Preintegration complex
PB1	Polymerase basic 1 protein
PB2	Polymerase basic 2 protein
pH1N1	Pandemic H1N1
PIC	Pre-integration complex
PR	HIV-1 protease
RdRP	RNA-dependent RNA polymerase
RPV	Rilpivirine
RT	HIV-1 reverse transcriptase
SIV	Simian immunodeficiency virus
SPR	Surface plasmon resonance
ssRNA	Single-stranded RNA
STD	Saturation transfer difference
TDF	Tenofovir
tDNA	Target DNA
TINS	Target-immobilized NMR screening
TMAO	Trimethylamine-N-oxide

vDNA Viral DNA

WaterLOGSY Water-ligand observed via gradient spectroscopy

List of Tables

List of Illustrations

Chapter 1: Fragment-Based Drug Discovery

Synopsis

Fragment-based drug discovery (FBDD) has become a powerful tool for the discovery of novel pharmacophores for subsequent lead and drug development. This chapter briefly describes the theoretical and experimental aspects of the fragment-based approach with strong emphasis on the aspects governing the design and curation of a fragment library.

1.1 Introduction

Drug discovery during the early 1990s was dominated by high-throughput screening (HTS) of libraries, consisting of large, drug-like compounds. The hits identified from these screenings were highly potent (μM to nM) however, subsequent lead optimization proved to be challenging due to their large size and hydrophobic nature (Congreve et al., 2005). Additionally, these libraries allowed for very inefficient sampling and coverage of chemical space. For compounds consisting of 30 non-hydrogen atoms, chemical space was estimated to consist of 10^{60} molecules (Bohacek et al., 1996). Thus, even screening a million compounds would allow only a tiny fraction of the chemical space to be sampled for compounds of this size. Other major drawbacks of this approach were the need for robust, high throughput biochemical assays for activity screening, low hit rates, false positives, and difficulties with lead optimization as well as securing intellectual property rights.

Over the past decade, FBDD has gained popularity as an effective and efficient drug discovery technique. The fragment-based approach involves screening of small molecule libraries against a target protein to identify weakly potent, bioactive molecules for further lead development. The smaller, less complex nature of fragments increases the probability of binding to the target protein resulting in higher hit rates. The low chemical complexity of the fragments allows for efficient search of diverse chemical space (Hadjuk et al., 2007; Hesterkamp et al., 2008; Erlanson et al., 2004).

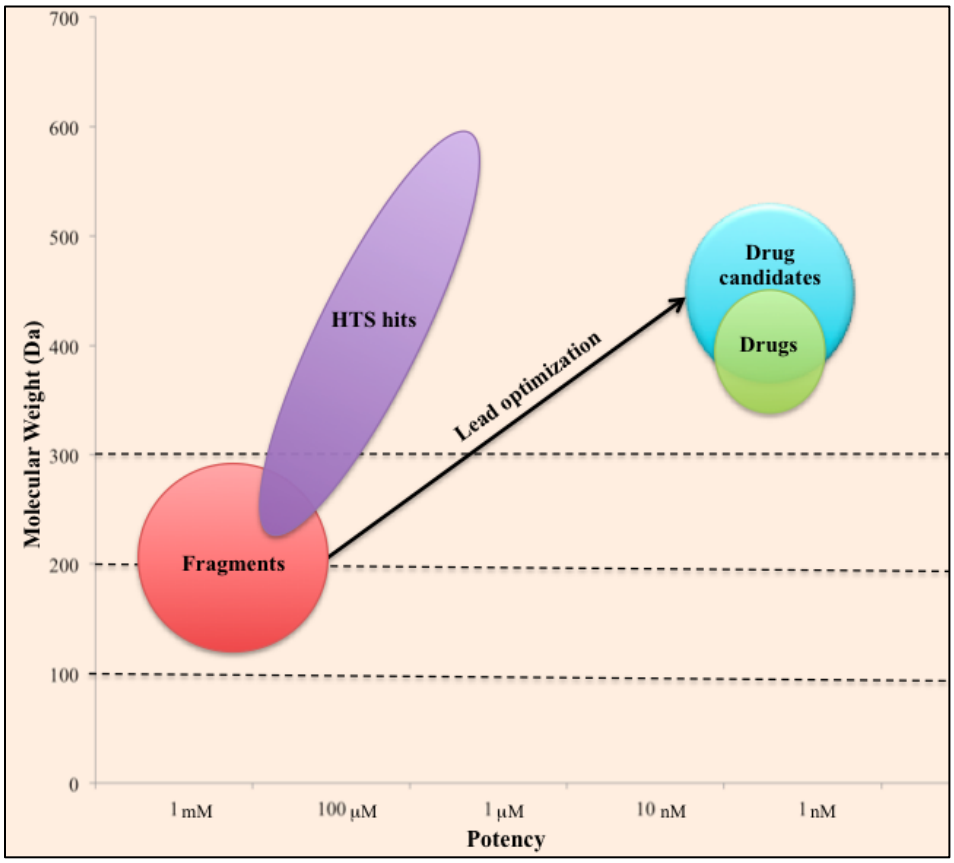


Figure 1: Lead development from HTS and FBDD as a function of molecular mass and potency (adapted from Rees et al., 2004).

Unlike conventional HTS, hits identified from fragment screening do not require deconstruction and can be efficiently enhanced for specificity and inhibitory activity (Figure 1). This is facilitated by the synthetically tractable nature of fragments for follow-

up chemistry. Incorporation of favorable physiochemical properties is readily possible when employing a fragment-based approach.

As shown in Figure 2, the development of a drug candidate from a FBDD platform consists of three key stages: 1.) design and curation of a fragment library, 2.) fragment screening using a wide range of biophysical techniques for the detection of fragment binding to the target protein, and 3.) fragment expansion, which involves iterative cycles of synthesis guided by a combination of bioactivity, structural, and *in silico* data. Each of these stages will be described in detail in the following sections.

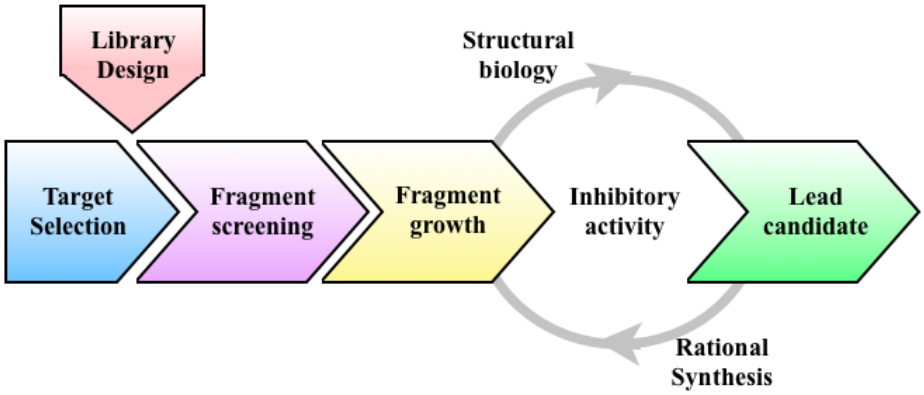


Figure 2: Fragment-based drug discovery pipeline.

1.2 Historical Perspective on Fragment-Based Drug Discovery

Fragment-based drug discovery was first conceptualized by Jencks in 1981, where it was suggested that the binding energy could be efficiently gained by connecting small, weak binders into larger molecules (Jencks, 1981). This approach was first applied four years later for the discovery of inhibitors targeting β-hydroxy- β-methylglutaryl coenzyme A reductase (Nakamura and Abeles, 1985). However, it did not gain popularity until Fesik and colleagues successfully incorporated fragment screening in their "SAR by NMR" approach (Shuker et al., 1996).

The fragment-based approach gained further momentum in 2001 after a model was used to rationalize the low success rate of HTS (Hann et al., 2001). It was hypothesized that the probability for a compound being a hit is dependent upon the complexity of the compound. Thus, if the compound is too small, the hits will not be easily detected during screening. Likewise, a low hit rate would be expected if compounds were too complex in nature. For a complex molecule to bind tightly to a target protein and form promising interactions, the molecule has to be positioned in a manner such that all of the molecular interactions formed contribute positively to the overall binding energy of the molecule (Hann et al., 2001). Thus, screening with larger molecules requires highly sensitive assays capable of detecting binding at low μM to nM concentrations and introducing chemical modifications may be challenging due to complicated synthetic chemistry and changes may alter the binding mode of the original hit, which, in turn, can dramatically reduce compound potency.

1.3 Ligand Efficiency

Owing to their small size of fragment hits, their binding affinities are generally weaker (typically mM) than those of large molecule hits. However, when the binding energy is normalized to the number of non-hydrogen atoms, fragments often bind more efficiently than larger molecules (Figure 3). This concept of energy normalization per atom was first introduced in 1984. A statistical analysis of a limited set of drugs was conducted to determine the theoretical binding potential based on the functional groups present (Andrews et al., 1984). Later, Kuntz et al. applied this concept to investigate the contribution of individual atoms to the binding energy (Kuntz et al., 1999). This led to the development of ligand efficiency (LE), a useful metric to evaluate the quality of

compound binding based on the atomic contribution to the overall binding energy of a molecule (Hopkins et al., 2004).

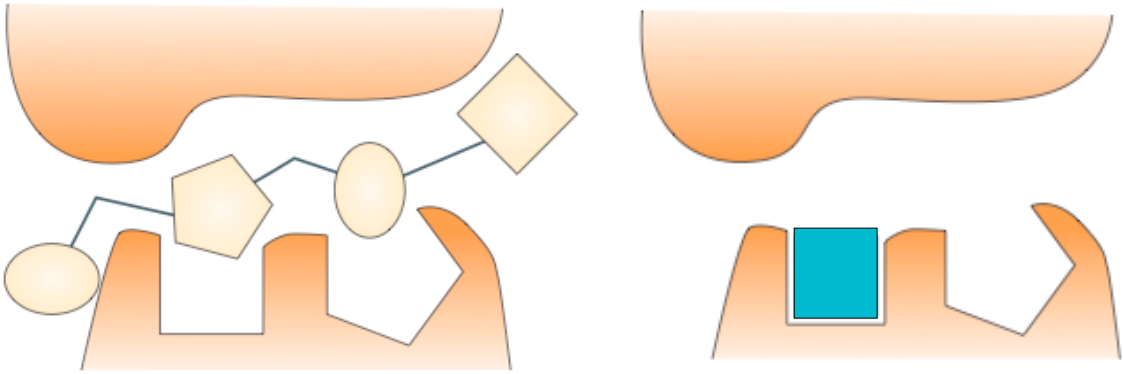


Figure 3: Schematic diagram depicting the efficiency of hits from conventional HTS (left) and FBDD (right) (figure adapted from Rees et al., 2004).

Ligand efficiency (in kcal/mol•NHA) is calculated by Equation 1, in which **Δ**G° is the standard free energy of binding, R is the universal gas constant, T is the absolute temperature in Kelvin (K), K_D (M) is the dissociation constant, and *NHA* is the number of non-hydrogen atoms. Ligand efficiency enables comparisons between different chemical series and scaffolds by normalizing the potency of a lead with respect to molecular weight. It is a commonly used metric in the fragment-based approach to guide fragment-to-lead development such that each non-hydrogen atom added during development contributes favorably to the binding energy.

$$LE = \frac{\Delta G^\circ}{NHA} = \frac{-RT \ln K_D}{NHA} \qquad \textbf{Equation 1}$$

1.4 Fragment Library Design

The success of the FBDD approach heavily relies on the design of the fragment library to be screened. The design of good screening library was fueled by the findings of Teague et al. (1999). Here, analysis of the data from high-throughput screens conducted at AstraZeneca revealed that polar, low molecular weight compounds were attractive for

lead development. Lead optimization frequently led to larger, more complex compounds, thus screening programs should focus on "lead-like" or fragment compounds rather than drug-like compounds (Teague et al., 1999; Oprea et al., 2001).

As with any screening approach, the quality of hits identified during the course of a screening campaign is dependent on the quality of the library. Thus, in case of fragment screening, careful library construction is required to ensure that a large percent of chemical space is covered in an efficient manner. A good fragment-screening library consists of compounds that are low in complexity and cover a large chemical space by ensuring chemotype diversity. Thus, library design continues to be a crucial yet challenging aspect of the fragment-based approach.

1.4.1 Fragment Selection and Diversity

The diversity of a library is dependent on several factors including fragment source, library purpose, and limitation of screening technique. Sources of fragments can range from known drug or biologically active molecules, natural products, or synthesis of novel chemical scaffolds. Fragment selection can be used to generate a general screening library or be biased for a particular target to generate a more focused library (Howard et al., 2006; Bodoor et al., 2009). The latter strategy may be efficient in finding hits, however the novelty of the chemical scaffold and thereby patentability will be reduced (Stockman and Dalvit, 2002). Construction of a general or a multi-target library heavily relies of maximizing chemical diversity to ensure the hits identified from any given screen will represent a range of chemotypes. In this case, design relies on *in silico* molecular and 3D shape descriptors to ensure each chemotype is equally represented within this library.

1.4.2 Complexity

Lower complexity is favored for screening as it significantly reduces the chemical space and increases the probability of a binding event. However, this also significantly reduces the potency of the identified hit molecule (Hann et al., 2001; Leach et al., 2006; Leach and Hann, 2011). Additionally, the probability of multiple binding events increases for compounds that have extremely low potency, thus making it difficult to detect an unambiguous binding mode for a fragment and engineering compound specificity.

To limit complexity, Astex Pharmaceuticals proposed the Rule of Three (Ro3), which recommends molecular mass < 300 Da, ≤ three hydrogen bond donors, ≤ three hydrogen bond acceptors, and calculated logP (clogP) <3, rotatable bonds ≤ 3, and a polar surface area ≤ 60 $Å^2$ (Congreve et al., 2003; Congreve et al., 2008). Design of initial fragment libraries by groups in academia and industry normally adhere to these guidelines. However, recently the Rule of Three has attracted some criticism within the field. Koster et al. found only four out of the eleven hits identified satisfied the Ro3 during a screening campaign against endothiapepsin with a fragment library that was not designed based on the Ro3 (Koster et al., 2011).

Recently, the average molecular mass of fragments has dropped to ~200 Da. To minimize reorientation of the bound fragment during lead development, a molecular weight of 150 Da has been identified as the lower limit for complexity (Babaoglu et al., 2006). Fragments with molecular weight less than 150 Da often bind only at high concentrations and at multiple sites and orientations (Siegal et al., 2007). The multiple solvent crystallographic structure (MSCS) method takes advantages of this property to identify promising druggable hot spots in the target protein (Mattos and Ringe, 1996;

English et al., 1999). Use of different solvents allows one to explore the plasticity and surface hydration of a particular site. In addition, the bound solvent molecules can provide experimental positions for functional groups that can be incorporated into ligand design. A molecular modeling approach can be used in which the binding of individual solvent molecules is optimized and subsequently linked to form a ligand with high binding affinity and specificity (Mattos et al., 2006). While this approach is theoretically viable, lead development using MSCS has met with limited success due to difficulty in reliably predicting optimal binding modes for linked molecules.

.1.4.3 Fragment Solubility and Reactivity

Since fragments bind to a protein target with very low affinity, screening is conducted at higher concentrations and requires fragments to be highly soluble. Additionally, fragments should be synthetically tractable for follow-up chemistry and avoid highly reactive, unstable, or toxic scaffolds, such as alkylating or acylating groups (Rishton, 1997; Rishton, 2003; Siegal, et al., 2007). Baell and Holloway (2010) have identified several substructures, dubbed pan-assay interference compounds (PAINS), which were false positive in a high-throughput screen. A common feature in the identified substructures was the presence of a Michael acceptor, such as rhodanines. These molecules act as soft electrophiles capable of reacting with the nucleophilic residues of the target either reversibly or irreversibly. Subsequently, their propensity to form covalent bonds makes them highly unattractive for drug discovery purposes.

1.4.4 Technical Considerations

Design of fragment libraries has been found to be challenging and to date no clear fragment library has established as a gold standard in the field. Library design often

requires maintaining a balance between high diversity, low complexity, and technical considerations. The screening technique is yet another variable that plays an important role in the design process. Library size is often restricted by the throughput of a screening technique whereas the minimal complexity of fragments is dependent upon the sensitivity of the said screening technique. Cross-reactivity between fragments also needs to be taken into consideration for techniques that use cocktails of more than one fragment for high-throughput screening.

Numerous commercially available fragment libraries take these factors into consideration and are readily used by both academics and big pharma. Although inconvenient, rigorous quality control to routinely validate compound identity and monitor compound purity and stability is required (Davis and Erlanson, 2013). Great amounts of resources are often wasted simply due to mislabeled or improperly identified fragments. For example, more than a dozen vendors were found to be selling the wrong isomer of bosutinib, a clinical stage kinase inhibitor (Halford, 2012; Davis and Erlanson, 2013). Additionally, several instances of false positives have been reported due to the presence of "silent" impurities from organic synthesis. For instance, several studies have reported high number of false positives due to residual gadolinium (Baell, 2013) as well as silver (Ward et al., 2012) and zinc (Hermann, 2013) ions.

Compound degradation is an important factor to consider when designing and maintaining a fragment library since fragments are frequently stored in dimethylsulfoxide (DMSO), a mild oxidant, for long periods of time (Kozikowski, 2003). For a test set of compounds dissolved in DMSO, substantially more compound degradation was observed after multiple rounds of freeze-thaw compared to storage at room temperature

(Oldenburg et al., 2005; Davis and Erlanson, 2013). Additionally, the highly hygroscopic nature of DMSO often causes water to be present in the compound stock solution. This can potentially be problematic with respect to fragment stability, solubility, and concentration. Alternative solvents, such as methanol or water, can be used to avoid these problems, however fragment solubility can limit fragment selection.

1.5 Biophysical Screening Methods

The low binding affinities of fragment hits allow for biophysical techniques to be used for screening. By now, an array of methods has been developed to rapidly identify weakly binding fragments against a target protein using biophysical approaches. Fragment libraries can be also screened for biological activity, however the expected lack of potency commonly necessitates the use of an *in vitro* enzymatic assay. Thus, biophysical methods are commonly used for initial identification of fragment hits, which can then be characterized by *in vitro* enzymatic assays for potential biological activity.

Selection of the biophysical method used for screening is highly dependent upon economical factors such as high throughput and low consumption of both protein and fragments. Since each methodology possesses its own advantages and pitfalls, additional factors such as the nature of the protein and fragment library also play a role in the selection process. For instance, fragments that are prone to aggregation are not suitable for nuclear magnetic resonance (NMR) or surface plasmon resonance (SPR). Likewise, protein size, monomeric state, and stability in necessary reagents for accurate detection of fragment binding can also affect the selection of the biophysical method used for screening. Several complementary screening techniques are used to ensure more reliable results from a fragment screening campaign.

1.5.1 Nuclear Magnetic Resonance (NMR) Spectroscopy

Several NMR-based methodologies have been found to be suitable for fragment screening. These methods are often categorized into either ligand-based or protein-based approaches. SAR by NMR, the first reported screening method for fragment screening, is a protein-based method that uses changes in the ^{15}N-labeled protein to identify the location and binding of low affinity fragments to the FK506 binding protein (Shuker et al., 1996). Here, fragment mixtures or "cocktails" of up to 30 compounds at a concentration of 400 μM are screened against an isotopically labeled protein target (Siegal et al., 2007). This approach is advantageous since it provides structural information about the ligand-binding site and has very low false positive rate. Although highly informative, protein based methods often are very resource intensive, requiring a fairly large amount of soluble, labeled protein.

Ligand-based NMR methods, on the hand, allow for rapid hit identification using relatively small amounts of protein. These methods, such as saturation transfer difference (STD), water-ligand observed via gradient spectroscopy (WaterLOGSY), and Carr-Purcell-Meiboom-Gill (CPMG), use proton NMR for the detection of fragment binding (Stockman et al., 2002). CPMG experiments determine fragment binding is associated with a reduction in the intensity of the proton signal. Differences in signal intensity for ligand (STD) or bulk water (WaterLOGSY) resonances between the bound and unbound states can also be used to identify fragment binding (Mayer et al., 1999). Use of ^{1}H NMR allows for rapid screening whereas high throughput is achieved by using fragment cocktails. Despite such advantages, tight binding fragments are often found to be false negatives since the rate of dissociation is not large enough to distinguish between the

bound and unbound states. Reliability of the results greatly depends on the solubility and aggregation of the target protein and fragments. Characterization of the binding pocket is not readily possible and requires the use of either protein-based methods or a displacer ligand with a known binding mode. In the latter case, compound binding is characterized by screening in the presence and absence of a known ligand (Brough et al., 2009).

Target immobilized NMR screening (TINS) is an alternative ligand-based approach that reduces protein consumption by immobilizing target protein onto a solid-state support (agarose or sepharose). Compound binding is detected by comparing 1D NMR spectra of protein in the presence and absence of fragment. TINS allows for efficient screening of fragment libraries, exhibits very low false positive rates, and is capable of detecting ligand binding with K_D from 60 to 5000 μM (Vanwetswinkel, 2005). It is important to note, however, unstable variants of the target protein, such as aggregates, can be formed by target immobilization and subsequently screened (Siegal et al., 2007).

Fluorine-19 (^{19}F) NMR is another ligand-based method that is slowly gaining popularity as a screening tool. Use of ^{19}F NMR is advantageous since ^{19}F is 100% abundant in nature. Compared to ^{1}H, analysis of ^{19}F spectra is much simpler and allows for automated analysis due to the absence of peak overlaps. Due to the large chemical shift range of ^{19}F, a larger difference in line width is observed for fluorinated molecules upon binding to the target protein (Dalvit, 2009). ^{19}F-based NMR experiments, such as fluorine chemical shift anisotropy and exchange for screening (FAXS), are highly sensitive with a large dynamic range (Vulpetti and Dalvit, 2012). This allows for screening large cocktails of fragments (up to 35 compounds per cocktail) at low concentrations. Fluorinated compounds also have favorable medicinal chemistry

properties; however, the usefulness of this technique is limited by the need for a fluorinated library and the low chemical diversity of the commercially available fluorinated compounds (Vulpetti et al., 2009; Jordan et al., 2012).

1.5.2 X-ray Crystallography

X-ray crystallography is another powerful biophysical method for fragment screening. The successful application of solvent mapping to identify hot spots on protein surfaces (Mattos and Ringe, 1996) first demonstrated the applicability of X-ray crystallography for the identification of small molecule binding in a directed search for many potential binding sites on a target protein. Crystallographic fragment screening was first used to identify ligand binding to triose phosphate isomerase from *Trypanosoma brucei* (Verlinde et al., 1997). Since then, this approach has been broadly accepted by pharmaceutical companies and developed into suitable platforms for fragment screening. Astex with its Pyramid approach is an example of such application (Williams, 2011).

Protein crystallization proves to be a major challenge for the successful implementation of X-ray crystallography for fragment screening. Ideally, crystallographic screening requires protein crystals to be highly reproducible and diffract X-rays to high resolution (better than 2.5 Å) to reduce ambiguities that may occur upon ligand binding. Crystal engineering and proteolytically stable protein constructs can be used to attain this for proteins that are either highly flexible or poorly diffracting. To prevent false negatives, the protein in the selected crystal form must be in a biologically relevant conformation, and binding sites must not be occluded by protein-protein crystal contacts or compounds used for crystallization and cryoprotection. Additionally, the protein crystallization condition should not promote chemical reactivity of the fragments used for screening.

For high-throughput purposes, crystallographic screening entails soaking of fragment mixtures or cocktails into preformed crystals. Fragment screening is typically conducted at high compound concentrations (10 mM and higher) to ensure full occupancy for low-affinity ligands. Cocktails are often designed with high shape diversity to allow for easy detection and deconvolution. Fragments are soaked in a solution based on the crystallization condition to ensure crystal stability and longevity during the soak. The solvent used for solubilizing fragments must also be to taken into consideration when optimizing the soaking solution as it may affect crystal stability. For instance, most protein crystals can only tolerate up to 10% (v/v) of DMSO. Soaking at higher concentrations of DMSO can either cause the crystal lattice to collapse or have a detrimental effect on diffraction resolution or quality. Incorporating DMSO into the protein crystallization condition can circumvent sensitivity of protein crystals to DMSO (Perryman et al., 2010). For the sake of ease, it is also advantageous to incorporate cryoprotectant into the fragment-soaking solution. Thus, optimization of the soaking conditions to ensure protein stability, fragment solubility, and technical ease is required to initiate a crystallographic fragment-screening campaign.

An alternative approach for crystallographic fragment screening is the co-crystallization of the protein in the presence of the fragment. This method is resource-intensive since each compound requires a separate crystallization experiment. Co-crystallization with fragment cocktails can be challenging due to issues with fragment solubility and long-term stability. Furthermore, ligand binding can alter the mobility of the protein thus requiring a change in the crystallization condition. This approach cannot identify weak fragment binders since high concentration of fragment can interfere with

crystallization (Davies and Tickle, 2012). Thus, co-crystallization is commonly used when fragment soaking either affects crystal stability or fails to identify any hits.

X-ray crystallography is a highly sensitive technique that allows for reliable detection of fragment binding with dissociation constants greater than 5 mM (Davies and Tickle, 2012). Simultaneously, it also provides structural information that enables rapid and efficient assessment of hits with respect to synthetic tractability for structure-based drug design. Despite such advantages, application of X-ray crystallography as the primary screening method has been under-appreciated and considered impractical due to its relatively low throughput and highly resource-intensive nature. However, recent advancements in crystal transport, robotic crystal mounting, powerful detectors that allow for rapid data collection, and automated data collection software have significantly improved the throughput and reduced the amount of labor associated with this approach. The high throughput combined with an efficient strategy for data collection and rapid hit identification (described in Chapter 2) makes X-ray crystallography an attractive approach for primary screening in that it not only identifies fragment binding but also provides a three-dimensional structure that can facilitate rapid structure-based optimization.

1.5.3 Surface Plasmon Resonance

Surface plasmon resonance (SPR) has emerged as a popular, label-free biophysical technique for fragment screening as it combines sensitivity with high throughput and low protein consumption. Detection of ligand binding is determined by changes in the refractive index of the solid support onto which the target protein is immobilized. SPR methods use at least 10- to 100-fold less protein than other fragment

screening techniques, with a typical SPR campaign using as little as 25 to 50 μg of protein per screening campaign (Neumann et al., 2007). Advances in biosensor technology have enhanced sensitivity of the technique to detect small molecules with molecular mass as low as 100 Da. Unlike other methods, SPR can be also used to characterize fragment binding with respect to kinetics and thermodynamics, specifically for fragments with slow binding off-rate. For weakly bound fragments with fast on- and off-rates, K_D can be derived from a binding model created by steady-state responses measured at different concentrations. This is often time consuming as it requires multiple injections.

The need for immobilization of protein onto a sensor chip can prove to be problematic for SPR experiments as it may occlude the binding site of interest. False positives are frequently identified due to non-specific binding events or binding to the surface matrix. Using a reference surface with either a standard reference protein or an inactive form of target protein can rectify this. Alternatively, competitive binding experiments could be used to determine location of fragment binding (Perspicace et al., 2007; Navratilova and Hopkins, 2010).

1.5.4 Isothermal Calorimetry

Isothermal calorimetry (ITC) is slowly gaining popularity as a screening tool. Fragment identification is characterized by a change in the either enthalpy or entropy upon ligand binding. Non-specific binding can be distinguished by extrapolating stoichiometry of ligand binding from an ITC experiment. Scott et al. (2009) reported that fragments binding via large enthalpic interaction can be readily developed into superior lead candidates. Thus, ITC can also be used to prioritize fragment hits for development.

ITC is frequently used as a secondary rather than a primary screening technique to validate fragment binding. This is predominantly due to a low sensitivity towards weak binders, high protein consumption, and low throughput of ITC in comparison to other screening tools (Ladbury, 2010; Torres et al., 2010).

1.5.5 Mass Spectrometry

Mass spectrometry (MS) has been successfully used to identify hits from fragment-screening campaigns (Maple et al., 2012; Swayze, 2002). Fragment binding is determined by an increase in molecular weight of a protein-ligand complex. Since MS analysis is conducted in a vacuum under high electric field, careful optimization is required to keep a protein in its native state and maintain stability of protein-ligand complexes. Additionally, MS is not compatible with detergents, thus small molecule aggregation can be extremely problematic during screening. Since MS analysis is conducted in the gas phase, where polar interactions are strengthened and hydrophobic interactions are weakened, hits identified from the screening may not necessarily be active in the solution phase (Hannah et al., 2010).

1.6 Fragment-to-Lead Development

Validated fragment hits serve as pharmacophore scaffolds for structural elaboration for improving potency through an iterative process of rational design and synthesis. Knowledge of the ligand-binding site and nearby pockets derived either from X-ray crystallography or protein-based NMR methods help facilitate this process. Currently, there are three major approaches—fragment linking, fragment merging, fragment evolution, and fragment self assembly—that can be employed to facilitate the development of potent leads from initial fragment hits.

1.6.1 Fragment Linking

Fragment linking utilizes a linker to join fragment hits bound to proximal sites with the target protein. Linking is favorable because the binding free energy of the linked molecule is less than the sum of the binding energies of the individual fragments. An ideal linker would allow the original fragments to maintain their preferred binding orientation while making favorable interaction with the protein. To minimize the entropic penalty of binding for a linked compound, reducing the flexibility of the linker is also recommended (Rees et al., 2004; Scott et al., 2012). In reality, these restrictions make fragment linking extremely challenging.

1.6.2 Fragment Merging

Fragment merging improves potency by incorporating structural components of overlapping fragment hits thereby modifying the core. Structural information of natural substrates and known ligands can also be used to identify important binding motifs and interactions that can enhance potency (Rees et al., 2004; Scott et al., 2012).

1.6.3 Fragment Evolution

Fragment evolution or growth involves the addition of functional groups to the original fragment hit to improve potency and binding. Here, the original hit acts as an "anchor" and often maintains its binding mode during the evolution process. Fragment evolution is relies heavily on structural information gained from either X-ray crystallography or NMR. *In silico* docking is also commonly used to guide chemical expansion of the initial fragment hit. Additionally, ligand efficiency can be used to ensure that new functionalities are highly efficient and beneficial (Hopkins et al., 2000; Rees et al., 2004; Scott et al., 2012).

1.6.4 Fragment Self-Assembly

The fragment self-assembly approach uses the protein to select for fragments that bind to nearby sites and chemically react to form a larger molecule (Rees et al., 2004). This approach does not require information regarding the binding affinities of the two fragments. Click chemistry is particularly useful in the case of fragment self-assembly. However, this restricts chemical diversity by allowing only compounds with suitable reactive functionalities to be screened. Careful library design and optimization of the reaction conditions is also necessary to avoid protein degradation for the successful implementation of this approach. Success is also highly contingent upon the binding of fragments in orientations that are proximal to each other and conducive for chemical reactivity.

Chapter 2: Fragment Screening Against HIV-1 Reverse Transcriptase Using X-ray Crystallography

Synopsis

HIV/AIDS continues to be a pressing public health concern worldwide. HIV-1 reverse transcriptase (RT), a key enzyme in viral replication, has been an attractive target for anti-retroviral therapy. While the highly effective HAART combinations in common use usually contain two and even three RT drugs, there is still a need for additional agents that are effective against drug-resistant variants. The highly flexible nature of RT as established by numerous structural and biochemical studies suggested the possibility of targeting new allosteric pockets for future drug therapy. Fragment screening by X-ray crystallography was conducted to probe the allosteric nature of RT. A high-throughput screening method using X-ray crystallography allowed for structural characterization of 16 allosteric binding sites, of which seven were of biological interest. Using a fluorescence-based enzymatic assay developed in-house, fragment binding to three of these seven sites, located near the polymerase and RNase H regions of HIV-1 RT, was found to be inhibitory. This chapter describes each of these sites with respect to structure and potential druggability. Crucial variables that were found to improve the efficiency of high throughput of X-ray crystallography as a screening technique are highlighted as well.

The work discussed in this chapter has been published in (* denotes co-first authors):

Bauman, J.*, Patel, D.*, Dharia, C., Fromer, M. W., Ahmed, S., Frenkel, Y., Vijayan, R. S. K., Eck, T. J., Ho, W. C., Das, K., Shatkin, A. J., Arnold, E. (2013) Detecting allosteric sites of HIV-1 reverse transcriptase by X-ray crystallographic fragment screening, *Journal of Medicinal Chemistry,* **56**: 2738-2746.

Having claimed more than 25 millions lives since its discovery in 1981, acquired immunodeficiency syndrome (AIDS) and its causative agent human immunodeficiency virus (HIV) continues to constitute a major global epidemic. In 2011, there were approximately 34 million people living with AIDS/HIV and 2.5 million newly reported cases (WHO HIV/AIDS, 2013) worldwide. Recent advances in the development of effective antiretroviral therapy have allowed for HIV-positive people to live longer and healthier lives. However, the HIV/AIDS epidemic continues to pose a public health challenge particularly in impoverished regions, such as sub-Sahara Africa where one out of 20 individuals is living with HIV/AIDS (WHO HIV/AIDS, 2013). Emerging resistance to current available drug therapy due to the highly dynamic and evolving nature of the virus poses yet another obstacle.

2.1 HIV Strains

Phylogenetic analysis identified two genetically distinct types of HIV within the human population: HIV-1, closely related to simian immunodeficiency virus (SIV) from chimpanzees, and HIV-2, which is related to SIV from sooty mangabeys (Hirsch et al., 1989). HIV-1 is currently the prevalent strain within the global population. Four major subtypes of HIV-1 have been identified: main (M), non-M and non-O (N), outlier (O), and pending the identification of further human cases (P) (Tebit et al., 2011). The majority of the HIV-1 infections reported worldwide belong to group M whereas groups O and N are limited to West and Central Africa, particularly Cameroon (Buonaguro et al., 2007). Group P was recently established after a virus isolated from a Cameroonian woman was found to be related to SIV in gorillas and distinct from HIV-1 viruses belonging to groups M, N, and O (Plantier et al., 2009).

Eight distinct groups of HIV-2 (A-H) have been identified, however only groups A and B are epidemic with group A being more prevalent. Although HIV-2 is typically confined to Western Africa, HIV-2 group A infections have been reported in India, South America, Europe, and the United States (Sharp et al., 1999). Other subtypes of HIV-2 have only been documented at an individual level.

2.2 HIV Virion

HIV is a lentivirus belonging to the *Retroviridae* family of viruses. Mature HIV-1 virions have a spherical shape with a diameter of approximately 100-120 nm (Barre-Sinoussi et al., 1983). HIV is enveloped by a lipid bilayer studded with roughly 72 copies of Env, which plays an important role during viral infection (Figure 4). Each Env consists of a cap made from three copies of glycoprotein 120 (gp 120) that is anchored by a stem consisting of three copies of glycoprotein 41 (gp 41) (Kwong et al., 2000).

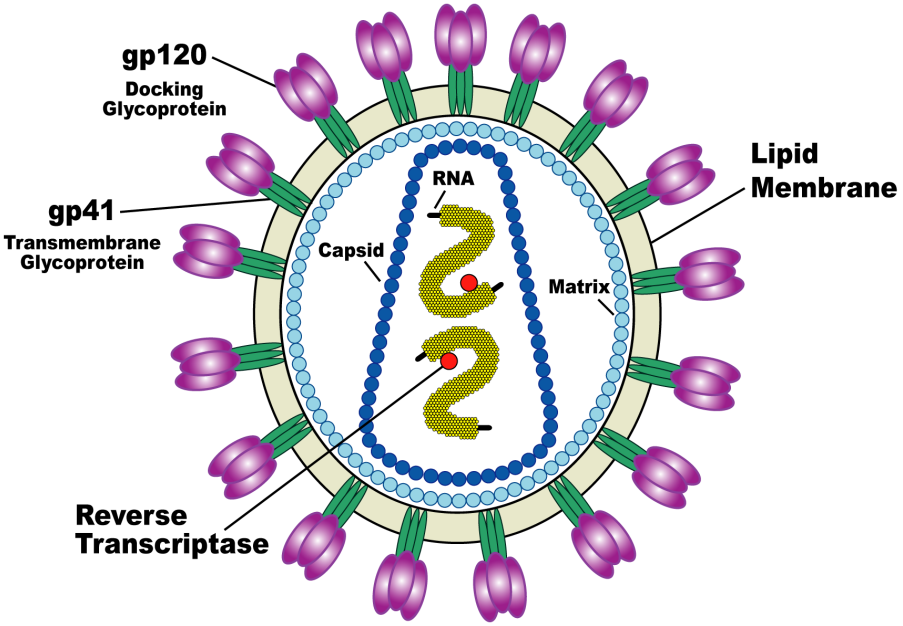


Figure 4: Anatomy of an HIV virion

An inner club-shaped capsid, consisting of roughly 2,000 copies of viral capsid protein, or p24, resides within the viral envelope. Enclosed in this capsid are two copies of

positive-sense, single-stranded RNA (~9.7 kbp in size) encoding the viral genome, which is packed with copies of reverse transcriptase and integrase, enzymes necessary for viral replication. An inner coat of matrix consisting of viral protein p17 separates the capsid and the envelope.

2.3 HIV Viral Genome

The HIV viral genome consists of nine open reading frames flanked by long terminal repeats at each end (Figure 5). Starting from the 5' end, *gag-pol* gene encodes the Gag and Pol polyproteins, which are proteolyzed into key structural and enzymatic proteins. Proteolysis of the Gag polyprotein yields the matrix, capsid, nucleocapsid, and p6 proteins, which are involved in virus assembly, budding, and release (Ganser-Pornillios et al., 2008). The *pol* gene encodes for a polyprotein that produces reverse transcriptase (RT), integrase (IN), and protease (PR) upon cleavage. Proteolysis of the Env gp160 polyprotein, which is encoded by the *env* gene at the 3' end of viral genome, produces gp120 and gp41 (Hunter et al, 1990; Freed et al., 1995; Checkley et al., 2011).

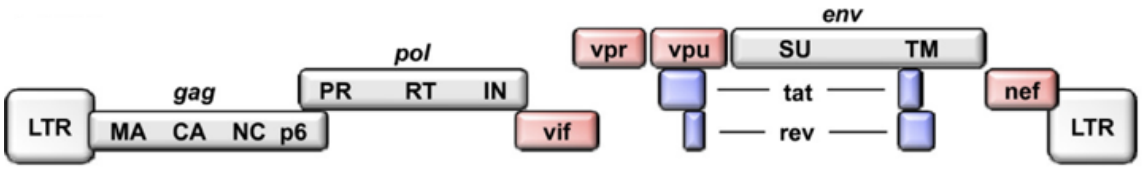


Figure 5: Schematic representation of the HIV-1 viral genome

HIV-1 also encodes for six additional proteins---Tat, Rev, Vif, Vpr, and Nef. Tat and Rev are regulatory proteins involved in viral transcription. Tat is a transcriptional activator that allows for full-length transcripts to be produced by promoting the elongation phase of HIV-1 transcription (Kao et al., 1987; Feinberg, et al., 1991). Rev is involved in shuttling of unspliced or singly-spliced viral RNA transcripts from the nucleus to the cytoplasm (Seelamgari et al., 2004). Nef, a 27 kDa myristoylated protein,

is involved in the down regulation of CD4 receptor (Aiken et al., 1994), perturbation of T-cell activation (Luria et al., 1991), and stimulation of HIV infectivity (Miller, 1994). Vpr, Vpu, and Vif are accessory proteins that have been found to be expendable for *in vitro* virus replication (Trono, 1995). Vpr is involved in the transport of the pre-integration complex (PIC) into the nucleus and is capable of inhibiting cell division (Malim and Emerman, 2008). Vpu enhances virion release and initiates ubiquitin-mediated degradation of CD4 (Bour et al., 2003). Vif prevents incorporation of APOBEC3G into virion by reducing the translation and enhancing the degradation of APOBEC3G (Stopak et al., 2003).

2.4 HIV Life Cycle

The HIV viral life cycle involves a series of complex interactions with the host cell (Figure 6). Viral replication is initiated by the interaction between HIV envelope, specifically Env, and the host cell surface receptors to allow for viral entry. Binding of gp120 to the CD4 cell receptor anchors the virus to the host cell and induces a conformational change within the V3 region of the HIV envelope. This allows for binding to the cellular co-receptors, specifically CCR5 and CXCR4 (Gomez and Hope, 2005), resulting in the fusion of the cellular membrane and viral envelope and subsequent entry of the viral capsid into the cell. Upon entry, uncoating of the capsid is initiated to release the viral genome into the cytoplasm. This is not a spontaneous process and requires assistance from several cellular proteins, such as cyclophilin A (Auewarakul et al., 2005; Aiken, 2006) and peptidyl-prolyl isomerase (Pin 1) (Misumi et al., 2010).

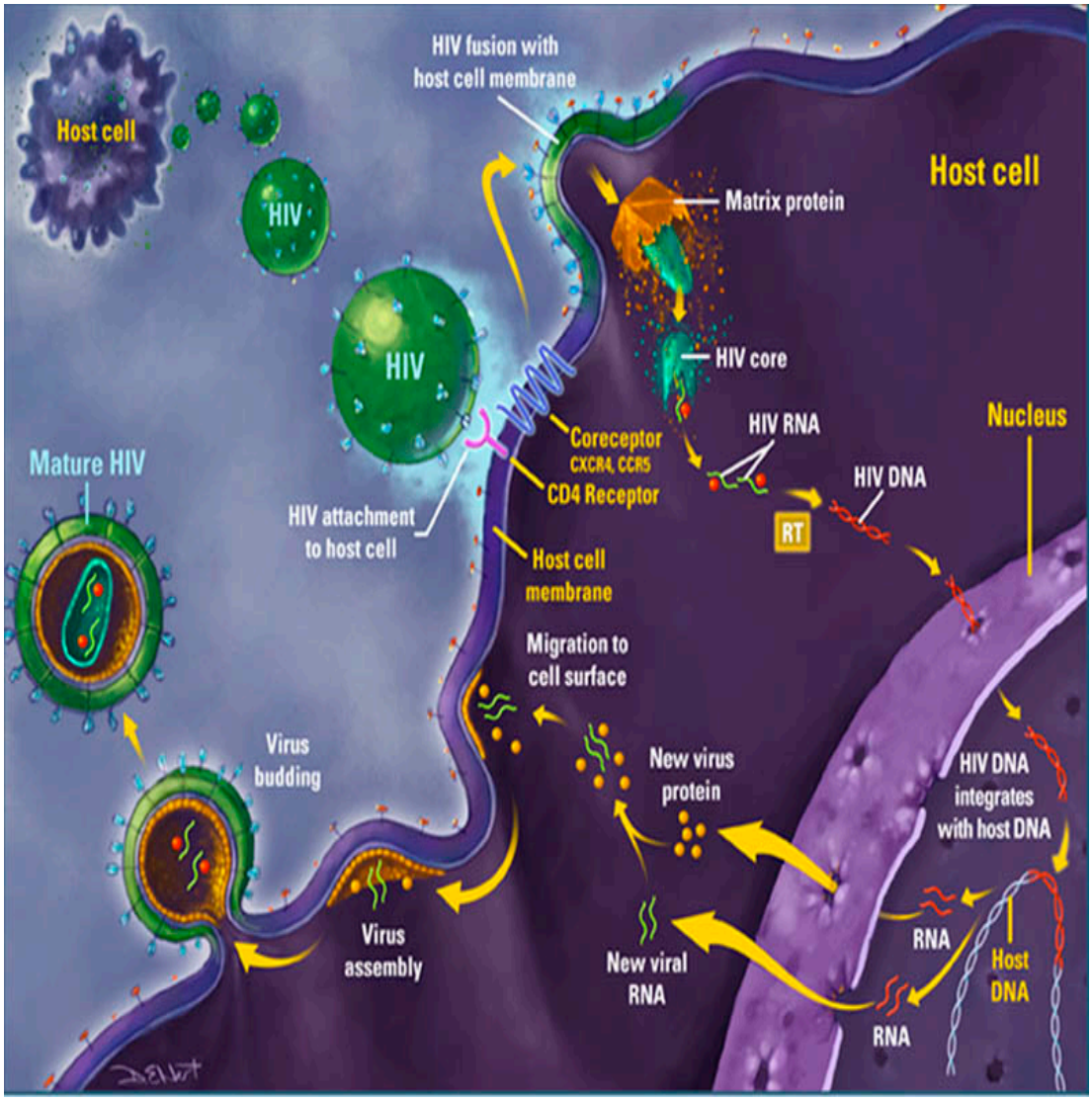


Figure 6: HIV viral life cycle (courtesy of HIV webstudy).

After the release of the viral core, HIV reverse transcriptase uses cellular nucleotides and $tRNA^{Lys3}$ to synthesis a double-stranded DNA based viral genome using the single-stranded HIV RNA released from the capsid as the template (Hu and Hughes, 2012). Upon completion of reverse transcription, a pre-integration complex, consisting of viral DNA, reverse transcriptase, integrase, matrix, Vpr, and cellular co-factors, is formed and subsequently transported into the nucleus for integration into the host DNA. The integrated virus DNA is now referred to as a provirus and can use cellular transcription

machinery for the synthesis of structural, regulatory, and accessory proteins necessary for viral production. Transcription of proviral DNA into viral mRNA is controlled by several factors, including HIV Tat and Rev (Nekhai and Jeang, 2006). Viral mRNA is transported into the cytoplasm where it is translate into viral proteins by cellular ribosomes. The Env polyprotein is translated by the ribosomes in the endoplasmic reticulum whereas Gag and Gag-Pol polyproteins are translated by ribosomes in the cytoplasm (Bell and Lever, 2013). Polyproteins are then transported to the cellular membrane and proteolytically cleaved by cellular and viral proteases to form mature, functional viral proteins for assembly and viral budding.

2.5 HIV-1 Reverse Transcriptase

HIV-1 reverse transcriptase (RT) is vital for viral replication (Figure 7). RT serves as an RNA-dependent DNA polymerase for the synthesis of single-stranded DNA from the viral RNA template. It also has DNA-dependent DNA polymerase activity to complete the synthesis of the newly transcribed DNA and RNase H activity to degrade the RNA template.

RT itself is a 117 kDa heterodimer consisting of p66 (66 kDa) and p51 (51 kDa) subunits (Figure 7). The p51 subunit is the product of the proteolytic cleavage of the C-terminal RNase H domain from the p66 subunit (di Marzo Veronese et al., 1986). Despite strong sequence similarity, crystal structures of RT revealed that the subunits to have very different conformations. The N-terminal region of p66 resembles an open right hand containing three subdomains, aptly referred to as the fingers (residues 1-85 and 118-155), palm (residues 86-117 and 156-237), and thumb (residues 238-318) subdomains (Kohlstaedt et al., 1992). The palm subdomain contains the polymerase active site with

the three catalytic aspartic acid residues (Asp110, Asp185, and Asp186) and the characteristic YMDD motif. The connection (residue 319-426) subdomain separates the thumb and the C-terminal RNase H (residues 427-560) domains. The p66 subunit is the home to both the polymerase and RNase H activities of RT. The p51 subunit lacks polymerase activity as the catalytic residues are buried. Instead, it stabilizes the properly folded p66 subunit to ensure proper enzymatic function.

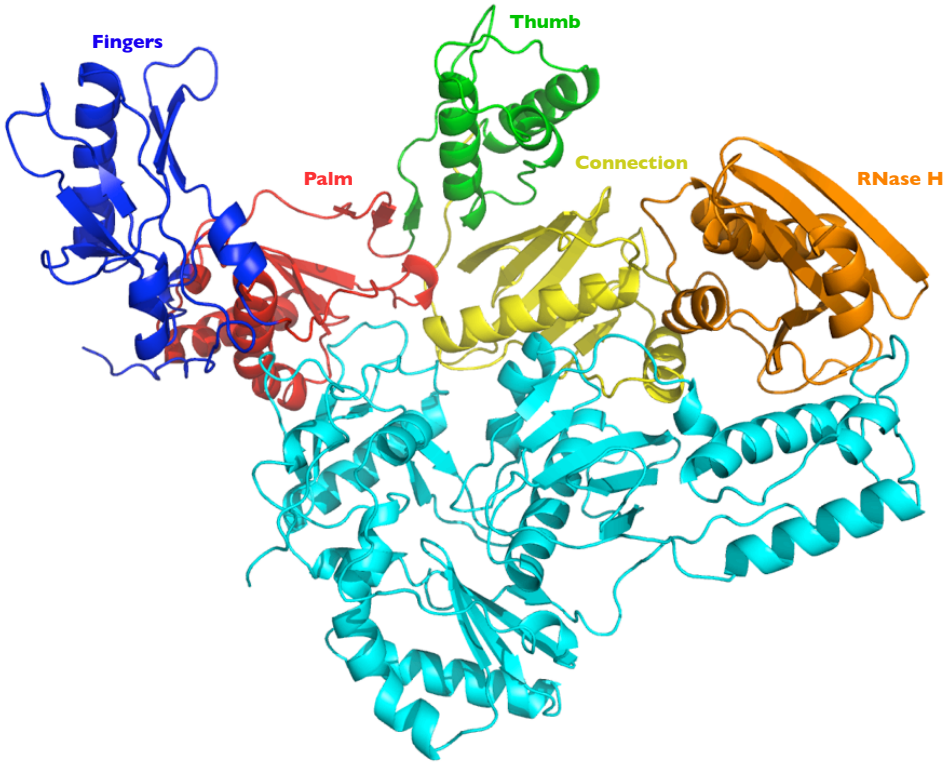


Figure 7: Cartoon representation highlighting structural features of HIV-1 reverse transcriptase. The p66 subdomains are color-coded fingers (blue), palm (red), thumb (green), connection (yellow), RNase H (orange), and the p51 subunit (cyan). Crystal structure figures were made with MacPyMol (Schrödinger, New York, NY) unless otherwise noted.

2.6 Antiviral Therapy

Currently, a total of 26 FDA-approved drugs inhibit viral replication by preventing viral entry or impeding the enzymatic activities of key viral proteins, specifically reverse

transcriptase, integrase, and protease (FDA antiretroviral drugs used in treatment of HIV infection, 2013) (Table 1). Since 1996, antiviral regiments rely on combination therapy,

Drug Class	Generic Name (Acronym)	Brand Name
Fusion Inhibitor	Enfuvirtide (T-20)	Fuzeon
CCR5 Antagonist	Marviroc (MVC)	Selzentry
Reverse Transcriptase (RT) Inhibitors		
Nucleloside RT Inhibitors	Abacavir (ABC)	Ziagen
	Didanosine (ddI)	Videx/
	Emtricitabine (FTC)	Emtriva
	Lamivudine (3TC)	Epivir
	Stavudine (d4T)	Zerit
	Tenofovir DF (TDF)	Viread
	Ziodvudine (ZDV, AZT)	Retrovir
Non-nucleoside RT Inhibitors	Delavirdine (DLV)	Rescriptor
	Efavirenz (EFV)	Sustiva
	Etravirine (ETR)	Intelence
	Nevirapine (NVP)	Viramune
	Rilpivirine (RPV)	Edurant
Protease Inhibitors	Atazanavir (ATV)	Reyataz
	Darunavir (DRV)	Prezista
	Fosamprenavir (FPV)	Lexiva
	Indinavir (IDV)	Crixivan
	Nelfinavir (NFV)	Viracept
	Ritonavir (RTV)	Norvir
	Saquinavir (SQV)	Invirase
	Tipranavir (TPV)	Aptivus
Integrase Inhibitors	Raltegravir (RAL)	Isentress
	Dolutegravir (DTG)	Trivicay
	Elvitegravir (EVG)	Stribild

Table 1: Current FDA-approved inhibitors for treatment of HIV/AIDS.

referred to as highly active antiretroviral therapy (HAART), to treat HIV infection. The

goal of HAART is to obtain synergy between the inhibitors of different targets, lower individual drug dosages and thereby reduce the toxic side effects, and diminish the emergence of drug resistance (De Clerq, 2010). HAART usually relies on a combination of three or more RT and PR inhibitors but can be modified to incorporate other drugs including newly approved inhibitors of IN. Currently, there are 13 FDA approved anti-HIV drugs that inhibit the enzymatic activity of RT. These drugs can be classified as either--nucleoside reverse transcriptase inhibitors (NRTIs) or non-nucleoside reverse transcriptase inhibitors (NNRTIs).

2.6.1 Nucleoside Reverse Transcriptase Inhibitors

NRTIs became the first class of antiretroviral drugs to gain FDA approval for the treatment of HIV with the development of zidovudine (AZT) in 1987. Currently, there are eight NRTIs—zidovudine (AZT), stavudine (D4T), zalcitabine (ddC), didanosine (ddI), lamivudine (3TC), abacavir (ABC) and emtricitabine (FTC) and one nucleotide tenofovir (TDF)—that are available for treatment (Figure 8). NRTIs are structural analogs of dNTPs except they lack a 3'-OH. They inhibit the polymerase activity of RT by competing with the natural dNTP substrate for incorporation into the viral DNA. The lack of a 3'-OH prevents further elongation of viral DNA, thus terminating viral transcription (Goody et al., 1991; Balzarini et al., 1998).

NRTIs are delivered to the cell as nucleoside prodrugs and activation requires the addition of α, β, and γ phosphates by cellular kinases. Tenofovir requires the addition of only two phosphate groups as it is a phosphonate analog. Inhibitor potency of NRTIs is affected by the rate of activation and subsequent rate of incorporation (Cihlara and Rayb, 2010).

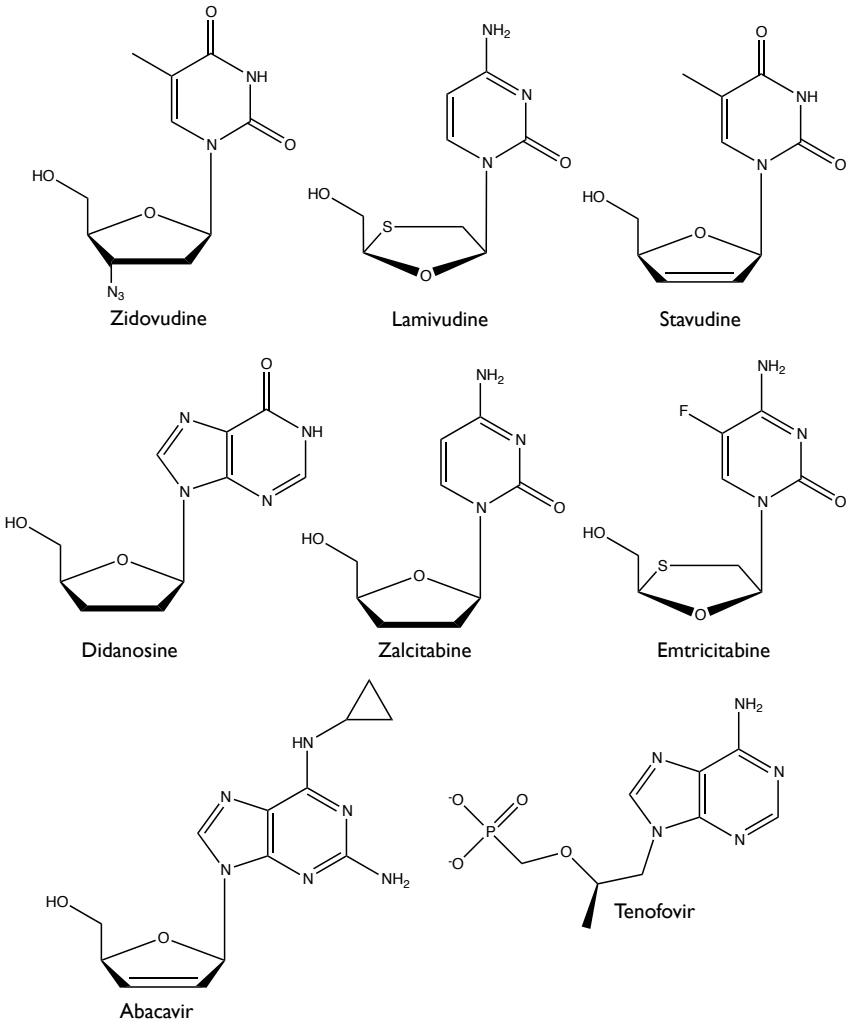


Figure 8: Chemical structures of NRTIs

2.6.2 Non-nucleoside Reverse Transcriptase Inhibitors

NNRTIs are allosteric inhibitors of RT (Balzarini, 2004; Jochmans et al., 2009) with high specificity for HIV-1 and significantly less adverse side effects compared to NRTIs. Unlike NRTIs, which are modified derivatives of dNTPs, NNRTIs are chemically diverse compounds that bind to an allosteric pocket, referred to as the NNRTI-binding pocket (NNIBP). NNRTI binding stabilizes the NNIBP, which is located ~10 Å away from the polymerase active site between the β6-10-9 and β12-β13-β14 sheets in the palm subdomain of the p66 subunit (Jacobo-Molina et al., 1993; Ding et al., 1995a, 1995b; Ren

et al., 1995; Esnouf et al., 1995). Formation of the NNIBP causes the catalytic aspartate residues to shift 2 Å and the side chains of Tyr181 and Tyr188 rotate away from the hydrophobic center of the pocket and towards the direction of the catalytic active site (Hsiou et al., 1996). The NNIBP also includes residues Leu100, Lys101, Lys103, Val106, Thr107, Val108, Val179, Tyr181, Tyr188, Val189, Gly190, Phe227, Trp229, Leu234, and Tyr318 of p66, and Glu138 of p51 (Nanni et al., 1993; Tantillo et al., 1994; de Bethune, 2010).

Currently, there are five FDA approved NNRTIs available for antiviral therapy (Figure 9). First generation NNRTIs--nevirapine and efavirenz--are an essential component of HAART therapy. However, low genetic barrier to drug resistance and adverse side effects led to the development of etravirine and rilpivirine, second generation NNRTIs, which maintain high efficacy and tolerability against wildtype and mutant strains.

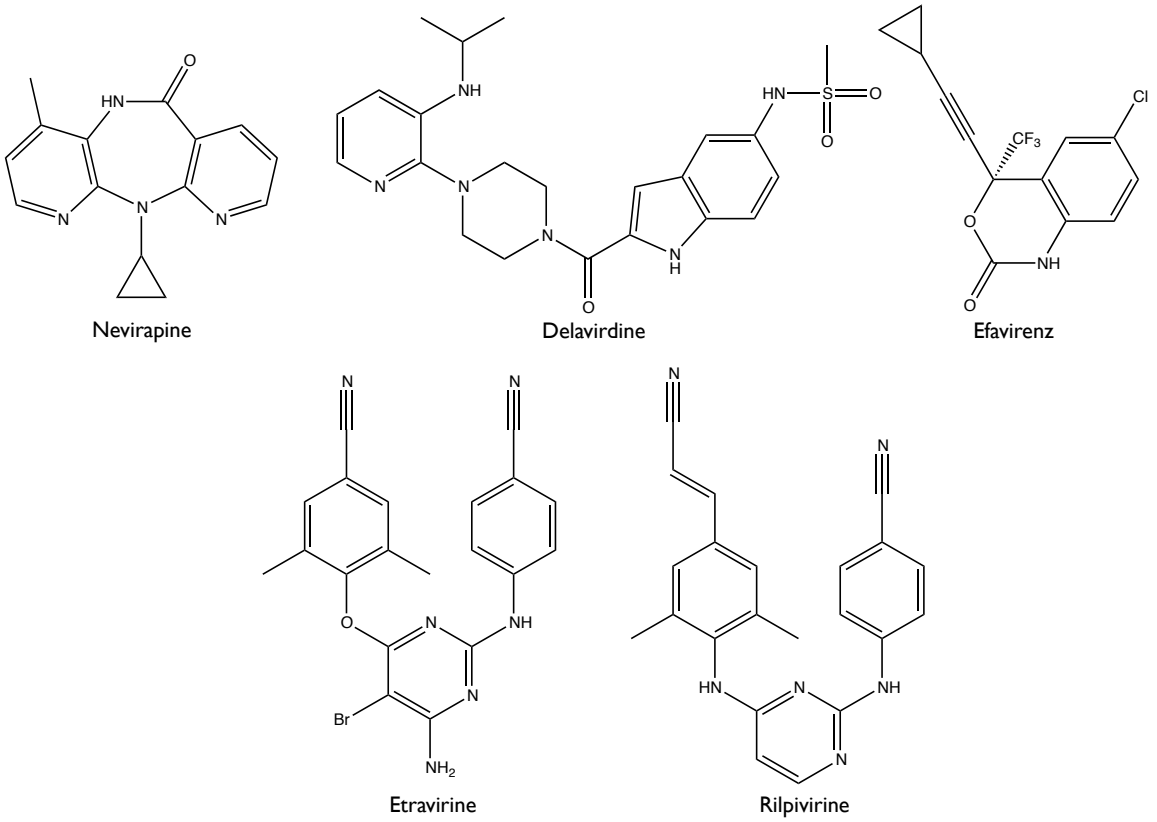


Figure 9: Five FDA-approved NNRTIs

Despite rapid emergence of drug resistance, nevirapine, the first FDA approved NNRTI, is used as part of HAART therapy due to its synergism with NRTIs (Havlir et al., 1995), low cost production, and good bioavailability. Delavirdine also selects rapidly for resistance and requires a relatively high and inconvenient dose regimen of 400 mg, 3 times a day to maintain reasonable efficacy (Scott and Perry, 2000). It also lacked 2nd generation potency and resistance reliancy. As a result, delavirdine is no longer clinically used. Efavirenz is the leading prescribed NNRTI and a principal component of HAART regimens. Significant decrease in the antiviral potency of efavirenz usually requires two or more mutations within the HIV-1 RT. In particular, efficacy of efavirenz is strongly compromised by the Lys103Asn mutation within RT (Vrouenraets et al., 2007; Sahlberg and Zhou, 2008).

The second-generation NNRTIs, etravirine and rilpivirine, are highly potent and retain efficacy against a broader spectrum of NNRTI drug resistant strains of HIV. Both etravirine (ETV, TMC125) and rilpivirine (RPV, TMC278) (Janssen et al., 2005) were discovered through optimization of diarylpyrimidine class of RT inhibitors (Janssen et al., 2005). The high level of potency against both wildtype and mutant HIV-1 RT variants observed for both TMC125 and TMC278 has been attributed to strategic flexibility of the molecules. This concept, first introduced by Das et al. in 2004, suggests that the inherent flexibility of the molecule ("jiggling and wiggling") can accommodate emerging mutations while maintaining potent inhibition of RT activity (Das et al., 2004 and Das et al., 2005). Subsequently, TMC125 and TMC278 are effective against commonly observed single NNRTI drug resistance mutations that line the NNRTI pocket, including Lys103Asn, Tyr181Cys, Gly190Ala, Leu100Ile, Val179Ile, and Tyr188Leu.

2.7 HIV-1 RT Dynamics

The highly dynamic nature of RT facilitates its ability to be a multifunctional enzyme. Structural comparison of the different conformational states of RT found the fingers and thumb subdomains to be highly flexible. Within the apo RT structure, the thumb and fingers subdomain adopt a "closed conformation". However, in the presence of nucleic acid, the thumb repositions and adopts an "open conformation" to allow for the formation of the nucleic acid binding cleft and to hold the nucleic acid in place (Jacobo-Molina et al., 1993; Ding et al., 1998). Domain movement is not limited to the thumb subdomain. Crystal structure of the ternary complex (RT-DNA-dNTP) revealed that the fingers subdomain adopt a closed conformation, whereas it remains open in the pre- and post-translocation complexes. Thus, structural comparison of the various RT complexes

revealed dynamic movement of the fingers and thumb subdomains from open to closed state is necessary for DNA polymerization.

The highly flexible nature of RT is not strictly limited to inter-domain movements; key biochemical studies have characterized the effect of protein dynamics of RT on its enzymatic activity. Liu et al. (2008) demonstrated that RT can spontaneously slide over long distances of RNA:DNA and DNA:DNA duplexes in order to easily target the primer terminus for polymerization, access multiple sites, and compensate for low processivity. This sliding is believed to be a thermally driven diffusion process and does not require energy from nucleotide hydrolysis (Liu et al., 2008). Enzymatic activity of RT is regulated by the dynamic equilibrium formed by the two different binding orientations of RT to nucleic acid, aptly termed "RNase H competent orientation" and "polymerase competent orientation" (Abbondanzieri et al., 2008). Single molecule FRET analysis indicate that enzyme dissociation is not necessary for this flipping to occur and the kinetics of flipping can be altered by nucleotides and inhibitors (Abbondanzieri et al., 2008). Nucleotides were found to decrease the flipping rate from the polymerase competent to the RNase H competent orientations whereas NNRTIs stabilizes the RNase H competent orientation. The ability to shuttle and flip requires a series of complex conformational changes to increase replication efficiency while maintaining enzymatic activity.

2.8 Aim of the Study

Emerging drug resistance to the currently available RT inhibitors presses the need for more efficacious treatment against wildtype and mutant HIV-1 strains. The highly dynamic nature of RT as established by both structural and biochemical studies suggests

the possible existence of additional allosteric pockets for drug design. Fragment screening by X-ray crystallography was used to probe the flexibility of RT for the identification of novel allosteric sites with promising potential for drug design. Crystallographic screening of 742 fragments revealed binding to 16 sites within RT of which seven were of great interest. Enzymatic assays revealed three of the seven sites to be inhibitory. The following sections describe the development of a high-throughput screening methodology using X-ray crystallography and the dual-activity monitor assay (JDAM). Seven sites of interest are also described in detail.

2.9 Development of HTS Screening Methodology

Using X-ray crystallography for as the primary screening method for hit identification is time-consuming and extremely laborious. Thus, a high-throughput screening methodology was developed to allow for rapid hit identification without wasting the available resources. Several key variables, described in the following sections, were found to have a tremendous impact on the success of an X-ray crystallography-based fragment-screening campaign.

2.9.1 Crystal Optimization

High-resolution crystals are necessary for fragment screening by X-ray crystallography. Thus, this approach would have been highly problematic for HIV-1 RT, which produced crystal forms that diffract to 2.5-3.0 Å. However, Bauman et al. successfully used crystal engineering to create two enzymatically active, non-drug resistant variants of RT, RT52A and RT69A, which produced crystals that diffracted X-rays to better than 2 Å resolution. The high-resolution diffraction of these constructs was maintained in the presence and absence of NNRTI bound to RT (Bauman et al., 2008).

Thus, these constructs were amenable for a fragment screening campaign using X-ray crystallography.

Initial screening with crystals of RT69A proved to be challenging since the majority of these crystals were destroyed during fragment soaking. Varying the soaking conditions by decreasing the fragment and DMSO concentrations did not improve crystal stability in the fragment solution. The poor stability of the unliganded RT69A crystals was attributed to fragment binding to the NNRTI pocket, which as previously discussed entails a large conformational changes within RT. To limit protein flexibility, crystals of rilpivirine in complex with RT52A were used for fragment screening instead. rilpivirine stabilized the open conformation of the NNRTI binding pocket and improved crystal stability during fragment soaking. These crystals are highly reproducible and diffracted X-rays better than 2 Å resolution. They were also extremely robust with 93% yielding a high quality X-ray diffraction after fragment soaking.

2.9.2 Library Design

A fragment library of 775 compounds comprising 500 compounds purchased from Maybridge (Cornwall, UK), 175 compounds purchased based on the published recommendations of Christophe Verlinde and Wim Hol (Verlinde et al., 2009), and an additional 100 compounds generously gifted by James Williamson (The Scripps Research Institute, La Jolla, unpublished). The 175 fragments comprising the Verlinde library were purchased from Maybridge, Sigma-Aldrich or Acros. The fragments were purchased with nominal purity of >99% and used without further purification. A new lot of highly pure fragment (>99% purity) from Sigma-Aldrich, Acros, or Tocris was purchased to validate fragment hits from the Williamson library.

To enable high throughput screening, the fragments were grouped into 143 cocktails with roughly four to eight compounds (100 mM each in d_6-DMSO). Fragment cocktails were designed to have maximum structural diversity to facilitate deconvolution in the electron density maps, cocktail solubility, and no predicted chemical reactivity among the constituents. Additional compounds were purchased throughout the screening process based on derivatization of fragment hits and *in silico* docking (using GlideXP from Schrödinger) to newly discovered binding pockets.

2.9.3 Optimization of Soaking Condition

Fragment screening by X-ray crystallography usually involves soaking crystals first with fragments then with a cryoprotectant. In this case, DMSO had the dual purpose of acting as the solvent for the fragments and a cryoprotectant for RT-rilpivirine crystals. This minimized crystal damage and reduced the amount of time necessary for soaking and freezing. Crystals were found to tolerate 20% (v/v) DMSO, which in turn allowed for soaking of fragment at concentration of 20 mM each. Ethylene glycol at a final concentration of 5% (v/v) was added to this solution as an additional cryoprotectant.

Fragment solubility was a major concern soaking at a high fragment concentration. Initial experiments revealed that several hydrophobic fragments bound to RT in the presence of L-arginine (Figure 10). It is speculated that the solubilization is due to the guanidinium group of L-arginine, which can form hydrogen bond and/or π-stacking interactions with fragments (Flocco et al., 1994).

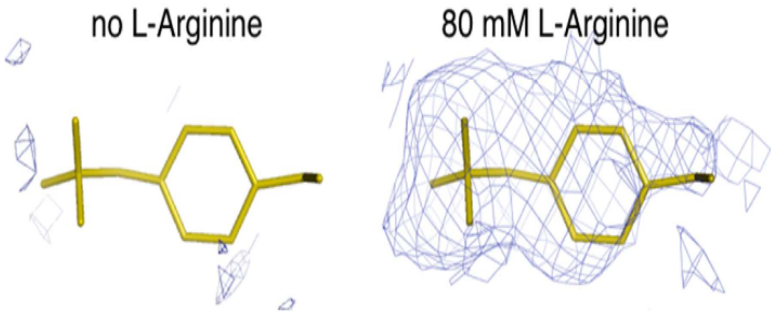


Figure 10: Importance of L-arginine. Electron density (calculated using $F_{obs(fragment\ soaked)}$-$F_{obs(DMSO\ blank)}$ coefficients) at the Knuckles pocket after crystals of RT-rilpivirine were soaked in solutions containing compound **2**, with and without 80 mM L-arginine present, respectively.

No noticeable difference in the strength of fragment binding (as assessed by the quality of the electron density) was rilpivirine observed upon varying the time crystal spent in the soaking solution. An ideal soaking time of one to two minutes was determined from initial soaking experiments that explored a time range from 30 seconds to 10 minutes.

2.9.4 High-throughput Data Collection and Processing

A crystal structure of RT52A-rilpivirine soaked in 20% (v/v) DMSO without the presence of any fragment was determined at a resolution of 1.8 Å and revealed 16 binding sites for DMSO within RT (Figure 11). This structure was used as a blank reference to avoid misinterpretation of solvent binding as fragment binding. The "DMSO blank" coordinates and scattering factors were then used to create $F_{obs(fragment\ soaked)}$-$F_{obs(DMSO\ blank)}$ (F_o-F_o) difference maps to identify fragment binding.

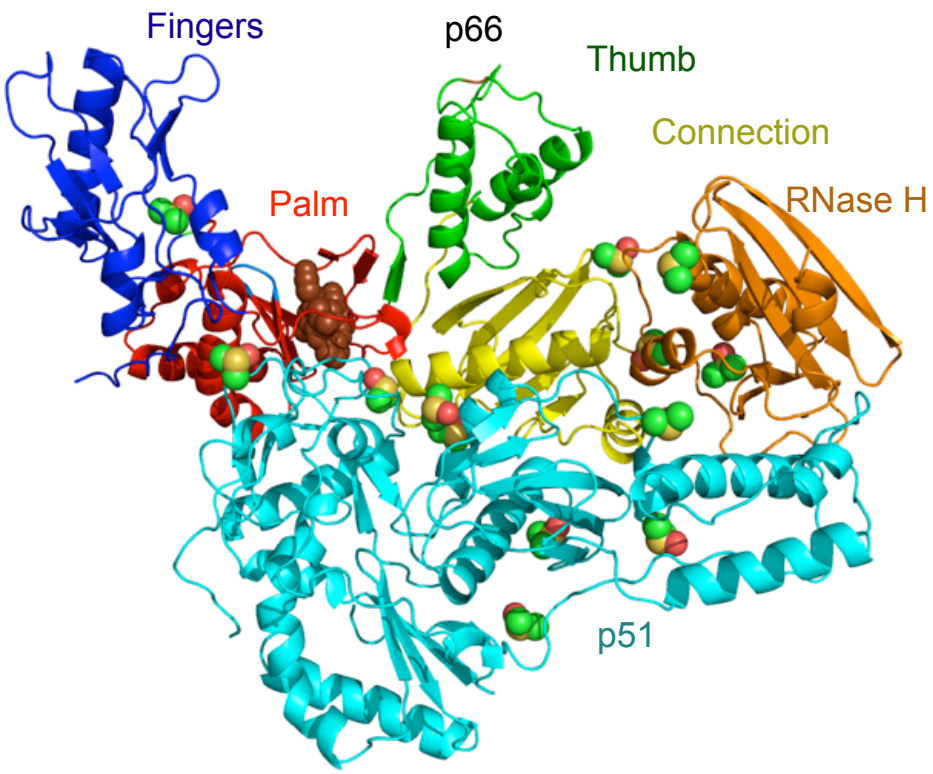


Figure 11: Color-coded cartoon of the RT-rilpivirine complex with 20% d_6-DMSO present. Rilpivirine (brown space filling) is bound at the NNRTI-binding pocket. The p66 subdomains are color-coded fingers (green), palm (red), thumb (blue), connection (magenta), RNase H (orange), and the p51 subunit (cyan). Ordered waters are shown as blue dots and d_6-DMSO molecules are shown as green, yellow, and red spheres.

In order to maximize data collection and minimize radiation damage, a quick initial data collection pass on each crystal was performed. Limiting exposure time and increasing oscillation range reduced the duration of dataset collection to less than five minutes while maintaining diffraction resolution between 2.0 to 2.3 Å (Figure 12). Once data collection was completed, the dataset was then immediately processed and a $F_{o(fragment)}$-$F_{o(DMSO\text{-}blank)}$ difference map was generated and visually inspected to identify any change in the electron density due to either fragment binding and/or repositioning of protein residues or the solvent. Crystals for which fragment binding was detected were subjected to a high-resolution data collection pass to improve the resolution of the dataset to 1.8 to 2.1 Å. The high-resolution dataset was then merged with the quick-pass dataset

to improve the density for the bound fragment for further refinement. Fragment hits identified from cocktail soaking were subsequently verified by individual fragment soaking.

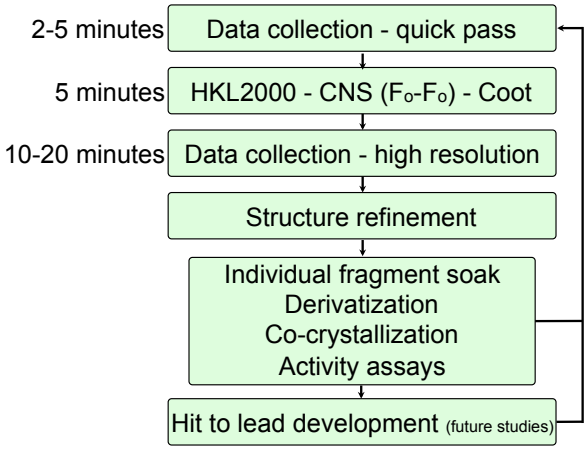


Figure 12: Flowchart for fragment screening by X-ray crystallography with approximate completion times indicated.

2.9.5 Dual Activity Monitor Assay (JDAM)

The dual activity monitor assay, JDAM, was developed to elucidate if fragment binding inhibited either the polymerase or RNase H activities of RT (Figure 13). A 142-nucleotide, purified T7 polymerase runoff transcript was annealed to a DNA primer. In the presence of dNTPs, RT extends the DNA primer, complementing the RNA template using the RNA-dependent DNA polymerase activity, and then degrades the RNA using its RNase H activity. A DNA "molecular beacon" type probe, *i.e.*, a single-stranded oligonucleotide with a stem-loop structure, was used to detect activity. The probe consists of a 25-nucleotide loop that is complementary to the newly synthesized DNA strand and a five base-pair stem that holds a fluorophore and quencher proximal to each other. Upon synthesis of the DNA strand and hydrolysis of the template RNA, the stem-loop structure of the probe collapses when it binds to the complementary DNA. This subsequently causes the fluorophore and quencher to separate, resulting in fluorescence. A decrease in

the amount of fluorescence in the presence of a fragment relative to a non-inhibited reaction is interpreted as inhibition.

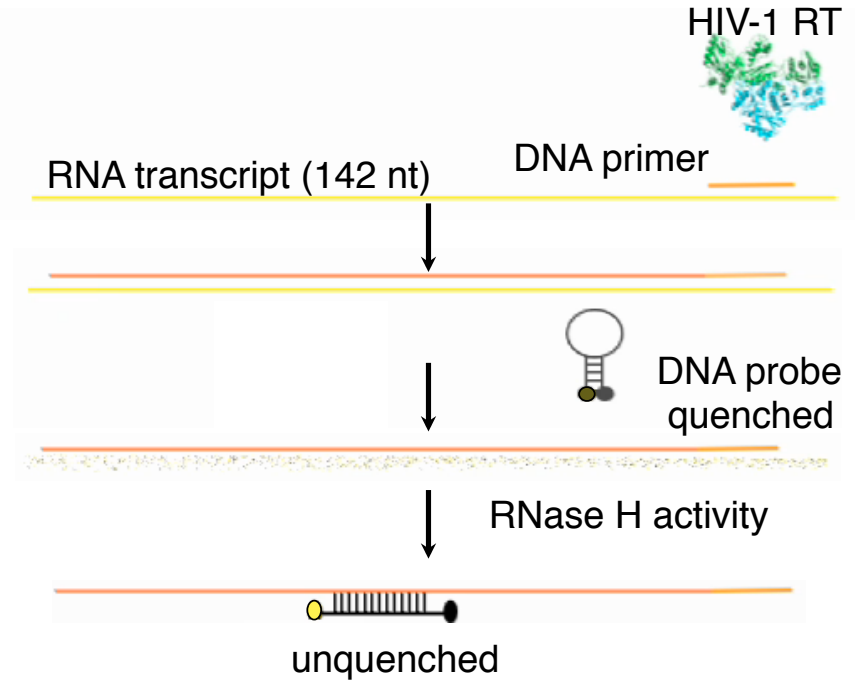


Figure 13: Schematic drawing depicting the JDAM activity assay.

The Z' score for JDAM was determined to be an acceptable 0.74 (Zhang et al., 1999). All fragments and derivatives found bound to RT were tested for inhibition using JDAM. Each inhibitory fragment was then co-crystallized with RT in the absence of rilpivirine to confirm the hit identity and to determine whether or not the fragment was an NNRTI.

2.10 Allosteric Binding Sites

A total of 606 datasets were collected with an average resolution of 2.1 Å from screening 742 fragments. Out of 606 datasets, 43% of the datasets collected were from cocktail soaks, 34% for individual compounds, 14% fragment derivatives, and 9% were co-crystals with individual fragments. The high percentage of individual soaks stemmed from difficulty with identifying weakly bound fragments.

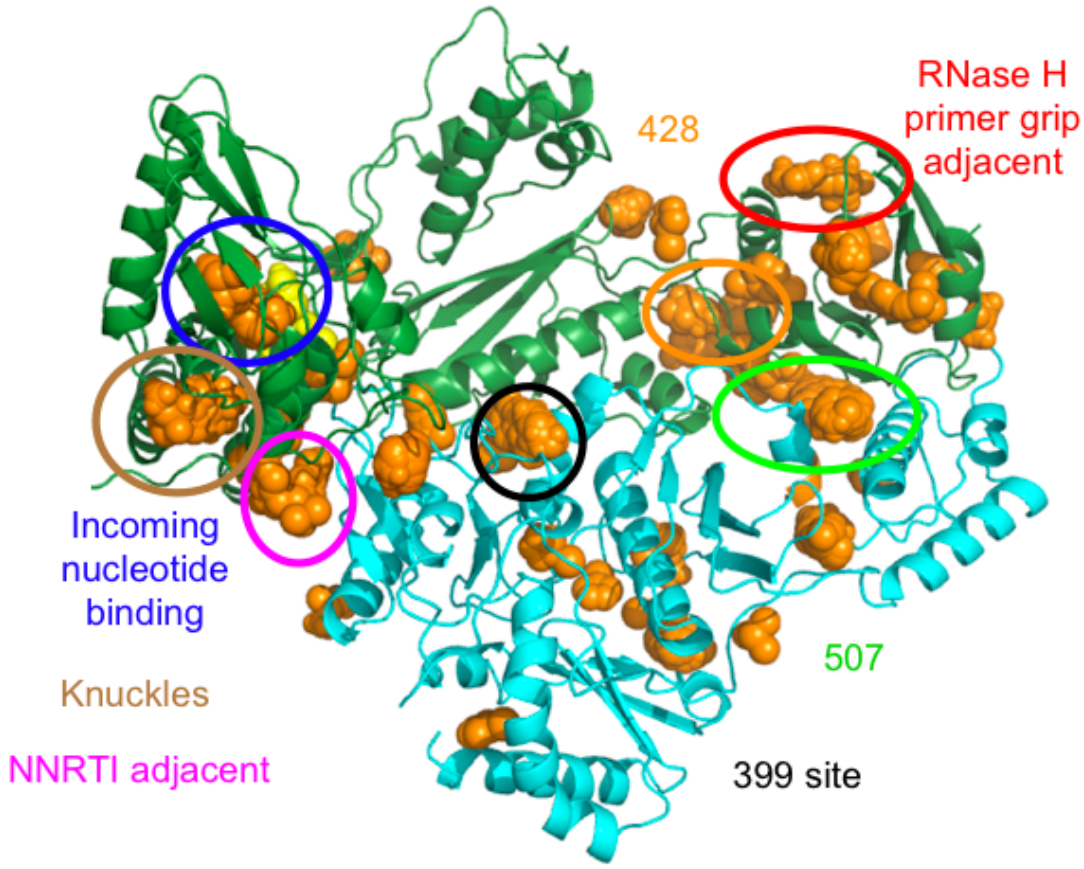


Figure 14: Drawing representing p66 (dark green), p51 (cyan), and rilpivirine (yellow spheres) with all fragments discovered in this study (orange spheres). The sites of fragment binding described in this report are circled.

A hit rate of 4.4% was determined after the identification of 34 bound fragments to 16 different sites throughout RT. Seven of the 16 sites were found to be biologically relevant (Figure 14). Fragments binding to three (NNRTI adjacent, Knuckles, and incoming dNTP) of the seven sites were found to be inhibitory in the dual activity monitor assay. Each of the seven sites is described in detail below.

2.10.1 NNRTI Adjacent Binding Site

The NNRTI Adjacent binding site (NABS) is located at the interface of p66/p51 interface next to Glu138 of p51. The pocket is tunnel-like with Thr139 (p51), Pro140 (p51), Thr165, Leu168, Lys172, and Ile180 forming an opening of the tunnel. This

extends into another cavity near the polymerase active site at Gln91 (Figure 15). This pocket is separated from the NNRTI-binding pocket by the β9 strand.

Of the compounds binding at this site, the best electron density and inhibition in JDAM was observed for **1**, which had an IC_{50} value of 350 μM with a ligand efficiency of 0.34 kcal/mol•NHA (non-hydrogen atom). **1** formed hydrogen bonds with the backbone carbonyl of Ile180 and the backbone amide nitrogen of Gln182 as well as hydrophobic interactions with Pro140 and Ile180. To accommodate fragment binding, side chains of Gln161 and Gln182 rotate, which allows Gln161 to participate in a hydrogen bond network. This network consists of a hydrogen bond between Gln161 and the backbone amide of Gln91, which also forms a hydrogen bond with a backbone carbonyl of Gln182. The formation of this network causes a 2.1 Å Cα backbone movement of Gln91 that propagates to the surrounding residues. Comparison of the bound and unbound structures revealed a 1.1 Å Cα backbone movement of Gln182, which in turn causes the catalytic YMDD motif to shift by a 1.5 Å. Thus, inhibition of **1** is attributed to its ability to distort the YMDD motif and alter the interaction network of this region.

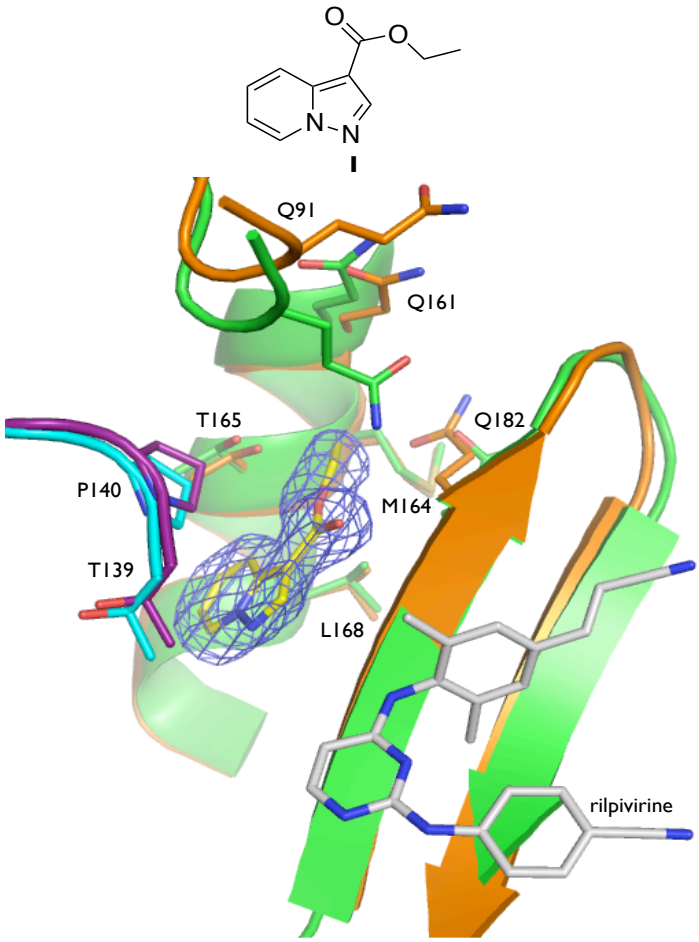


Figure 15: Crystal structure of RT-rilpivirine with **1** bound at the NNRTI Adjacent site. Residues that shift upon ligand binding are shown in purple (p51) and orange (p66) and in their non-bound positions as cyan (p51) and green (p66). Rilpivirine is shown on the right side of the panel (white). Electron density is shown for **1** from a *F*o – *F*c omit map (3.5σ contours).

Druggability of the NABS was assessed with respect to rapid fragment to lead development as well as emergence of possible drug resistance. The proximity of the NNRTI pocket suggested the possibility of fragment linking. The presence of a strongly ordered water molecule, approximately 4.1 Å away from **1** and 3.7 Å from rilpivirine, supported the possibility of designing agents to bind to both the NNRTI and NNRTI Adjacent sites. Fragment evolution of the initial fragment hit could also be an alternative

lead development strategy. This approach can allow for the development of lead molecules against wild-type and known and predicted mutant RT variants.

Mutational analysis of 1,809 sequences of RT from clinical isolates was conducted to predict possible mutations that can arise for residues comprising the NABS pocket (HIV Sequence Base, 2009). Residues with <1% mutations compared with the consensus were defined as highly conserved. With the exception of Lys172, which is commonly Arg, residues forming NABS that are not involved in NNRTI resistance were highly conserved. The high conservation of residues forming key interactions (Pro140, Ile180, and Gln182) suggests this pocket may be attractive for lead development. However, NNRTI-resistance mutations such as Glu138Lys, Thr139Ala, Thr165Ile, and Tyr181Cys, should be taken under consideration during development.

2.10.2 Knuckles Site

Similar to the NNRTI-binding pocket, the Knuckles pocket, located between the p66 finger and palm subdomains, is stabilized by fragment binding. Without fragment present, the Knuckles pocket is a buried cavity near the polymerase active site. When fragment binds to the open form of this cavity, it stabilizes Ser117 such that it is 2.8 Å away from its position in the unliganded form (Figure 16). Fragment binding accompanies the fingers subdomain to shift 1.2 Å as well as a 3.2 Å polypeptide backbone shift of key residues (Tyr115 and Phe116) involved in binding of the incoming dNTP during polymerization.

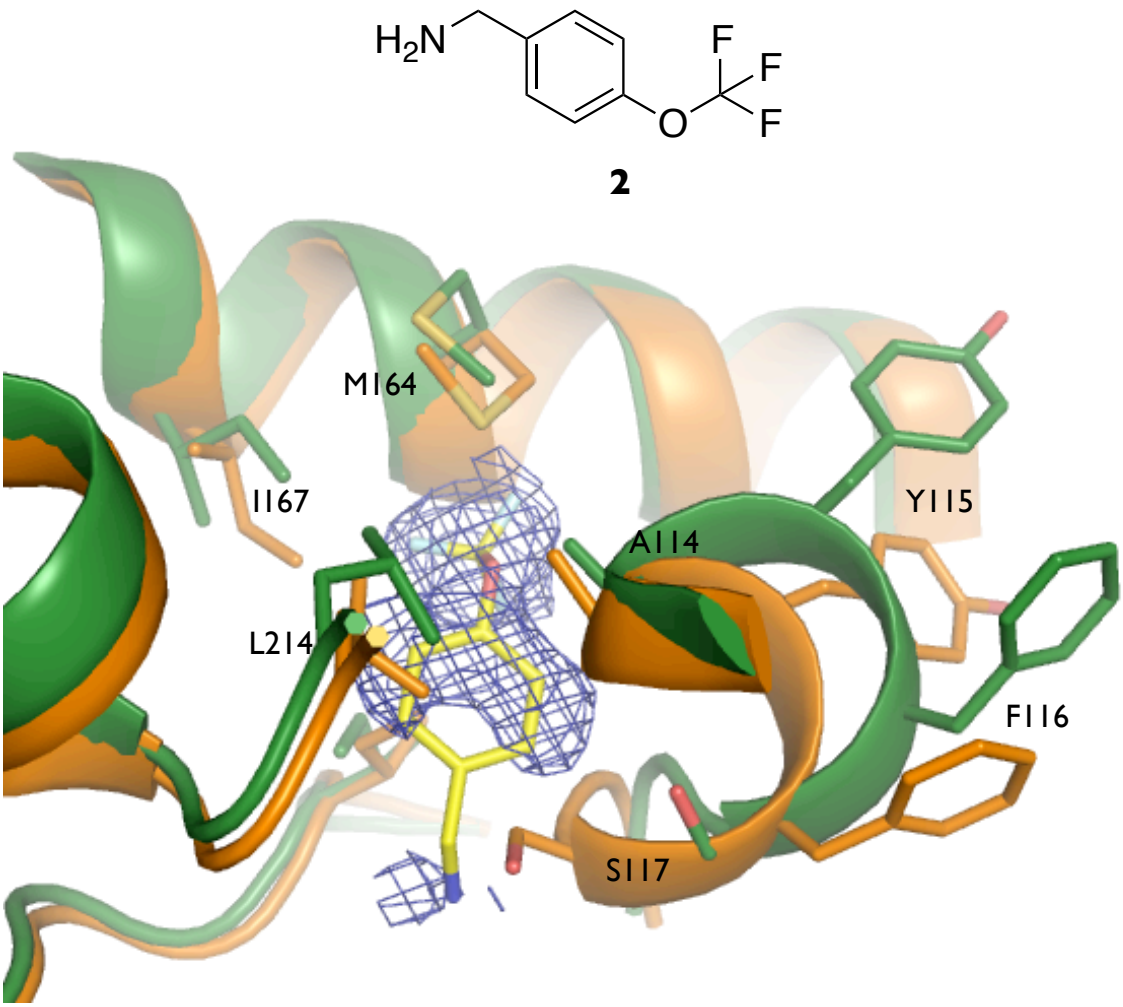


Figure 16: Crystal structure of RT-rilpivirine with **2** bound at the Knuckles site. Residues that shift because of ligand binding are shown in orange and in their nonbound positions as green. Electron density is shown for **2** from a *F*o – *F*c omit map (3.5σ contour).

The strongest electron density was observed for the binding of **2** (Figure 16). All derivatives of this fragment shared electron density for the trifluormethoxy. Thus, **2** could be successfully modeled despite the lack of electron density for the full molecule. Fragment binding to this site is predominantly characterized by hydrophobic interactions. The aromatic ring of **2** forms hydrophobic interactions with Ile5, Ala114, Leu214, and Val118 of p66. The trifluoro moiety also forms halogen bonds with Met164 and Ser117. Mutational analysis revealed that residues comprising this pocket are high conserved with the exception of Leu214, which is predominantly phenylalanine.

Derivatives based on **2** were purchased and assayed for inhibitory activity. Of these compounds, highest inhibition was observed upon the replacement of the methylamino with a hydroxyl (**3**) at the para position of the benzene ring. **3** had an IC_{50} of 600 μM with a corresponding ligand efficiency of 0.37 kcal/mol NHA whereas the remaining compounds were inhibitory at concentrations > 1 mM.

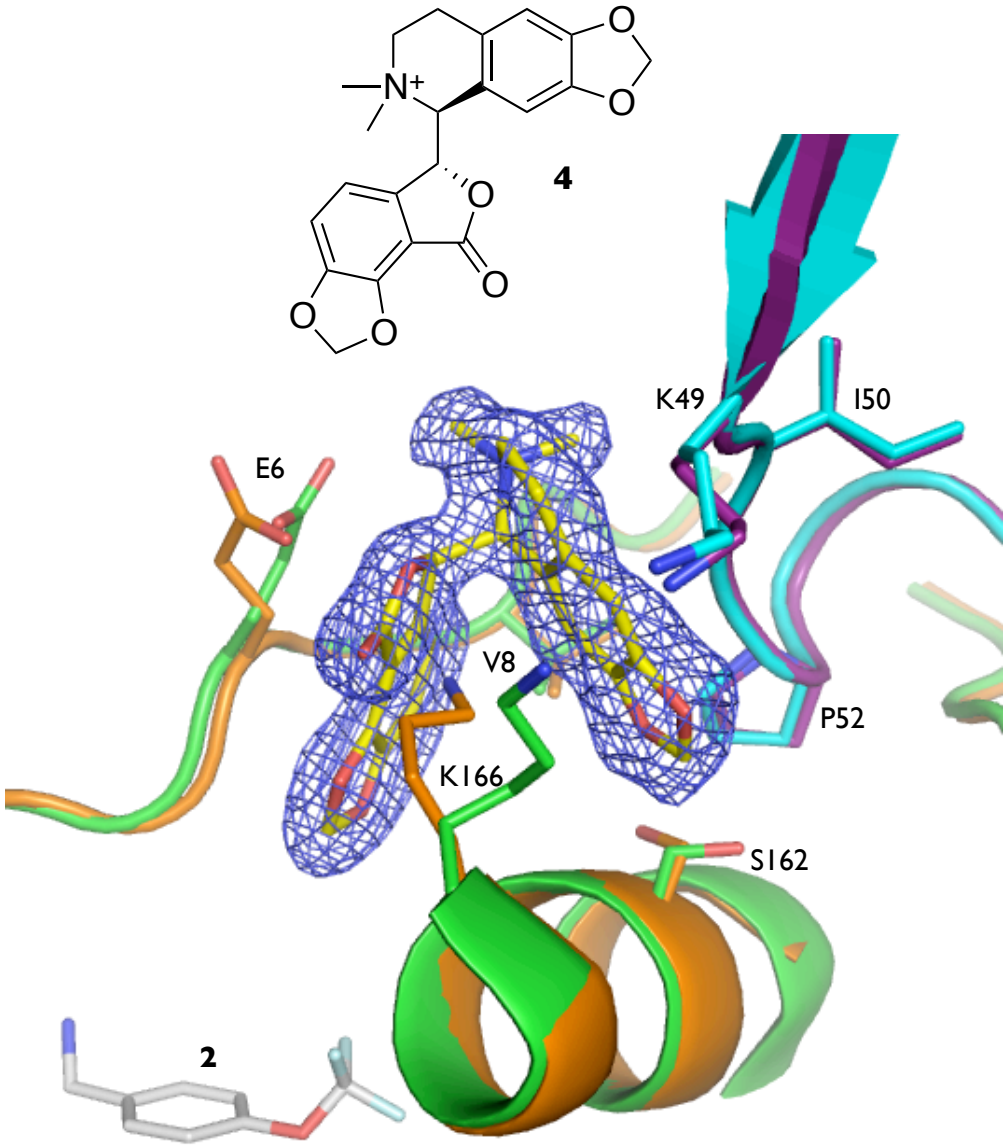


Figure 17: Crystal structure of RT-rilpivirine with **4** bound to the Knuckles site and the position of **2** shown at the bottom of the panel. Residues that shift because of ligand binding are shown in purple (p51) and orange (p66) and in their nonbound positions as cyan (p51) and green (p66).

The promising inhibitory activity of the identified fragment hits and the highly conserved nature of the Knuckles pocket make it attractive for drug design. As shown in Figure 17, fragment growth can occur either in the direction of the solvent exposed Ser117 or Ile167. Binding of **4** ~5 Å away from **2** makes fragment expansion by Ile167 an attractive possibility. Further lead development can take advantage of the hydrogen bonding interaction **4** has with Ser163 as well as the hydrophobic interactions with Lys166, Gly51, and Glu6.

2.10.3 Incoming Nucleotide Binding Site

The incoming nucleotide binding site, which comprises Tyr115, Phe116, and Gln151, expands to accommodate fragment binding. Pocket expansion occurs upon binding of **5** due to the reorientation of Gln151 (Figure 18). Structural analysis of **5** bound to RT revealed electrostatic interactions between the piperazine ring of **5** and the carboxylate of Asp185, an invariant residue required for polymerization. The carboxylate of **5** also forms hydrogen bonds with the amine of Lys73, hydroxyl of Tyr146, and backbone amide of Gln151. The aromatic ring of **5** also forms edge-to-face π-stacking interactions with Tyr115 and Phe116. Compound **5** was found to be an excellent candidate for lead development with an IC_{50} value of 200 μM and LE of 0.31 kcal/mol•NHA.

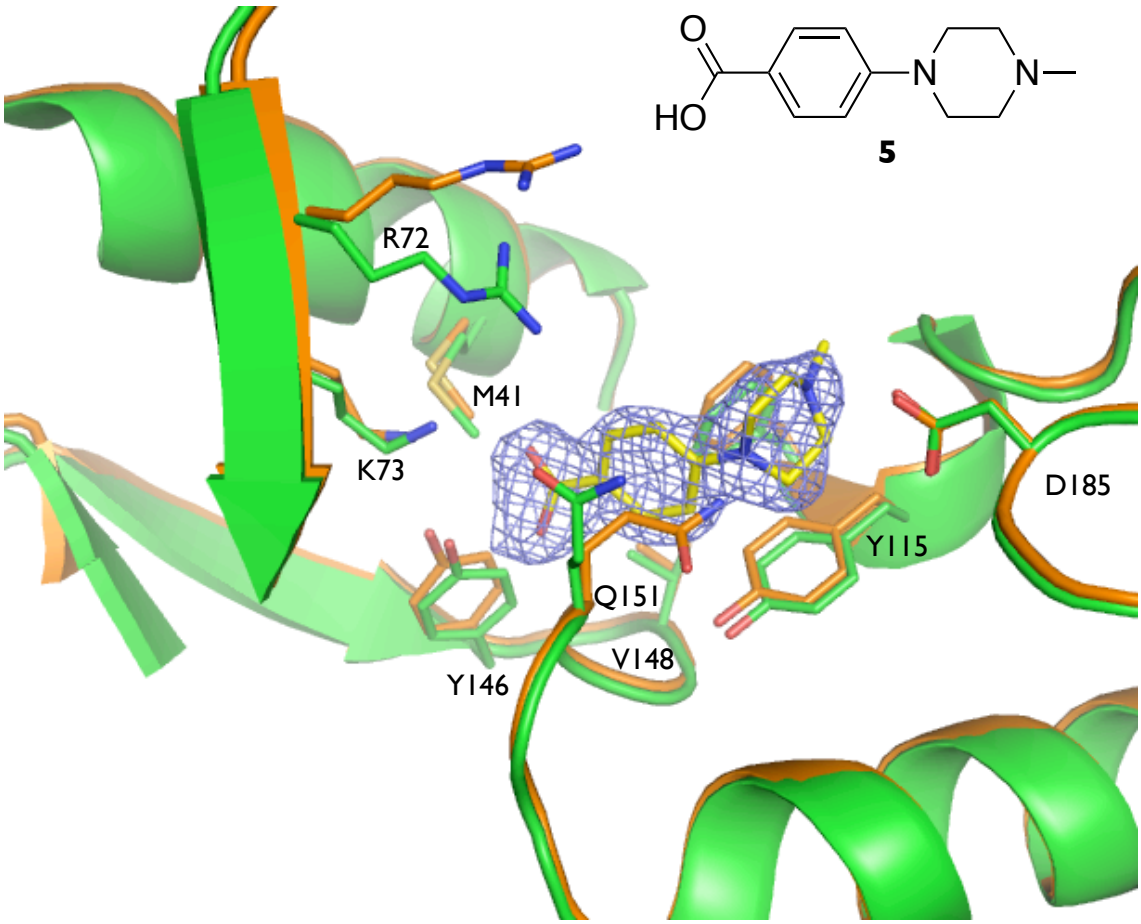
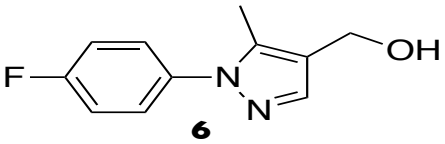


Figure 18: Crystal structure of RT-rilpivirine-**5** bound to the incoming dNTP site. Residues that shift because of ligand binding are shown in orange and in their nonbound positions as green. Electron density is shown for **5** from a F_o-F_c omit map (3.0σ contours).

This site is extremely attractive for future drug design efforts, as compound binding to this would competitively inhibit polymerization by interfering with dNTP binding. Additionally, this site is highly conserved thus evolution of drug resistance would have negative implication on RT activity and, in turn, viral replication. Several groups have reported that mutations at Tyr115 to polar amino acids lead to severe defects RT activity whereas substitution with hydrophobic amino acids lead to an intermediate effect. Thus, development of **5** can potentially lead into a new class of RT inhibitors that prevent polymerization by competing with dNTP binding.

2.10.4 399 Site

The stability of the p66/p51 heterodimer is essential for proper RT activity. This has been biochemically demonstrated by the development of peptides and small molecules that disrupt dimerization. However, further development of these dimerization inhibitors has been hindered due to lack of crystal structures with the bound agents. Fragment screening revealed binding of **6** to a pocket located at the p66/p51 interface, referred to as the 399 site. This pocket consists of Trp24, Pro25, Glu399, Thr400, Trp402, and Thr403 from p51 as well as Thr377 and Ile380 from the p66 subunit (Figure 19).



6

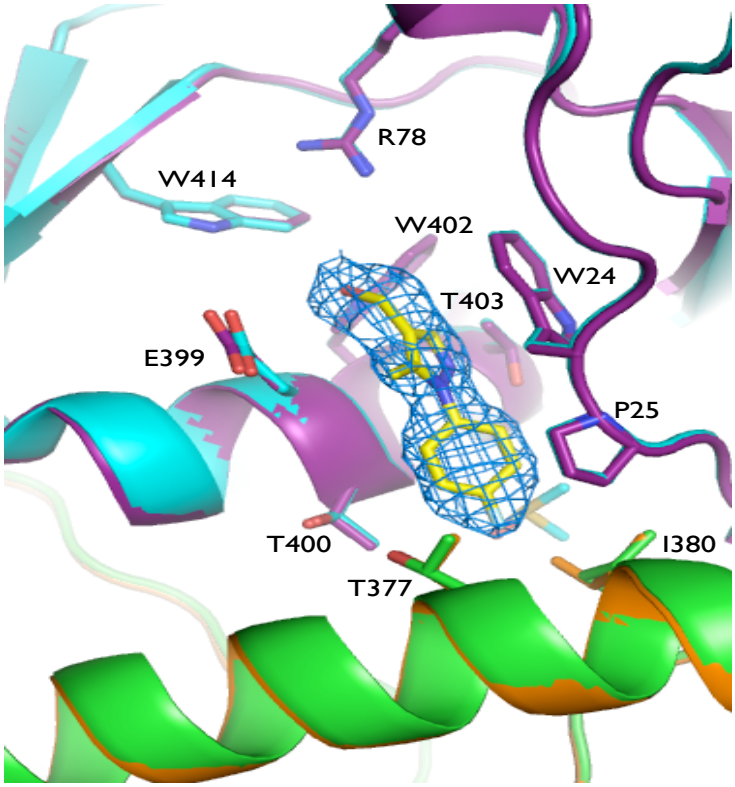


Figure 19: Crystal structure of **6** bound at the 399 site. Residues that shift upon ligand binding are shown in purple (p51) and orange (p66) and in their nonbound positions as cyan (p51) and green (p66). Electron density is shown for **6** from a F_o-F_c omit map (1.5σ contours).

Compared to the fragments discussed thus far, **6** only forms a limited number of interactions with the protein. It forms a hydrogen bond with the backbone carbonyl of Glu399 and hydrophobic interactions with Trp24 and Pro25. No enzymatic inhibition was observed for **6** or any other fragments found bound to this site. This is primarily attributed to the limited number of fragment-protein interactions in combination with the large size and the apparent rigidity of the site in several RT structures. It is possible that larger compounds with the fragment hit as the pharmacophore could be potentially inhibitory.

2.10.5 428 Site

Located in the connection subdomain of p66, the 428 site consists of residues Glu396, Thr397, Thr400, Leu425, and Gln428, whose side chain rotates 70° to allow for binding of **7** (Figure 20). Fragment binding is stabilized by hydrogen bonds between the hydroxyl of **7** and Thr400 and the amine of **7** and the carboxylate of Glu430. Additional hydrophobic interactions between **7** and Gln428 and Leu425, respectively, are also observed. **7** was found to inhibit 30% of RT activity at 1 mM in the JDAM assay. The presence of a DMSO molecule located 2.6 Å away from **7** suggests potential for improving the inhibitory activity of **7** via a fragment expansion approach.

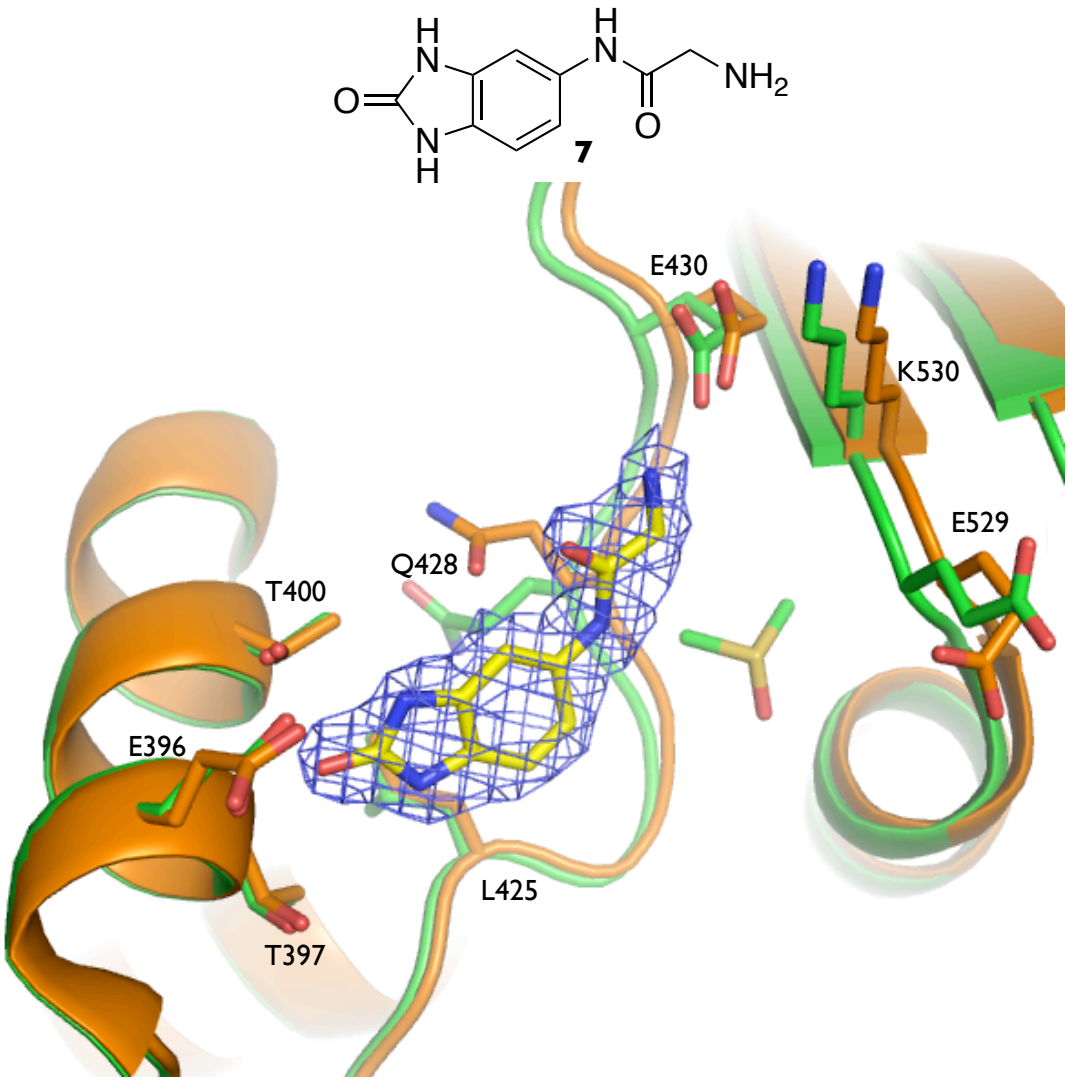


Figure 20: Crystal structure of RT-rilpivirine with **7** bound at the 428 site. Residues that shift due to ligand binding are shown in orange and in their non-bound positions as green. Electron density is shown for **7** from a F_o-F_c omit map (2.5σ contours).

2.10.6 507 Site

The 507 site is a large solvent-exposed pocket at the base of the RNase H domain located ~9 Å away from the 428 site. Binding of **8** to this site was fortuitous as it was predicted to bind to the Knuckles site based on docking studies. Strong electron density was observed for **8**, which forms hydrogen bonds between with the hydroxyl of Tyr532 and the backbone carbonyl of Leu533 (Figure 21). It also forms hydrophobic interactions with Lys259 from p51 and Leu429, Leu533, Ala534, and Trp535 from the p66 subunit.

An IC_{50} of 150 μM using JDAM and a ligand efficiency of 0.19 kcal/mol•NHA was determined for **8**. Despite its modest inhibitory activity, the low ligand efficiency of **8** suggests this hit needs be further optimized.

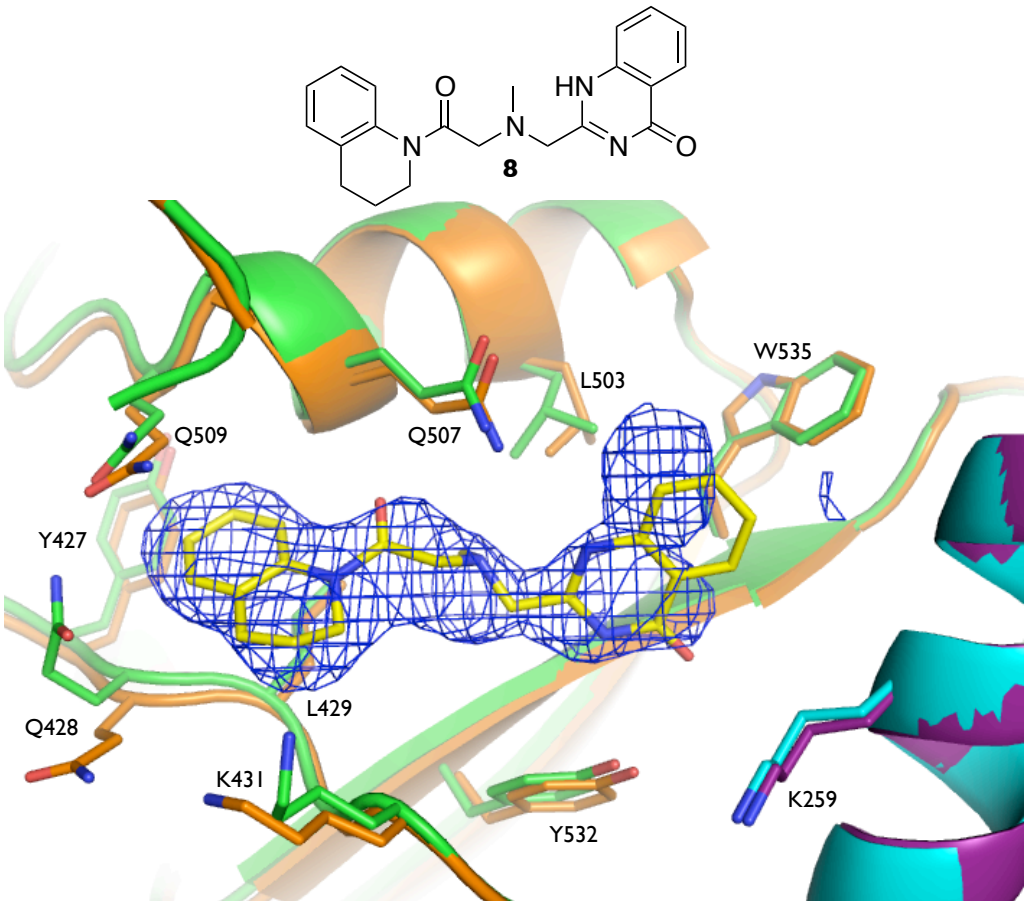


Figure 21: Crystal structure of RT-rilpivirine with **8** bound at the 507 site. Residues that shift due to ligand binding are shown in purple (p51) and orange (p66) and in the unbound form in the RT52A-ripilvirine structure as cyan (p51) and green (p66). Electron density is shown for **8** (yellow) from an F_o-F_c omit map (2.5σ contours).

Co-crystallization of **8** in complex with RT without rilpivirine present was conducted revealed binding of **8** to the NNRTI-binding pocket (Figure 22). Key interactions includes hydrogen bonding with the backbone amide of Lys101 and hydrophobic interactions with Lys103, Val106, Val179, Ile180, Tyr181, Tyr188, Phe227, Trp229, Leu234, and Tyr318. Based on the quality of electron density, the modest inhibitory activity observed for **8** is attributed to its ability to bind as an NNRTI.

Exploration of the inhibitory potential of the 507 site requires further investigation by screening for compounds that would exclusively bind the 507 site and not the NNRTI-binding pocket.

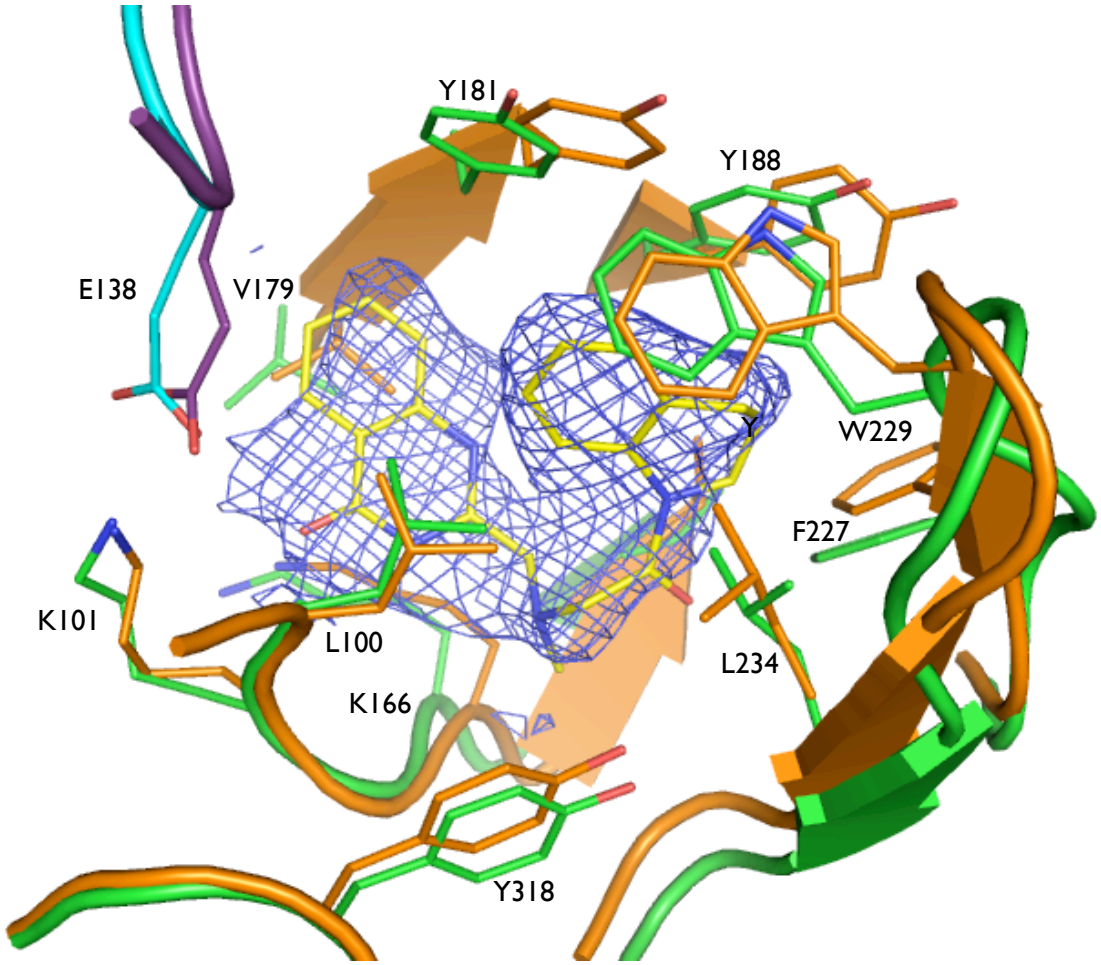


Figure 22: Crystal structure of **8** bound at the NNRTI-binding pocket without rilpivirine present. Residues that shift due to ligand binding are shown in purple (p51) and orange (p66) and in the RT52A-ripilvirine structure as cyan (p51) and green (p66). Electron density is shown for **8** (yellow) from an F_o-F_c omit map (4.0σ contours).

2.10.7 RNase H Primer Grip Adjacent

The RT crystallographic fragment-screening campaign revealed fragment binding (**9**) to a site composed of Leu469, Lys476, Gln480, Tyr483, Glu516, Leu517, and Gln520 (Figure 23). Fragment binding to this pocket, referred to as the RNase H Primer Grip Adjacent site, occurs upon the side-chain rotations of Gln480 and Gln520. Although no

inhibitory activity was observed for fragments binding to this site, a fragment, specifically **9**, formed favorable interactions with the protein. This includes the formation of a small hydrogen-bond network between carboxylate of **9** and Gln480 and Lys476, which is part of the RNA primer grip. The pyrazole of **9** also forms an additional hydrogen bond with Gln520 as well as hydrophobic interactions with Lys476, Gln480, Tyr483, Leu517, and Gln520. Key interactions with residues that are associated with RNA binding combined with room for extension into the RNA primer grip or the RNase H active site suggests that fragment expansion may lead to favorable inhibition.

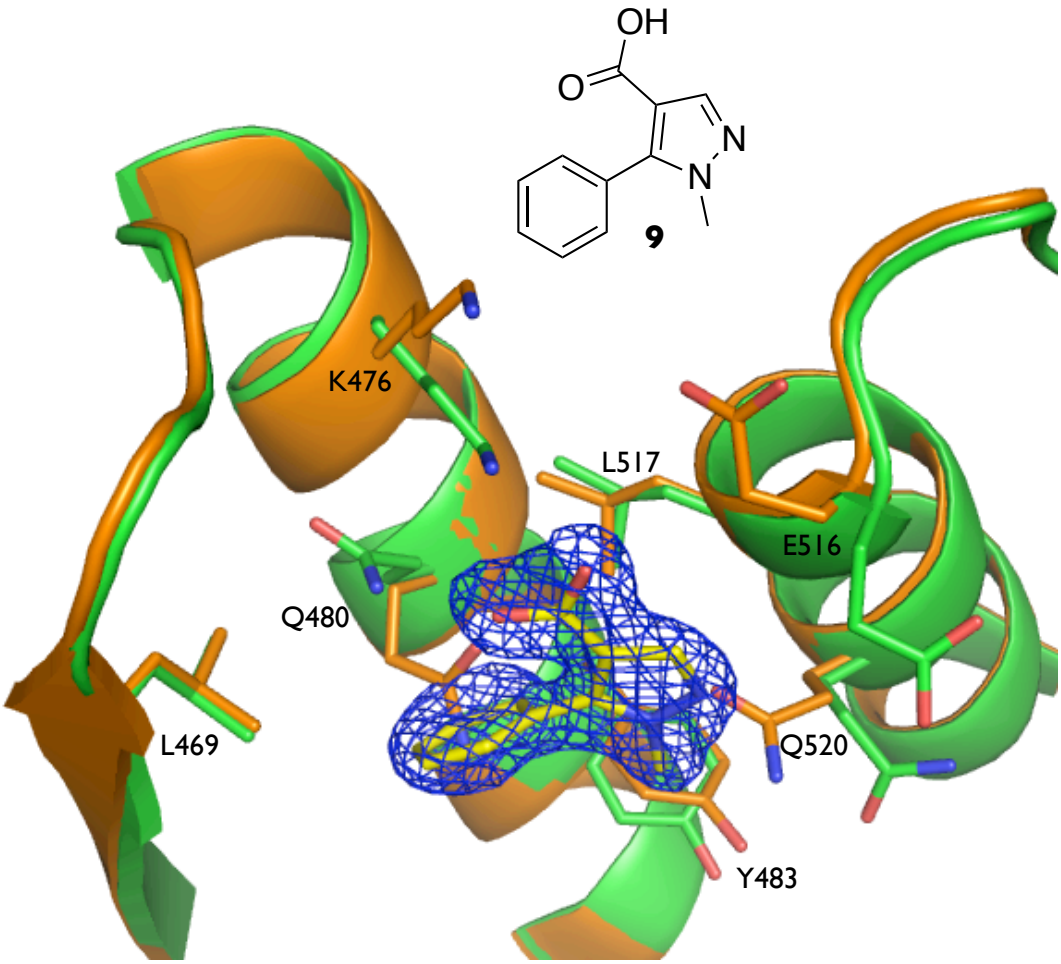


Figure 23: Crystal structure of RT-rilpivirine with **9** bound at the RNase H primer grip adjacent site. Residues that shift due to ligand binding are shown in orange and in the unbound form in the RT52A-ripilvirine structure as green. Electron density is shown for **9** (yellow) from a F_o-F_c omit map (3.0σ contours).

2.11 Conclusion

Fragment screening by X-ray crystallography was successfully conducted to probe for promising druggable allosteric pockets within HIV-1 RT. The success of this approach heavily relied on the selection of a suitable crystal form to limit the highly flexible nature of RT and optimization of the soaking conditions to ensure fragment solubility and crystal stability. High-throughput screening was achieved by taking advantage of a low-resolution data collection pass coupled with rapid hit identification using F_o-F_o maps generated using "DMSO only blank" dataset as the reference.

The crystallographic screening campaign identified seven sites of biological interest near the polymerase and RNase H active sites. Fragment binding to three of these sites (the NNRTI Adjacent, incoming dNTP binding, and Knuckles) were found to be inhibitory using a high-throughput enzymatic activity assay (JDAM). Availability of the structural data indicates room for fragment expansion into more lead-like compounds with strong inhibitory activity. Analysis of more than a thousand RT sequences from clinical isolates revealed that the residues forming these three sites are conserved, thereby limiting the potential ability of the virus to develop drug resistance to inhibitors targeting these sites without having a detrimental effect on RT activity.

Fragment screening revealed a higher hit rate for halogenated fragments in comparison to non-halogenated fragments. Seven out of the 29 fluorinated fragments and four out of the 17 brominated fragments screened were identified as hits, thereby giving hit rates of 24.1% and 23.5%, respectively. Our fragment library also contained two chlorinated fragments of which one was found to bind. The higher hit rates observed for

halogenated compounds suggest that biasing fragment libraries with halogenated compounds could be potentially be beneficial.

Brominated and fluorinated compounds, in particular, are advantageous due to their use for anomalous X-ray diffraction (Blaney et al., 2006) and isotope enrichment for NMR spectroscopy (Dalvit et al., 2002), respectively. Despite being weaker than hydrogen-bonding and hydrophobic interactions, halogen-bonding interactions with protein residues can strengthen fragment binding. Additionally, halogens are an excellent handle from the point of view of synthetic chemistry.

2.12 Materials and Methods

2.12.1 Expression, Purification, and Crystallization

HIV-1 RT constructs RT52A (crystallization optimized mutant) and RT35A (wild-type) were constructed, expressed, and purified as described by Bauman et al., 2008. Crystallization was performed using the hanging-drop method with EasyXtal DG-Tools (Qiagen, Valencia, CA) crystallization trays. Prior to crystallization, RT52A (20 mg/ml) was incubated with rilpivirine (TMC278/Edurant) at 1:1.5 molar protein-to-drug ratio at room temperature (~23°C) for 30 minutes. RT52A-rilpivirine crystals were produced in hanging drops at 4°C with a 1:1 ratio of protein and well solution containing 11% PEG 8000, 4% PEG 400, 50 mM imidazole pH 6.6, 10 mM spermine, 15 mM $MgSO_4$, 100 mM ammonium sulfate, and 5 mM tris(2-carboxyethyl)phosphine and an experimentally optimized concentration of microseeds from previously generated crystals (preseeding).

2.12.2 Fragment Soaking, Data Collection, and Processing

The fragment/cryo soaking solutions were prepared with crystallization well solution with the addition of 80 mM L-arginine, 5% (v/v) ethylene glycol, and 20% (v/v) d_6-DMSO (containing 20 mM final concentration each fragment). Crystals of RT52A-rilpivirine were harvested two weeks to four months after crystals formed. The crystals were placed in fragment/cryo soaking drops for one to two minutes before flash cooling in liquid nitrogen. Data collection was performed at the Cornell High Energy Synchrotron Source (CHESS) F1 beamline and the National Synchrotron Light Source (NSLS) X25 and X29 beamlines. The diffraction data were indexed, processed, scaled and merged using *HKL2000I* (Otwinowski et al., 2011). Initial datasets from crystals were collected to minimize the X-ray expsoure by increasing the oscillation range per image and decreasing exposure time. $F_{obs(fragment\ soaked)}$-$F_{obs(DMSO\ blank)}$ maps (as described previously) were immediately calculated using CNS (Brunger, et al., 1998) and visualized with Coot (Emsley et al., 2010). Datasets for crystals containing bound fragments were then recollected to improve maximum X-ray diffraction resolution. High-resolution datasets containing bound fragments were further refined using PHENIX (Adams, et al., 2010) and Coot (Emsley et al, 2010). Crystal structure figures were made with MacPyMol (Schrödinger, New York, NY) unless otherwise noted.

2.12.3 Dual Activity Monitor Assay

The 142-nucleotide RNA strand with the following sequence was produced with the MEGAshortscript kit and purified with the MEGAclear kit (Ambion, Austin, TX): GTCTATCTGGCATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAA GTAGATAAATTAGTCAGTGCTGGATAACTCGAGTCTGGTAAAGAAACCGCTG CTGCGAAATTTGAACGCCAGCACATGGACTCGTCTACTA. The DNA primer has

the sequence TAGTAGACGAGTCCATGTGC. The probe was (FAM)CCGGGGAATTGGAGGAAATGAACAAGTAGCCGG(Iowa-black). All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). RT (construct RT35A–wild type) was preincubated with inhibitor for 30 minutes at 37°C. The preincubation mixture contained 100 nM RT35A, 10 mM dNTPs, 1 mg/ml BSA, 1 mM DTT, 10 mM CHAPS, 1% (v/v) d_6-DMSO (with inhibitor), and 50 nM probe. Reactions were started with the addition of pre-mixed primer (225 nM) and RNA (200 nM). Reactions were incubated at 37°C and then polymerization was stopped with the addition of 500 nM RNase A (Qiagen, Valencia, CA) after 2 minutes. Fluorescence was immediately measured in a Varioskan microplate reader (Thermo Scientific, Rockford, IL). Dose-response curves were calculated by fitting data to a sigmoidal four-parameter logistics equation using Prism (GraphPad). Ligand efficiency was calculated using LE (in kcal/mol•NHA) = $-RT\ln(IC_{50})/NHA$.

Chapter 3: Identification of Novel Chemical Scaffolds Targeting the Endonuclease Domain of Influenza Type A Using Fragment Screening by X-ray Crystallography

Synopsis

There continues to be great pressure to develop highly efficacious drug therapy targeting influenza Type A viral infection due to emerging resistance to the currently available antivirals and the constant threat of possible pandemic strains. The RNA-dependent RNA polymerase (RdRP) of influenza, specifically inhibition of its endonuclease activity, is an attractive target for new drug therapy. Several structural and biochemical studies have established that the endonuclease activity of RdRP resides in the N-terminal domain of PA (PA_N) subunit from RdRP. Fragment screening by X-ray crystallography was conducted to identify promising new chemical scaffolds for endonuclease inhibitors, specifically targeting H1N1 PA_N. This chapter describes a few of the interesting fragment hits, some of which coordinated to a previously unseen third metal within the endonuclease active site. Additionally, this chapter also describes the development of a new class of endonuclease inhibitors, the hydroxypyridinones, through iterative cycles of structural biology, computational modeling, and chemical synthesis.

The work described in this chapter related to the development of the hydroxypyridinone class of endonuclease inhibitors was done in collaboration with the LaVoie laboratory at Rutgers University and the Martinez-Sobrido laboratory at University of Rochester. Specifically, chemical synthesis was performed by Ajit Parhi and Amy Xiang from the LaVoie lab and the antiviral activity assays were conducted by Steven Baker from the Martinez-Sobrido laboratory. The information discussed in this chapter either has been published or is currently in press.

Publications:

Bauman, J. D., Patel, D., Baker, S., Vijayan, R. S. K., Xiang, A., Parhi, A., Martinez-Sobrido, L., LaVoie, E. J., Das, K., Arnold, E. (2013) Crystallographic Fragment Screening and Structure-Based Optimization Yields a New Class of Influenza Endonuclease Inhibitors. *ACS Chemical Biology*, in press DOI: 10.1021/cb400400.

Parhi, A., K., Xiang, A., Bauman, J. D., Patel, D., Vijayan, R. S. K., Das, K., Arnold, E., LaVoie, E. J. (2013) Phenyl substituted 3-hydroxypyridin(1H)-2-ones: inhibitors of influenza A endonuclease. *Bioorganic & Medicinal Chemistry*, in press DOI: 10.1016/j.bmc.2013.08.053.

Sagong, H. Y., Parhi, A., Bauman, J. D., Patel, D., Vijayan, R. S. K., Das, K., Arnold, E., LaVoie, E. J. (2013) 3‐Hydroxyquinolin-2(1H)‐ones As Inhibitors of Influenza A Endonuclease. *ACS Medicinal Chemistry Letters*, **4**, 547-550.

Viral evolution continues to be a major challenge for the development of antiviral therapy. This is clearly demonstrated by the influenza virus, commonly known as flu, which causes epidemics on a yearly basis as well as pandemics pose a serious public health concern. In the last century alone, there have been three major pandemics—the 1918 Spanish Flu, the 1957 Asian Flu, and the 1968 Hong Kong Flu—that have taken 50 million, 2 million, and 1 million lives, respectively (Monto, 2005). Vaccination is the primary defense against influenza pandemics however large-scale vaccine production is often times difficult during seasonal epidemics (Gomez et al., 2013). Thus, prophylactics and antiviral drug therapy are often the front line for the treatment of influenza infection. However, the efficacy of these drugs is greatly reduced by the rapid emergence of drug-resistant strains of the influenza virus. New therapeutic options are needed to not only address the emerging resistance to drug therapy but also curtail the possibility of another pandemic threat.

3.1 Influenza Virus

Influenza, a member of the *Orthomyxoviridae* family of viruses, has negative sense, single-stranded viral RNA genome consisting of eight segments (Palese and Shaw, 2007; Das et al., 2012). These segments encode for eleven viral proteins, specifically hemagglutinin (HA), neuraminidase (NA), matrix 1 (M1), matrix 2 (M2), nucleoprotein (NP), non-structural protein 1 (NP1), non-structural protein 2 (NP2), polymerase acidic protein (PA), polymerase basic protein 1 (PB1), and polymerase basic protein 2 (PB2) (Bouvier and Palese, 2008). Influenza virus can be classified into three major forms, referred to as types A, B, and C. All three types vary with respect to pathogenicity and host range, which includes humans. Both influenza type B and C viruses are typically

found in humans whereas influenza type A viruses have been isolated from a range of animals, including birds, swine, and horses. On the other hand, influenza type B viruses have been isolated from seals (Osterhaus et al., 2000; Bodewes et al., 2013) while influenza C viruses have been isolated from pigs and dogs (Guo et al., 1983; Manuguerra and Hannoun, 1992).

Of the three, influenza A and B viruses pose a greater threat to humans and, as a result, have been well studied in biological and clinical settings. Influenza type C is of limited clinical significance as it only causes mild symptoms and cases rarely have been reported (Muraki and Hongo, 2010). Influenza type B viruses are mildly pathogenic but do not mutate rapidly. Influenza type A viral infections pose a great threat to humans as it is fast spreading and has potential to be highly fatal.

Influenza type A viruses are highly dynamic and constantly changing. These changes can either occur by antigenic shift or antigenic drift (Zambon, 1999). Antigenic drift produces a new strain by slightly alterations of the virion surface resulting from point mutations within HA and NA, thereby preventing effective binding of neutralizing antibodies. Antigenic shift introduces a new viral strain to the human population through animal to human transmission. Alternatively, antigenic shift can occur through genetic reassortment, which produces a new human influenza A subtype through mixing of influenza A genes from human and animal. Wild aquatic birds are the natural reservoir for type A viruses, with a total of sixteen subtypes of HA (H1-H16) have been identified (Fouchier, et al., 2005) and nine types of NA (N1-N9) have been identified in wild birds and poultry (Hinshaw et al., 1982; Kawaoka et al., 1990; Rohm et al., 1996). Of these, only a handful of the subtypes have been to be infectious in humans, including H1, H2,

H3, H5, N1, N2, N3. Viral evolution produces several different subtypes, including seasonal epidemic strains such as H1N1, H2N2, and H3N2.

Influenza A and B viruses are structurally similar with spherical or ovoid shaped capsids with a diameter of 80 to 120 nm. Influenza C viruses are structurally distinct and can form long cordlike structures on the surface of the infected cells (Bouvier and Palese, 2008). The viral envelope contains a lipid bilayer in which is embedded, hemagglutinin, which forms spiky extensions that account for ~80% of the shell, the mushroom-like neuraminidase, and the M1 protein. The spikes formed by the homotrimeric HA binds to sialic acid found on the membrane surface of the host cell. Upon binding, receptor mediated endocytosis allows the virus to enter the host cell. The M2 protein, which is a proton channel, lowers the pH in the endosome to initiate viral uncoating and subsequent release of the viral genome into the cell. Neuraminidase facilitates the release of new viral particles by removing sialic acid. The organization of HA and NA in the outer shell is used to distinguish influenza viruses.

3.2 Anti-Influenza Drug Therapy

Current FDA-approved inhibitors of influenza target the activity of the M2 ion channel protein and neuraminidase. The M2 ion channel inhibitors, amantadine (Symmetrel) and rimantadine (Flumadine), are no longer recommended for treatment due to emergence of drug resistance in the circulating influenza strains, including the pandemic 2009 H1N1 commonly referred to as pH1N1 (FDA Influenza antiviral drugs and related information, 2013). Current antiviral therapy relies on inhibitors of neuraminidase activity, zanamivir (Relenza) and the widely used oseltamivir (Tamiflu). Two approaches for the development of new antiviral therapy can be applied to address

the circulation of oseltamivir-resistant strains. The first approach entails the development of next generation M2 ion channel and neuraminidase inhibitors that retain efficacy against the emerging drug-resistant strains. Alternatively, drug development could be pursued for new target viral proteins that are not only essential for viral activity but also highly conserved among all three types of influenza viruses.

The RNA-dependent RNA polymerase (RdRP) of influenza is an attractive target for new drug development as it plays an essential role in viral transcription and replication and is highly conserved among all three major influenza types. Furthermore, it is a highly specific target as no homologous protein has been found in human cells. Drug development targeting the RdRP has recently gained momentum due to the availability of structural and biochemical information.

3.3 RNA-dependent RNA Polymerase

Influenza viral replication relies heavily on the transcription activity of the viral RdRP, a heterotrimer consisting of the polymerase acidic (PA), polymerase basic 1 (PB1), polymerase basic 2 (PB2) subunits (Krug et al., 1989; Honda et al., 1987; Huang et al., 1990). Transcription of the viral RNA genome is dependent on the ability of RdRP to acquire 5'-methylated cap from host cell's pre-mRNA. This mechanism, commonly referred to as cap snatching, entails the binding of the 5'-capped pre-mRNAs from the host cell to the PB2 subunit of RdRP followed by the subsequent cleavage 10-13 nucleotides downstream by the endonuclease activity (Figure 24) (Beaton et al., 1981; Plotch et al., 1981). The cleaved 5'-capped pre-mRNA is then used as a primer to initiate viral transcription by the polymerase activity of the PB1 subunit (Plotch et al., 1981). The high degree of sequence conservation of RdRP among different types of influenza viruses

makes RdRP an attractive target for novel anti-influenza drug therapy. Specifically, inhibition of the either the cap-binding, endonuclease, or polymerase activities of RdRP is actively being pursued (Das et al, 2009; Krug and Aramini, 2009). Inhibition of the endonuclease activity of RdRP, in particular, is the focus of this study.

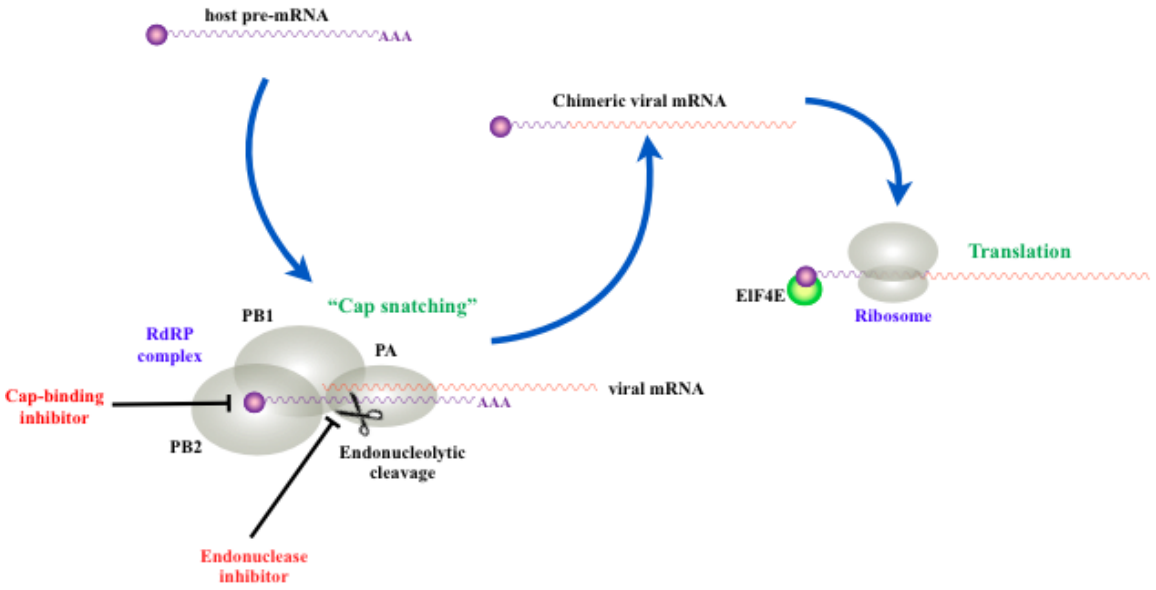


Figure 24: Schematic diagram demonstrating the cap-snatching mechanism of influenza RdRP and potential sites of inhibition.

3.4 Influenza Endonuclease

Endonuclease activity of RdRP was found to reside within the N-terminal domain of the PA subunit (PA_N) based on biochemical and structural studies. Crystal structures of the first 200 residues of PA_N from H5N1 (Yuan et al., 2009), H3N2 (Dias et al., 2009), and H1N1 (Kowalinski et al., 2012) revealed a domain fold and active site arrangement similar to the PD-(D/E)XK family of nucleases (Figure 25). Hydrolysis of ssRNA and ssDNa substrates by PA_N was found to be similar to that of ribonucleoprotein (RNP) (Dias et al., 2009), thus validating the observed endonuclease activity of RdRp to PA_N.

Furthermore, structure-based mutagenesis of key active site residues revealed that the endonuclease activity resides within PA_N (Yuan et al., 2009).

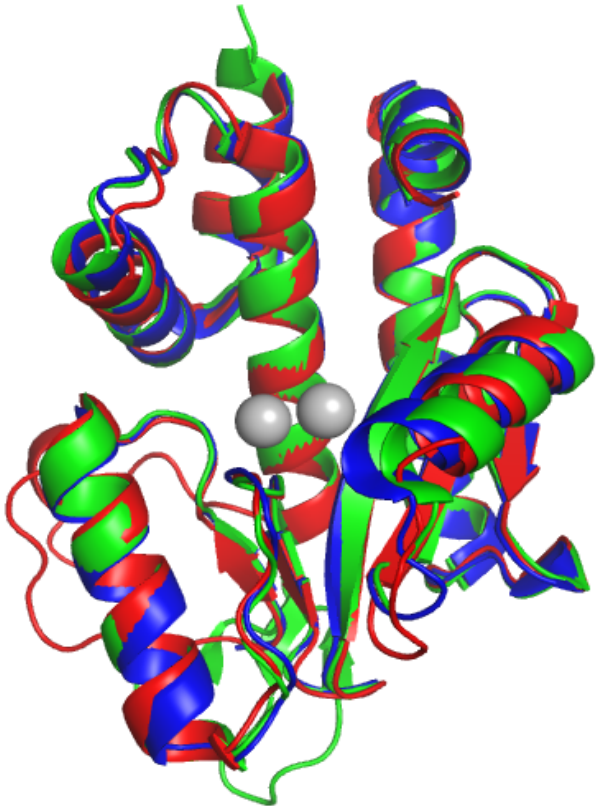


Figure 25: Superposition of PA_N crystal structures from H5N1 (blue; PDB code: 3EBJ), H3N2 (red PDB code 2W69), and H1N1 (green; PDB code: 4AVQ). Location of the active site metal ions is depicted as gray spheres. Unless otherwise noted all figures were prepared using MacPyMol (Schrödinger, New York, NY).

The first reported crystal structures for H3N2 and H5N1 PA_N failed to fully characterize the metal coordination within the endonuclease active site. The crystal structure of H5N1 PA_N revealed a single Mg^{2+} ion bound within the active site whereas two Mn^{2+} ions were seen within the active site of H3N2 PA_N. Initial biochemical studies using RNP found endonuclease activity was enhanced by divalent cations whereas no change in activity was observed in the presence of trivalent ions (Doan et al., 1999). Of the divalent cations assayed, manganese and cobalt greatly enhanced endonuclease activity of RNP while magnesium, zinc, and nickel ions were less efficient. A similar

trend was also observed when a thermal stability assay was used to study the effect of divalent cations on the stability of H3N2 PA_N (Dias et al., 2009). A two metal binding model for both Mn^{2+} and Mg^{2+}, was proposed based on ITC results from the titration of PA_N with varying concentration of metal ions (Crepin et al., 2010). Here, a higher binding affinity was observed for Mn^{2+} compared to Mg^{2+}, which could account for the efficient enhancement endonuclease activity of RNP of Mn^{2+} (Doan et al., 1999). The presence of two metal ions within the endonuclease active sites in crystal structures of H5N1 (Dubois et al., 2012) and H1N1 (Kowalinski et al., 2012) PA_N with nucleotides and inhibitors further confirmed the two metal binding mode of PA_N.

3.5 Endonuclease Inhibitors

The first generation endonuclease inhibitors, which include diketoacids (Tomassini et al., 1994), N-hydroxamic acid (Cianci et al., 1996), N-hydroxy imides (Cianci et al., 1996), and green tea catechins (Kuzuhara et al., 2009), were predominantly discovered through biological screening assays. Scientists at Merck were the first to demonstrate that diketo acids specifically inhibited the endonuclease activity without affecting elongation of mRNA (Tomassini et al., 1994). Structure activity relationships (SAR) of compounds in this study revealed that the diketo acid functionality was necessary for inhibitory activity while lipophilic groups enhanced the potency of the original hit. Coordination to two metals residing in the endonuclease active site was believed to be a central contributing factor to the inhibitory activity of the diketo acids. This notion was supported by the inhibition of HIV-IN and the RNase H activity of HIV-1 RT by diketo acids.

Flutimide, a metabolite from the fungus *Delitschia confertaspora*, and its derivatives were found to specifically inhibit endonuclease activity and more importantly replication of influenza A and B viruses in MDCK cells (Tomassini et al., 1996). SAR of this series revealed the importance of the N-hydroxy group for inhibitory activity, presumably due to its ability chelate metal ions. A similar correlation between metal-chelating motifs and inhibition of endonuclease activity was also observed in the N-hydroxy imide and N-hydroxamic acid series. A pharmacophore model of three oxygen atoms coordinating to two metal ions within the endonuclease active site was proposed based on the SAR of the N-hydroxy imide series (Parkes et al., 2003). Further development of these initial hits was hindered by the lack of structural information about the endonuclease active site.

Structural and biochemical characterization of the PA_N active site (Dias et al., 2009; Yuan et al., 2009) has renewed interest in the development of endonuclease inhibitors. Recently, several new classes of endonuclease inhibitors have been reported. These include green tea catechins (Kuzuhara et al., 2009), phenylethylphenylphthalimides (Iwai et al., 2010), marchatins (Iwai et al., 2011), and carbazoles (Yamada et al., 2012). Crystal structures of different variants of PA_N with ligands revealed that extension of the active sight inhibitors can be promising for lead development due to the presence of several subpockets nearby, as highlighted in Figure 26.

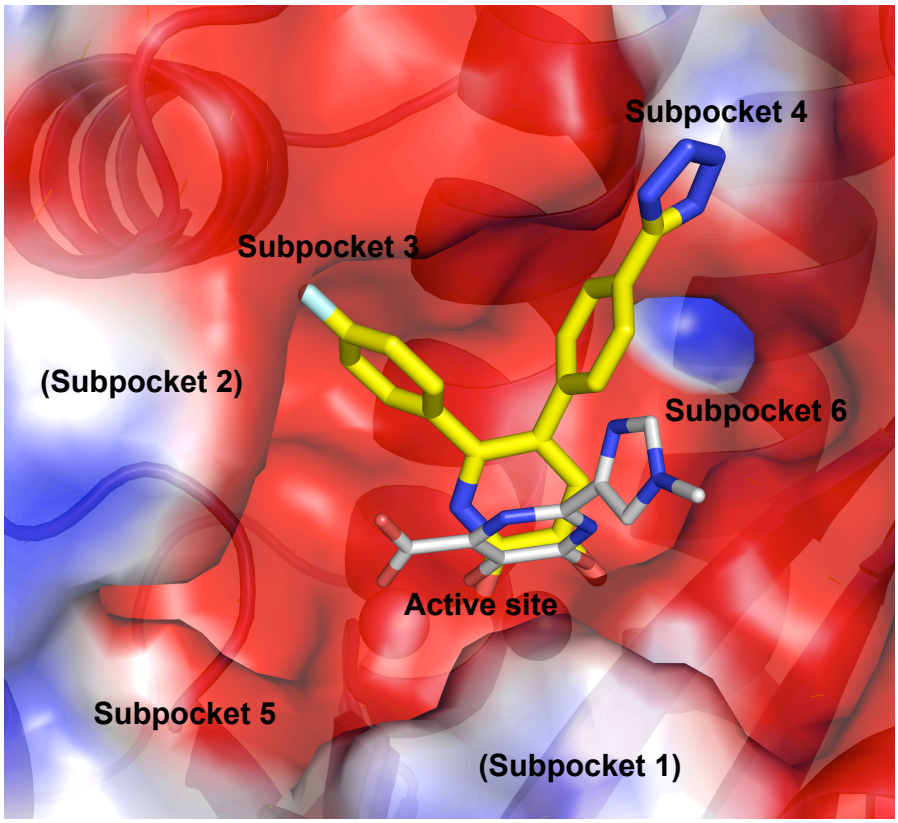


Figure 26: APBS-calculated electrostatic potential surface of the active site cleft of PA_N **51**, developed from a fragment hit described in this chapter, and a carboxamide derivative with three chelating groups crystallized by Dubois et al. (2012) shown in gray (PDB code: 4E5I). The subpockets are labeled as previously described by Dubois et al. (2012).

3.6 Fragment screening by X-ray crystallography

A small molecule fragment-screening campaign using X-ray crystallography was conducted to identify promising class of endonuclease inhibitors (PA_N). A high-resolution crystal form of pH1N1 PA_N was developed for HTS crystallographic screening. A total of eight fragments were identified from screening of the 775 compounds from our fragment library, giving a hit rate of 1%. Inhibitory activity of the identified hits was determined using a fluorescence-based enzymatic assay.

Fragments of interest will be described in detail. A SBDD platform relying on structural biology, enzymatic assays, computational docking, and chemical synthesis was also developed to enable rapid lead development of one particular fragment hit, 3-

hydroxypyridinone, into a compound with antiviral activity in cellular assays. Fragment binding also revealed the presence of a third metal ion within the active site. This was further explored to determine whether the binding of the third metal ion can occur naturally or is it solely due to fragment binding.

3.6.1 Construct Engineering

The initial crystal forms for influenza endonuclease were not suitable for fragment screening due to difficulties reproducing crystals and an occluded active site (Dubois et al., 2012). To search for an appropriate crystal form, the H5N1 endonuclease construct (obtained from Zhiyong Lou and Zihe Rao) was mutagenized and verified to be identical to the sequence of the 2009 pH1N1 PA_N (A/California/07/2009) and various C-terminal truncations were made to obtain proteolytically stable pH1N1 PA_N fragments for crystallization trials. A construct with an HRV14 3C protease cleavable N-terminal dual-hexahistidine tag truncated at residue 204, gave the maximum yield among the tested constructs and readily crystallized in multiple conditions. A crystallization condition with PEG 8000 as the precipitant in the presence of 5 mM Mg^{2+} and 1 mM Mn^{2+} produced crystals that reproducibly diffracted X-rays to 1.7 Å resolution. These crystals were robust and could be soaked into a solution containing fragments at a concentration of 10 mM and 10% (v/v) DMSO. The endonuclease active site was not occluded by crystal contacts, thus allowing for binding of small molecule inhibitors to PA_N. Superposition of this construct and 2009 pH1N1 PA_N revealed a backbone root-mean-square deviation (RMSD) of only 0.24 Å.

3.6.2 Fragment Hits

3.6.2.1 Laudanosoline

Strong electron density was observed for the binding of laudanosoline (**10**), a precursor for the synthesis of reticuline (McDonald, 1981), to the endonuclease active site. The laudanosoline solution screened was a racemic mixture of D and L isomers, however, the 1.60 Å crystal structure clearly indicates that only the D-isomer binds to PA_N.

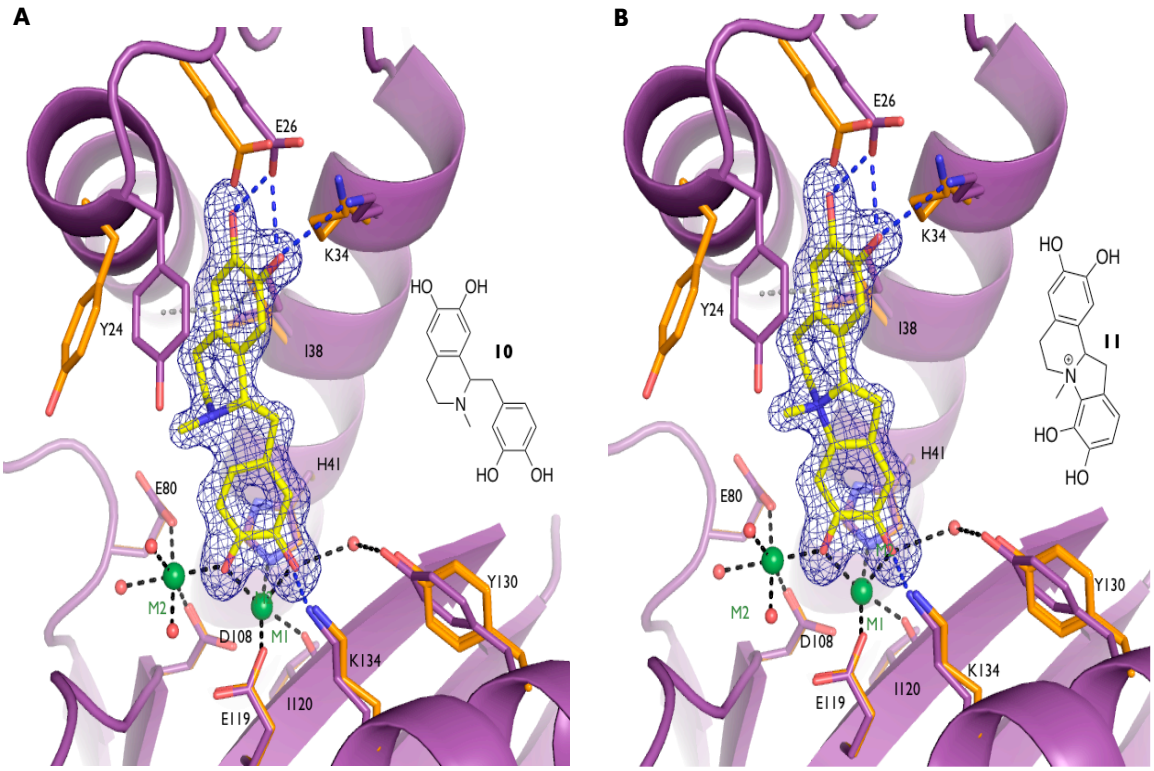


Figure 27: Crystal structure of pH1N1 PA_N soaked with racemic mixture of laudanosoline (purple). **A.)** Electron density with of bound fragment with laudanosoline (10) modeled. **B.)** Electron density of bound fragment with **11**, the product from the oxidation of **10**. Metal-coordinating bonds are depicted as black dashed lines, hydrophobic interactions as gray dashed lines, and hydrogen bonds as blue dashed lines. Residues in orange represent the conformation in the apo structure.

As shown in Figure 27A, the catechol of laudanosoline chelates to the two metal ions (M1 and M2) while coordinating to His41, Asp108, Glu119, and Lys134. The tetrahydroquinoline ring forms a hydrogen bond with the carboxylate of the Glu26 side chain. It also forms strong π-stacking interactions with Tyr24, a highly flexible residue

that forms strong hydrophobic interactions with nucleotides. An unusual and unacceptably short contact between the nitrogen of the tetrahydroquinoline ring and hydrogen from the carbon at the third position of the 4-methylbenzene-1,2-diol was observed, leading to questions about the identity of the bound ligand. In fact, the electron density within this region is more consistent with the bound molecule not being laudanosoline but rather an oxidized form of laudanosoline, shown in Figure 28 (McDonald, 1951). As shown in Figure 27B, this molecule fits the observed electron density without any steric strain. The refinement statistics for this dataset is described in the Materials and Methods.

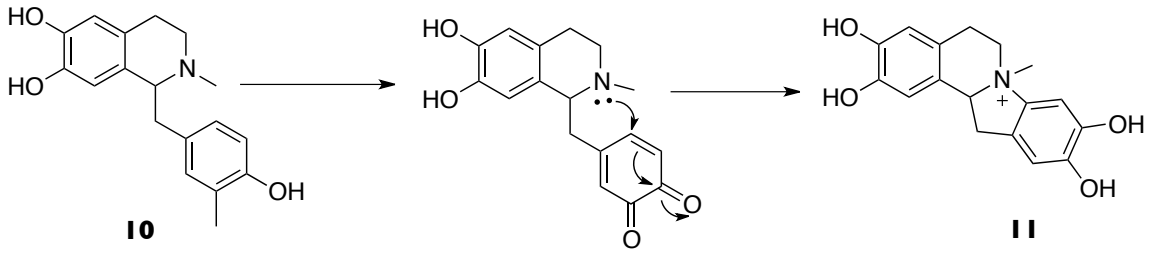


Figure 28: Chemical mechanism depicting the oxidation of laudanosoline (adapted from McDonald, 1951).

The laudanosoline solution was found to have an IC_{50} value of 2 μM in the enzymatic assay giving it a ligand efficiency of 0.33 kcal/mol•NHA. However, since this solution nominally contained a mixture of D and L isomers and the electron density indicates the bound molecule is likely to be the oxidized form of laundanosoline, a higher inhibitory potency is likely to be observed for the purified compound. Currently, work is being pursued to confirm the chemical identity of the bound compound and determine its inhibitory activity.

3.6.2.2 Non-Specific Fragment Binding

Two copies of one fragment, **12**, binds to the active site of PA_N (Figure 29). Although the electron density for both copies of **12** is not fully complete, it is clear that one molecule of **12** coordinates to the active site metal ions through its diketo-like functionality. The branched electron density for the second molecule of **12** indicates that this molecule is flipped compared to the first molecule of **12**. This further validated by contouring the electron density.

Binding of this dimer is accommodated without a significant change to active site. Compared to the apo structure, the side chains of Tyr24, Glu26, Lys34, Phe46, Glu80, and Phe105 slightly shift to expand the pocket for the dimer. Of these residues, significant change was observed for the side chain of Glu80, which shifted 2.9 Å with a ~99° rotation of the carboxylate. The aromatic ring of Tyr24, a residue involved in nucleotide binding, also rotated ~18° to shift 0.5 Å away to accommodate dimer binding.

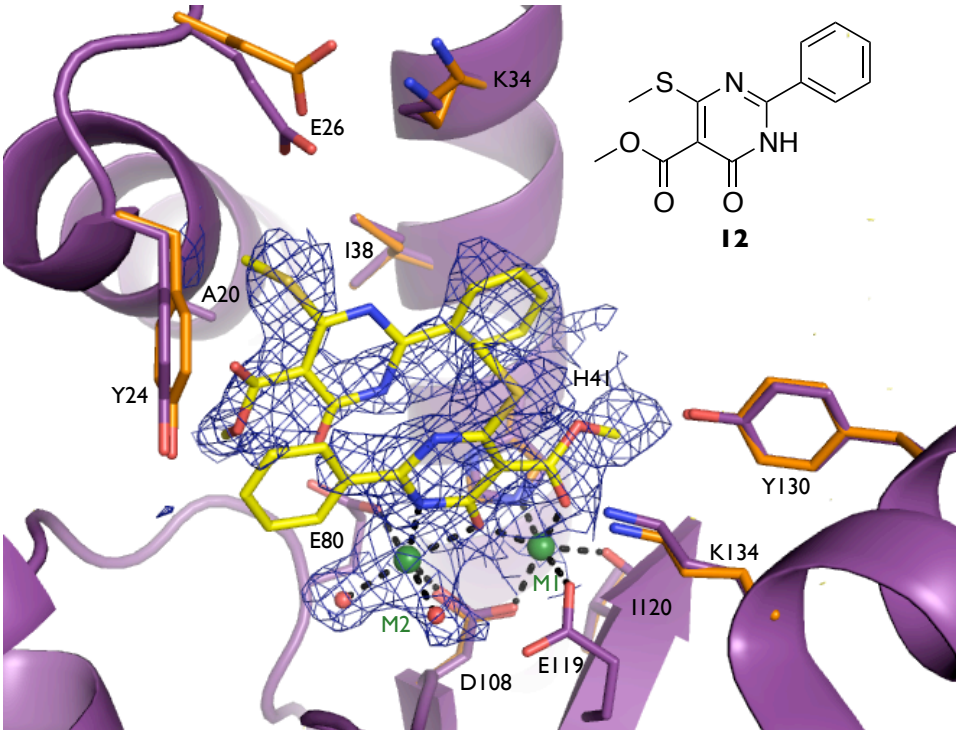


Figure 29: Crystal structure of **12** bound to pH1N1 PA_N (purple). Metal-coordinating bonds are depicted as black dashed lines. Residues in orange represent the conformation in the apo structure.

Binding of **12** as a dimer creates a network of large hydrophobic interactions within the PA_N active site. The metal chelating **12** molecule forms hydrophobic interactions with side chains of Phe105, Leu106, Lys137, Asp108, Glu119, Try130, and Lys134. This molecule, in turn, forms strong stacking interaction with the second, non-chelating molecule of **12**, which also participates in a hydrophobic-interaction network with Ala20, Tyr24, Glu26, Lys34, Ile38, and His41.

The ketone of the pyrimidine of **12** chelates to M1 and M2 and coordinates with the side chains of carboxylates of Asp108 and Glu119. It also participates in a hydrogen-bond network formed by the side-chain oxygen of Glu80 and the adjacent nitrogen of the pyrimidine ring, and the coordinating water of M2. The carbonyl from the ester substituent of this molecule chelates with M1 and forms hydrogen bonds with the side chains of Glu119 and Lys134 as well as the backbone carbonyl of Ile120. Binding of the non-chelating copy of **12** is stabilized by the carbonyl from the pyrimidine ring, which forms bidentate hydrogen bonds to the carboxylate of Glu80 and a water molecule. Two additional hydrogen bonds are formed between the oxygen of the ester functionality and the surrounding water molecules.

3.6.2.3 Fragment Binding Away from the Active Site

Fragment binding was not only limited to the endonuclease active site. A fragment, **13**, bound to a tunnel (subpocket 2) located near the endonuclease active site (Figure 30). The side chains of Glu23 and Arg84 form this tunnel with residues Leu16, Lys19, Ala20, Gly81, and Asp83 lining this binding pocket. Fragment binding is predominantly stabilized by hydrophobic-interaction network formed by the side chains of Leu16, Leu42, and Ala20 as well as the backbone carbonyls of Gly81 and Arg82.

Although no inhibitory activity was seen for **13** in the enzymatic assay, it can be used as a handle for extending compounds bound to the active site into subpocket 2.

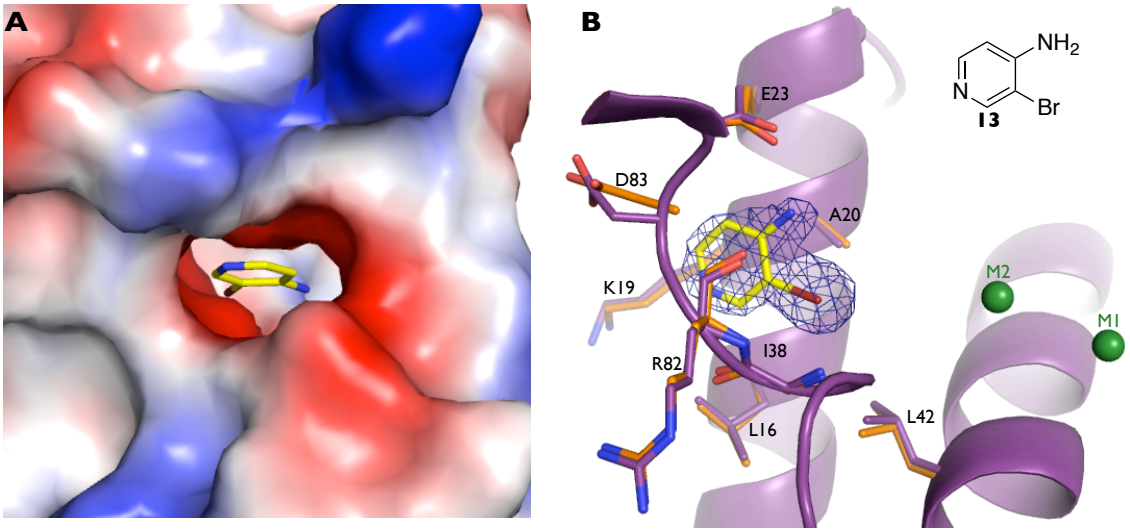


Figure 30: Crystal structure of **13** bound to pH1N1 PA_N (purple). (**A.**) Surface depiction of 12 bound to a tunnel within pH1N1 PA_N. (**B**.) Binding pocket of **13**. Residues in orange represent the conformation in the apo structure. Metal ions are represented as green spheres.

3.6.2.4 Fragment Screening and Third Metal Binding

Binding of **14** to subpockets 2 and 3 within PA_N led to the identification of a third metal binding at the active site (Figure 31). Subpocket 2, as described previously, consists of Leu16, Lys19, Ala20, Glu23, Gly81, and Asp83 whereas subpocket 3 is formed by Tyr24, Glu26, Lys34, and Ile38. The imidazole moiety of **14** stabilizes a previously undetected third metal ion (M3), which was found to be manganese based on anomalous diffraction at a resolution of 1.60 Å, close to the Mn K absorption edge (λ = 1.89 Å). The side chain of Glu80, which normally coordinates to M2 in the apo structure, rotates 90° to chelate to M3. Loss of coordination results in weaker electron density for M2 and resulting in a higher B-factor value of 98.3 $Å^2$ for M2 compared to 27.7 $Å^2$ for M1 and 30.3 $Å^2$ for M3.

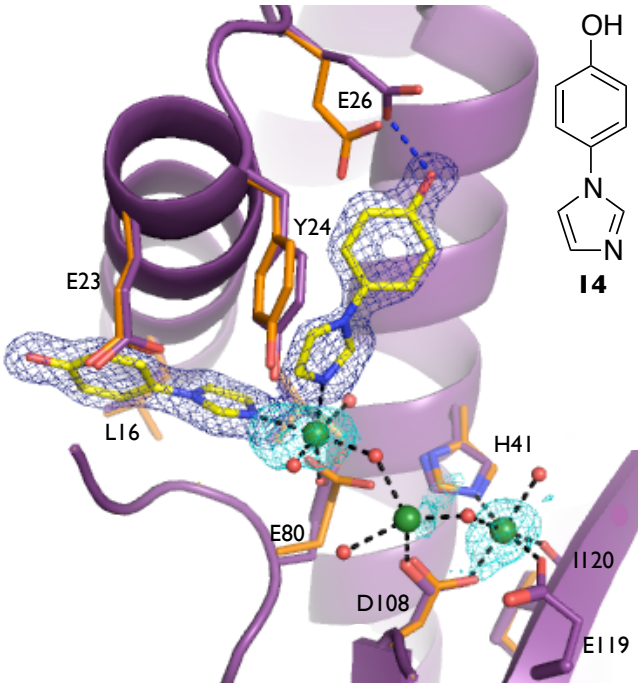
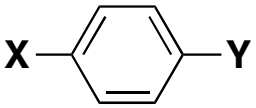


Figure 31: Crystal structure of **14** to pH1N1 PA_N (purple). Metal-coordinating bonds are depicted as black dashed lines and hydrogen bonds as blue dashed lines. Residues in orange represent the conformation in the apo structure.

The, inhibitory activity of **14** was found to be 1 mM, giving a ligand efficiency of 0.34 kcal/mol•NHA (non-hydrogen atom). Fragments with similar chemical scaffolds to **14** were individually assayed and soaked into crystals of PA_N. As shown in Table 2, substitution of the hydroxyl at the para position of the phenyl ring with an amine or methylamine causes a 10-fold improvement in inhibitory activity. An improvement in enzymatic inhibition is also observed upon the substitution of the imidazole ring with either a thienyl or furanyl rings. Soaking of these fragments into crystals of PA_N, however, was not successful thus the exact binding mode of these molecules is not known. It is possible that these molecules bind in a different manner than **14** however coordination to M3 would be lost due to the substitution of the heteroatom at the 3-position of the five-membered ring with a carbon.



	X	Y	IC_{50} (μM)	**Ligand efficiency** (kcal/mol•NHA)
14	OH	1H-imidazole	1000	0.34
15	CH_2OH	1H-pyrazole	1000	0.31
16	NH_2	4-thiadiazole	500	0.38
17	NH_2	2-thiophene	200	0.39
18	CH_2NH_2	2-furan	150	0.40
19	CH_2NHCH_3	1H-imidazole	100	0.39
20	NH_2	1H-pyrrole	100	0.42

Table 2: Endonuclease inhibitory activity for analogs of **14**.

3.6.3 Third Metal Ion Binding

Crystals of endonuclease were titrated with metal ions, specifically Mn^{2+} and Ca^{2+}, to ensure that binding of third metal ion is not just an artifact of fragment binding. A 200 μM Mn^{+2} soak revealed electron density for a third metal and its coordinating waters within the apo structure. Electron density for the third metal within the apo structure further strengthened upon increasing the manganese concentration 10 mM. No loss in coordination to M2 was observed due to the binding of M3 within the apo structure as apparent by B-factor values of 21.2 $Å^2$ for Mn1, 22.4 $Å^2$ for Mn2, and 119.5 $Å^2$ for Mn3 from the 10 mM manganese soak. No anomalous signal was detected for M3, which may be due to the high B-factor. Alternatively, it is possible that M3 is in fact a water molecule.

Soaking of PA_N crystals with a 100 mM solution of Ca^{2+} revealed the binding of M2 and M3 but not M1 (Figure 32). Bipyramidal pentavalent coordination to residues and coordinating waters confirmed the bound metals were calcium and not manganese or magnesium, which could be carried over from crystallization. The loss of M1 binding could potentially account for loss of endonuclease activity reported by Dias et al. (2009).

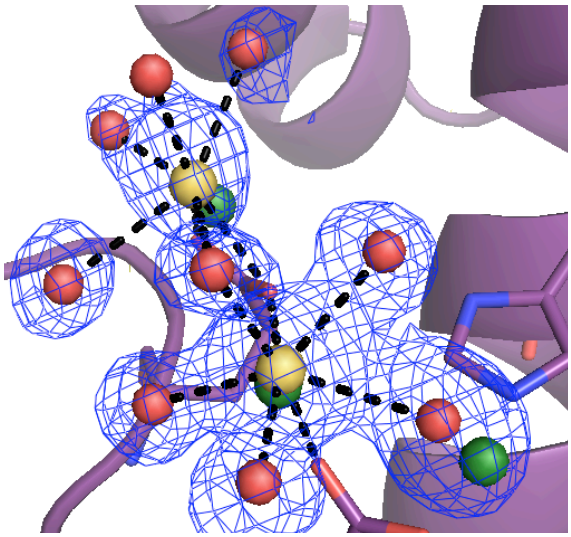


Figure 32: The active site cleft in the apo PA_N structure from a 100 mM $CaCl_2$ soak with electron density for the metal ions and coordinating waters. The locations of the metal ions from the soak of **14** are shown in green whereas the locations of the calcium cations from a 100 mM soak are shown in yellow. The electron density for the metal and coordinating waters is shown (omit map contour of 2.5σ).

The three-metal ion coordination observed for PA_N has also been reported for other endonucleases, including T5 flap endonuclease, endonuclease IV and EcoRV (Syson 2008; Ivanov 2007; Sam 1999). A three-metal mechanism for EcoRV catalysis has been proposed based on a combination of structural and molecular modeling data. Briefly, in this model, metal at site I is directly involved in catalysis whereas metal at site II provides structural support and metal at site III is believed to stabilize the negative charge of the transition state and recruit a water to protonate the leaving group (Kovall et al., 1999, Horton et al., 1998, Horton et al., 2004). Although this may not necessarily be the case with respect to influenza endonuclease, the presence of the third metal binding within active site suggests that further studies are required.

3.6.4 Development of the Hydroxypyridinone Class of Endonuclease Inhibitors

Crystallographic screening revealed ambiguous electron density for 5-chloro-3-hydroxypyridin-2(1H)-one bound to the active site of PA_N. A brominated derivative (**21)**

was used to clarify the exact binding mode of the fragment, anomalous X-ray diffraction at a resolution of 1.85 Å, close to the Br K absorption edge (λ = 0.92 Å). Soaking of this brominated compound revealed binding to not only to the active site of PA_N but also subpockets 2 and 3. The non-specific fragment binding allowed for coordination to all three metals within the active site. Binding of **21** at the active site allowed for coordination with M1 and M2 wherease binding to subpockets 2 and 3 stabilized M3 binding (Figure 33).

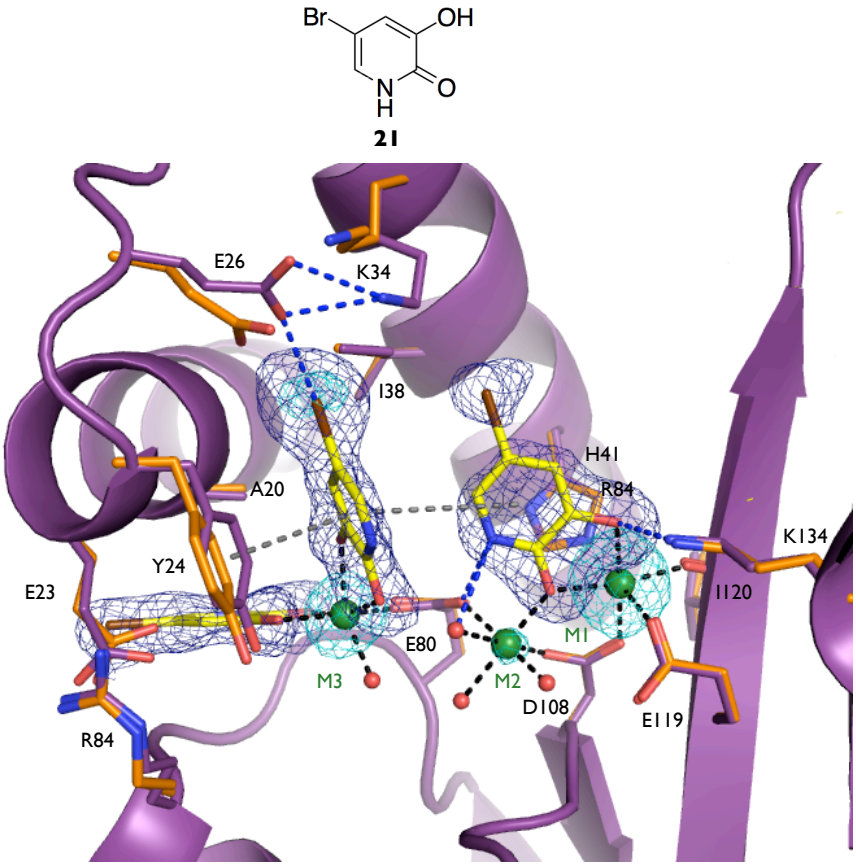


Figure 33: Crystal structure of **21** bound to pH1N1 PA_N (purple). Metal-coordinating bonds, hydrogen bonds, and hydrophobic interactions are depicted as black, blue, and gray dashed lines, respectively. Residues in orange represent the conformation in the apo structure.

Based on anomalous diffraction, M3 was determined to be manganese while M2 is believed to be magnesium since 5 mM magnesium and 0.1 mM manganese is present in

the soaking solution. Interaction between **21** and Glu80 allows for the coordination with M2 and M3. As a result, M3 has a lower B-factor value of 46.1 $Å^2$ compared to that of 98.3 $Å^2$ for M2 upon binding of **14**. In addition, the low B-factor values of 29.7 $Å^2$ for M1 and 48.9 $Å^2$ for M3 indicate that all three metals are bound to the active site.

21, when bound to subpockets 2 and 3 forms an interaction network between the hydroxyls of **21**, M3, and a coordinating water molecule. This includes a chelating interaction between M3 and the carboxylate of Glu80, which also coordinates to M2. Non-optimal hydrophobic interactions consisting of π-stacking between the side chains of Arg84 and Glu23, which interacts weakly with the bromine of **21** bound to subpocket 2. Hydrophobic interactions are also seen between **21** bound to subpocket 3 with side chains of Lys34 and Ile38. A halogen bond with a distance of 2.9 Å is formed between the carboxylate of Glu26 and the bromine of **21** bound to subpocket 3.

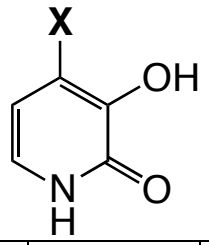
At the active site, **21** not only coordinates with M1 and M2 but also forms a hydrogen bond with Lys134 at a distance of 2.7 Å. Unlike previously published endonuclease inhibitors that form four coordinating interactions with the active site metals, this molecule of **21** only forms three—two with M1 and one with M2. A hydrogen bond between the nitrogen of the hydroxypyridinone ring with a coordinating water molecule of M2 further stabilizes this interaction network. The nitrogen of the hydroxypyridinone also forms a hydrogen bond at a distance of 2.8 Å with the side chain oxygen of Glu80. Binding of **21** at a 50° angle enables favorable π-stacking interactions with His41, a conserved residue. This type of interaction has not been seen within crystal structures of endonuclease inhibitors bound to PA_N (Dubois et al., 2012, Kowalinski et al.,

2012). The bromine of **21** also forms weak hydrophobic interactions with the side chains of Lys34 and Ile38, respectively.

A ligand efficiency of 0.73 kcal/mol•NHA was determined based on an IC_{50} value of 16 μM for **21** and assuming that the observed inhibition is solely due to binding at the active site. A hit-to-lead campaign was conducted based on a combination of favorable inhibitory activity, ligand efficiency, and synthetic tractability of the fragment hit. The availability of structural data allowed for computational docking using AutoDock Vina (Trott and Olson, 2010) for efficient exploration of chemical space. Substituents found to be promising in docking studies were prioritized for synthesis. Chemical space around the 4-, 5-, and 6- positions of the hydroxypyridinone ring scaffold was explored in this manner.

3.6.4.1 Substitution at the 4-Position of Hydroxypyridinone

Although both structural and modeling data suggested that expansion from the 4-position of the hydroxypyridinone ring could be promising, this was not the case experimentally. A dramatic loss in inhibitory activity was observed for small substituents (Table 3). Furthermore, obtaining crystal structures of PA_N bound to the compounds described in Table 3 proved to be challenging. It is possible that substituents at this position can affect the chelation properties or electronics of the hydroxypyridinone ring, resulting in the loss of enzymatic activity.



	X	IC_{50} (μM)	Ligand efficiency (kcal/mol•NHA)
22	-H	35	0.76
23	-CH_3	100	0.61
24	-Br	150	0.58
25	-c-propyl	200	—
26	-CN	>200	—
27	-F	>200	—

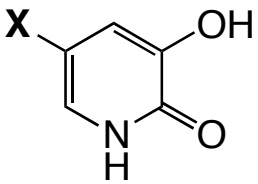
Table 3: Endonuclease inhibitory activity of 4-subsituted hydroxypyridinone compounds.

3.6.4.2 Substitution at the 5-Position of Hydroxypyridinone

Molecular modeling was used to explore potential substitutions at the 5-position of the hydroxypyridinone ring. As predicted by docking studies, the pocket readily accommodated larger substituents, which translated into an improvement in enzymatic potency. Further chemical modifications from the phenyl ring at the 5-position revealed that substituents at the meta- and para- positions in general improved inhibitory activity (Table 4).

Substitution with 3-benzonitrile (**34**) at the 5-position of the hydroxypyridinone ring resulted in a 20-fold improved in inhibitory activity when compared to **21**. A crystal structure revealed that a single molecule of **34** coordinated to all three metal ions in the active site of PA_N (Figure 34). The nitrile of **34** forms hydrophobic interactions with Lys34 and Ile38 while the benzonitrile ring makes similar interactions with Ala24 and Ile38. Comparison to the apo structure revealed that the active site of PA_N is flexible and alters to accommodate the binding of **34**. Specifically, a shift of 2.3 Å for the side chains of Glu26 and Lys34 not only enables the formation of a new salt bridge but also causes

subpocket 3 to expand. A coordinating water molecule of M3 stabilizes Glu23, a residue of subpocket 2, through a hydrogen bond interaction.



	X	IC_{50} (μM)	**Ligand efficiency** (kcal/mol•NHA)
28	H	35	0.76
29	$-CH_3$	8.6	0.77
30	-Br	5.0	0.81
31	c-hexane	2.0	0.56
32	-phenyl	2.1	0.56
33	3-fluorophenyl	0.74	0.56
34	3-benzonitrile	0.73	0.53
35	3-benzoic acid	1.79	0.46
36	4-fluorophenyl	0.73	0.60
37	4-benzonitrile	3.80	0.50
38	4-benzoic acid	0.51	0.54

Table 4: Endonuclease inhibitory activity of 5-subsituted hydroxypyridinone compounds.

Binding of **34** significantly alters the active site region containing Tyr24, an essential residue for nucleotide binding (Zhao et al., 2009, Kowalinski et al., 2012). When compared to the apo structure, a 0.95 Å backbone shift is observed for Tyr24. This causes the side chain of Tyr24 to shift 4 Å shift and rotate downwards by 45º, resulting in a chelating interaction between the hydroxyl of Tyr24 and M3. This orientation of the aromatic ring of Tyr24 also allows for weak hydrophobic interaction with the nitrile moiety of **34**. Since crystal structures indicate that Tyr24 is highly flexible and essential for nucleotide binding (Kowalinski et al., 2012), the ability of **33** to induce a change within the active site of PA_N that enables favorable interactions with Tyr24 is promising. Furthermore, strengthening this interaction can greatly improve the potency of the molecules in this series.

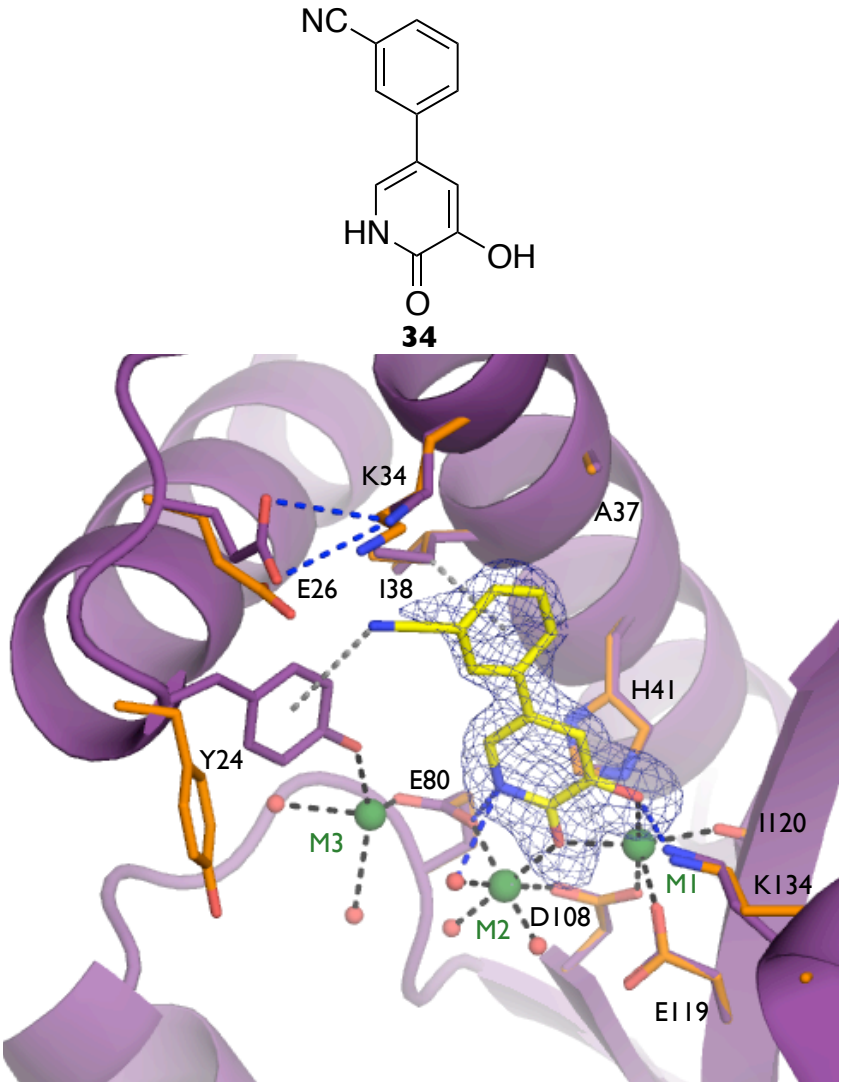


Figure 34: Crystal structure of **34** bound to pH1N1 PA_N (purple). Metal-coordinating bonds, hydrogen bonds, and hydrophobic interactions are depicted as black, blue, and gray dashed lines, respectively. Residues in orange represent the conformation in the apo structure.

3.6.4.3 Substitution at the 6-Position of Hydroxypyridinone

When compared to **21**, a 40-fold improvement in inhibitory activity was noted upon substitution of a phenyl at the 6-position of the hydroxypyridinone ring (**39**). Crystal structure revealed that three molecules of **39** chelated the three metal ions in a manner similar to **21** (Figure 35). As with **34**, a 2.4 Å shift between the side chains of Glu26 and Lys334 resulted in the formation of a salt bridge. The hydrophobic interaction

network previously seen with **34** is further strengthened by favorable contacts between the side chains of Glu26, Lys34, and Ile38 and the phenyl ring of **39** bound to subpocket 3. Additionally, **39** bound to subpocket 3 forms favorable π interactions with Tyr24 as well as with **39** bound at the active site. The hydroxypyridinone ring for **39** bound at the active site flips 180°, resulting in the loss of a hydrogen bond between the nitrogen of the hydroxypyridinone ring and a coordinating water molecule of M2.

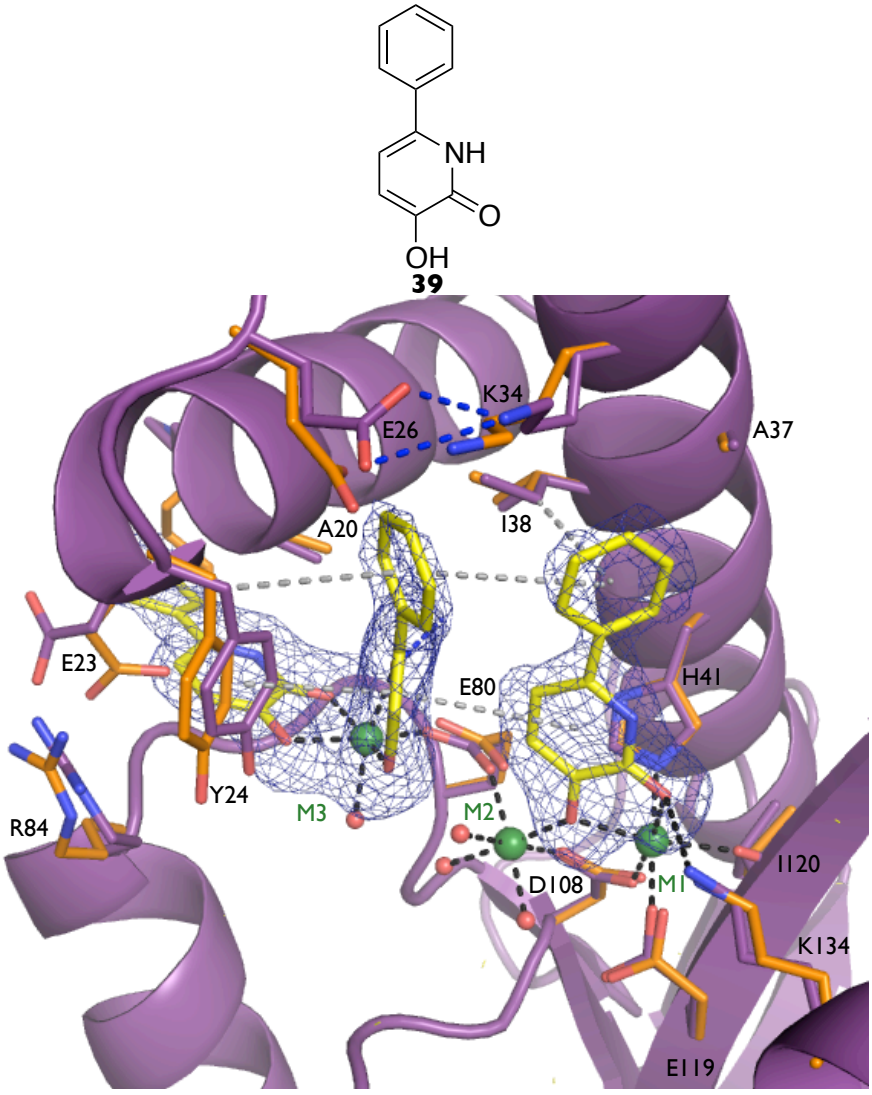
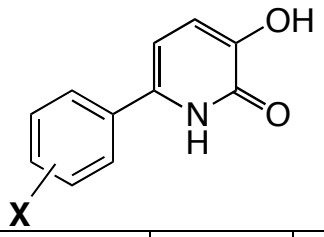


Figure 35: Crystal structure of **39** bound to pH1N1 PA_N (purple). Metal-coordinating bonds, hydrogen bonds, and hydrophobic interactions are depicted as black, blue, and gray dashed lines, respectively. Residues in orange represent the conformation in the apo structure.

Extension of the phenyl ring from the 6-position of the pyridinone ring was further explored to further enhance potency and binding. No significant differences were observed for substituents at the 3- or 4- positions of the phenyl ring (Table 5). However, SAR did reveal that electron withdrawing groups at these positions resulted in a dramatic loss in potency. Overall, fluorine was found to improve potency at both the 3- and 4-positions with the 4-fluoro having the highest potency in this series.



	X	**IC_{50} (μM)**	**Ligand efficiency** (kcal/mol•NHA)
39	H	0.87	0.59
40	3-F	0.53	0.57
41	3-amide	>200	—
42	3-OCH_3	0.54	0.53
43	4-F	0.29	0.60
44	4-CN	0.80	0.52
45	4-tetrazole	0.85	0.43
46	4-ethoxy carbonyl	>200	—

Table 5: Endonuclease inhibitory activity of hydroxypyridinone derivatives substituted with phenyl at the 6-position.

The strongest electron density was observed when the phenyl ring was substituted with 3-methoxy (**42**), which only bound to the endonuclease active site in the presence of all three metal ions (Figure 36). Binding of **42** was stabilized by chelating interactions between the hydroxyls of **42**, M1, and M2 together with a cation-π interaction between M3 and the anisole ring of **42**. Unlike **34**, flipping of the hydroxypyridinone ring was not observed for binding of **42**. Thus, the hydrogen bond between the nitrogen of the hydroxypyridinone and a coordinating water molecule of M2 is maintained. The oxygen of the methoxy substituent forms a hydrogen bond with the side chain of Lys34, which

also forms a salt bridge with Glu26. A hydrophobic interaction network is formed by the anisole ring of **42** and the side chains of Ala20, Tyr24, and Ile38.

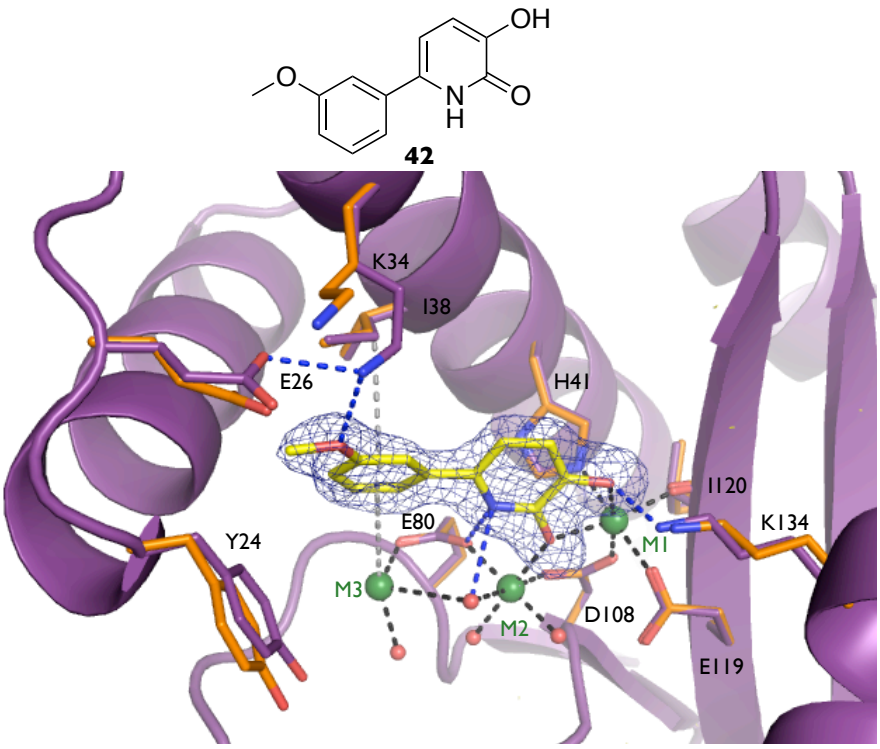
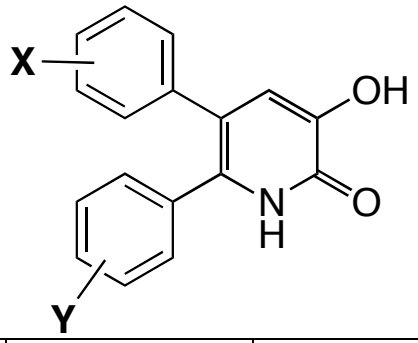


Figure 36: Crystal structure of **42** bound to pH1N1 PA_N (purple). Metal-coordinating bonds, hydrogen bonds, and hydrophobic interactions are depicted as black, blue, and gray dashed lines, respectively. Residues in orange represent the conformation in the apo structure.

3.6.5.4 Disubstitution at the 5, 6-Positions of Hydroxypyridinone

Based on structural information and the SAR developed for substitutions at the 5- and 6-positions of the hydroxypyridinone ring, disubstitution at these positions was explored to improve potency. As predicted, an improvement in potency was observed upon phenyl substitution at the 5 and 6 position of the hydroxypyridinone ring (Table 6). Significant improvement in potency was observed upon asymmetric substitution of the phenyl ring at the 5 and 6-positions of the hydroxypyridinone scaffold whereas symmetrical substitutions at these positions resulted in a reduction in potency. This was not surprising as the nearby pockets are very different in nature. Crystal structures

indicate that expansion from the 5-position favors moieties with groups capable of forming charged interactions whereas substituents capable of forming hydrophobic interactions are favored at the 6-position.



	X	Y	IC_{50} (μM)	**Ligand efficiency** (kcal/mol•NHA)
47	H	H	0.047	0.43
48	F	F	0.041	0.40
49	CN	CN	0.103	0.40
50	Tetrazole	Tetrazole	0.285	0.30
51	CN	F	0.136	0.41
52	F	Tetrazole	0.023	0.35
53	F	CN	0.054	0.37
54	Tetrazole	F	0.011	0.37

Table 6: Endonuclease inhibitory activity of 4-subsituted phenyl at both 5, 6-positions of the hydroxypyridinone ring.

Substitution with a 4-fluorophenyl at both the 5- and 6-positions (**48**) led to a 400-fold improvement in enzymatic inhibition. Crystal structure of **48** revealed that the 4-fluorophenyl from the 6-position of the hydroxypyridinone ring maintains a cation-π interaction with M3 and hydrophobic interactions with the side chains of Ala20 and Ile38 (Figure 37). The 4-fluorophenyl at this position also forms a weak hydrophobic interaction with Tyr24. Weak electron density for the side chain of Tyr24 indicates that the residue is still flexible and further optimization is required to strengthen this interaction. At the 5-position, the 4-fluorophenyl forms hydrophobic interactions with Lys34, Ala37, and Ile38. An ethylene glycol molecule from crystallization is displaced,

permitting the formation of a halogen bond between Arg124 and the fluorine at the 5-position. This allows for further extension into subpocket 4.

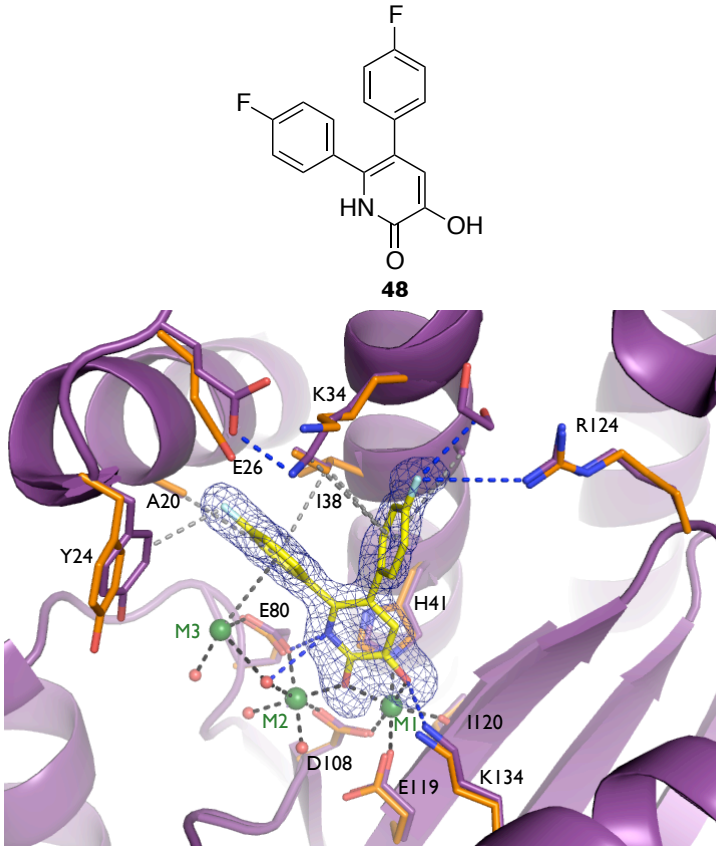


Figure 37: Crystal structure of **48** bound to pH1N1 PA_N (purple). Metal-coordinating bonds, hydrogen bonds, and hydrophobic interactions are depicted as black, blue, and gray dashed lines, respectively. Residues in orange represent the conformation in the apo structure.

Substitution of the 4-fluoro at the 6-position of **48** with a 4-tetrazolyl (**52**) causes a 180° flip of the hydroxypyridinone ring, resulting in the loss of a hydrogen bond between the nitrogen of the hydroxypyridinone ring and a coordinating water molecule of M2. As shown in Figure 38, **52** binds in such a way that the 4-(tetrazol-5-yl) phenyl occupies the pocket previously occupied by the 4-fluorophenyl at the 5-position of **48**. This allows the tetrazole to form bidentate hydrogen bonds with Arg124 and hydrophobic interaction with Lys34.

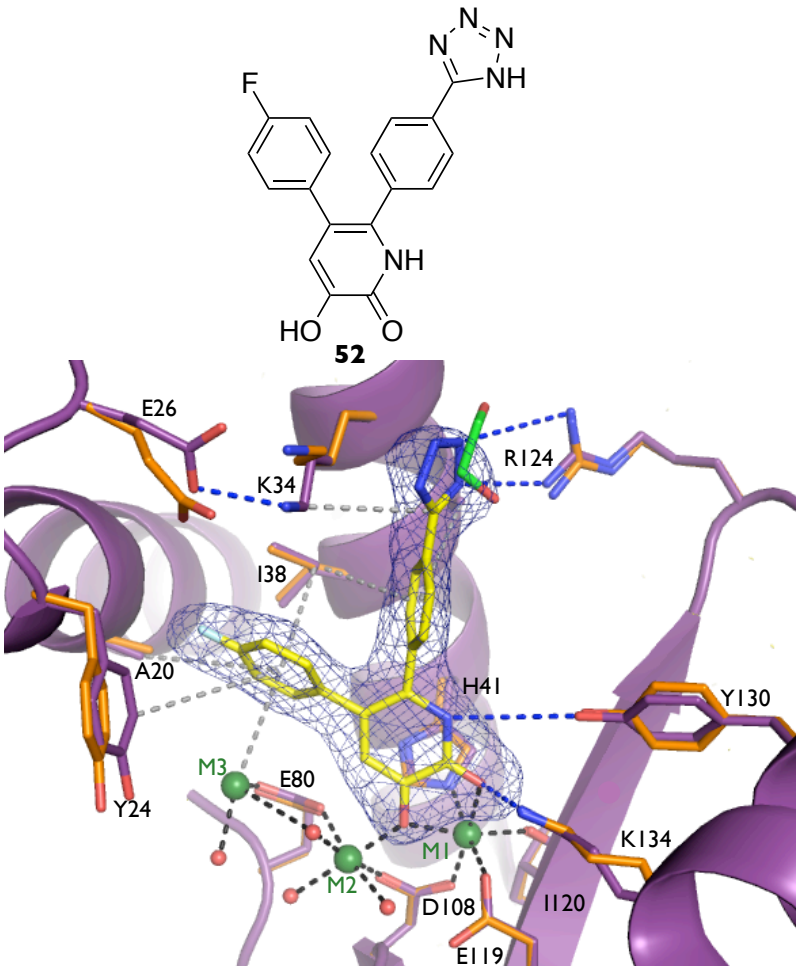


Figure 38: Crystal structure of **52** bound to pH1N1 PA_N (purple). Metal-coordinating bonds, hydrogen bonds, and hydrophobic interactions are depicted as black, blue, and gray dashed lines, respectively. Residues in orange represent the conformation in the apo structure.

Substitution with a 4-(tetrazol-5-yl)phenyl at the 5-position and 4-fluorophenyl at the 6-position (**54**) causes the hydroxypyridinone ring to revert to the standard orientation seen in previous structures of compounds in this series (Figure 39). This results in a 2-fold improvement in potency without any loss of interaction previously observed **52**.

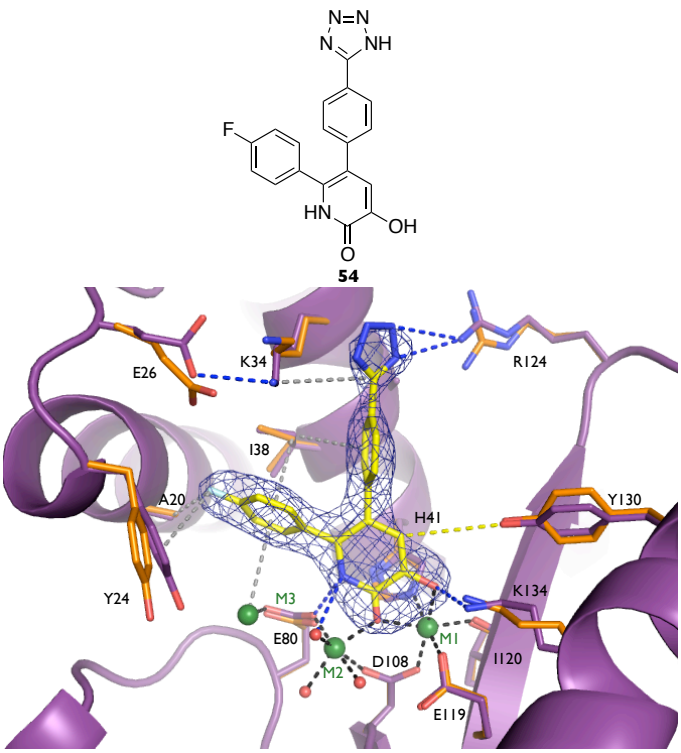


Figure 39: Crystal structure of **54** bound to pH1N1 PA_N (purple). Metal-coordinating bonds, hydrogen bonds, and hydrophobic interactions are depicted as black, blue, and gray dashed lines, respectively. Residues in orange represent the conformation in the apo structure.

The inhibitory activities of **52** and **54** against influenza A/Puerto Rico/8/1934 (H1N1; PR8) in MDCK cells was determined using a virus yield assay (Figure 40). Virus titers from supernatants were determined at 24-hours post infection using a fluorescent forming unit (FFU) assay (Schickli et al., 2001; Martinez-Sobrido et al., 2010; Baker et al., 2013). This was done for the concentrations of **52** and **54** along with DMSO and oseltamivir as the negative and positive controls, respectively. Although **52** did not cause significant reduction in viral growth, **54** was found to have an EC_{50} value of 11.4 µM. No cytotoxicity was detected (CC_{50} > 50 µM) for **54** as measured by MTT assay at 24 hours. Viral assay revealed that DMSO had no effect on viral replication at the tested concentrations whereas oseltamivir was found to have an EC_{50} value of 0.015 µM, which is similar to the previously reported EC_{50} using the PR8 (McSharry et al., 2004).

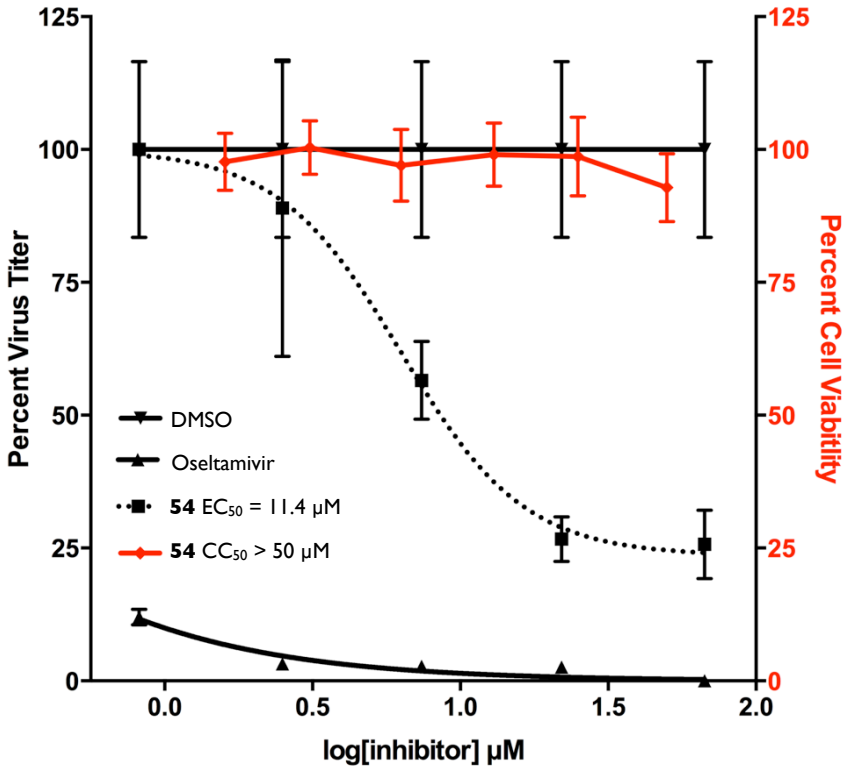


Figure 40: Viral yield inhibition assay and cytotoxicity assay. Virus yield assay using MDCK cells and the A/Puerto Rico/8/1934 (H1N1) influenza strain was used to determine the antiviral activity of **54**; oseltamivir and 0.1% (v/v) DMSO (only) were used as the positive and negative controls. Cytotoxicity for **54** (shown in red) was determined from MDCK cells after 24 hours using the MTT assay.

3.7 Conclusion

Fragment screening by X-ray crystallography using a high-resolution crystal form of pH1N1 PA_N was used for was successfully applied to identify chemical scaffolds that could bind to the endonuclease of influenza RdRP. Screening of 775 compound revealed a total of eight fragment hits, giving a hit rate of 1%. A total of eight fragments were identified from screening of the 775 compounds from our fragment library, giving a hit rate of 1%. Fragment binding was observed at the active site and allosteric pockets surrounding the active site cleft. Inhibitory activity of the identified hits was determined using a fluorescence-based enzymatic assay. Rapid lead development of a promising fragment hit, hydroxypyridinone, into a compound (**54**) with antiviral activity was pursed

using a SBDD platform consisting of structural biology, enzymatic assays, computational docking, chemical synthesis, and virology. This resulted in the development of **54**, which has an antiviral EC_{50} activity of 11 μM in a viral yield inhibition assay using MDCK cells. Fragment binding also revealed the presence of a third metal ion within the active site, which may have a functional role that needs to be explored further.

3.8 Materials and Methods

3.8.1 Plasmid Construction

The H5N1 PA_N sequence was mutagenized to match the 2009 pandemic H1N1 (pH1N1) sequence (PA 1-204). Mutagenesis was performed using overlap extension PCR with sequence and ligation independent cloning (Li and Elledge, 2007) into a pCDF2 vector (Novagen).

3.8.2 Protein Expression and Purification

pH1N1 endonuclease constructs were expressed in BL21 (RIL) cells (Stratagene). The BL21 cells were grown to an OD_{600} of 0.8 in 2xYZ media and induced with 0.15 mM IPTG at 17°C for 17 hours. Cells were harvested by centrifugation and purified on Ni-NTA (Qiagen) according to manufacturer's recommendations except the final elution was performed with 500 mM imidazole present. The dual-hexa-His tag was then removed by HRV14 3C protease cleavage followed by an additional IMAC purification step. Endonuclease was further purified by size-exclusion chromatography using HiLoad 26/60 Superdex 75 (GE Heathcare). The protein was concentrated to 10 mg mL^{-1} using a Ultrafree 10K (Millipore), aliquoted and stored in 100 mM NaCl and 20 mM Tris pH 8.0 at -80 °C.

3.8.3 Crystallography

Endonuclease at 5 mg/mL was mixed in equal volumes of precipitant solution containing 1 mM manganese chloride, 200 mM MES pH 6.7, 25% (w/v) PEG 8000, 100 mM ammonium sulfate, 10 mM magnesium acetate, 10 mM taurine, and 50 mM sodium fluoride. All crystallization was performed at 20°C.

Ligand soaks were performed by step-wise gradient shifting the crystal solution to 1 mM manganese sulfate, 200 mM HEPES pH 7.7, 25% (w/v) PEG 8000, 50 mM ammonium sulfate, 5 mM magnesium acetate, and 10% (v/v) ethylene glycol. 80 mM L-arginine was included to improve the solubility of the ligands. For fragment screening, cocktails of 4-8 compounds were used for the initial screening followed by single compound soaks to verify the bound fragment identity and to rule out cooperative binding. Ligands or fragments, in dimethyl sulfoxide, were soaked into crystals (final concentration 10 mM and 10% (v/v) dimethyl sulfoxide) for 2 hours before flash freezing in liquid nitrogen. X-ray diffraction data collection was performed at the Cornell High Energy Synchrotron Source (CHESS) F1 beamline, and the National Synchrotron Light Source (NSLS) beamline X25.The diffraction data were indexed, processed, scaled and merged using *HKL2000* (Otwinowski et al., 2011). Structure refinement were carried out using PHENIX (1.8.2-1309) (Adams et al., 2010) and COOT (0.7-rev4459) (Emsley et al., 2010) with riding hydrogens present when the X-ray diffraction resolution is better than or equal to 1.9 Å. Current crystallographic and refinement statistics for PA_N crystals soaked with laudanosoline are described in Table 7.

	PA_N-Laudanosoline Soak
X-ray source	CHESS F1
Wavelength (Å)	0.917
Space group	$C222_1$
Cell constants: a, b, c (in Å)	90.62, 101.30, 65.73
α, β, γ (in º)	90.0, 90.0, 90.0
Resolution range (Å)	1.60-50.00
Completeness (%)	99.5%
R_{merge}	0.050
Average I/σ (*I*)	27.1
Refinement Statistics	
No. of reflections	40,221
R_{work}	0.1739
R_{free}	0.1746
Ramachandran statistics (% of residues in favored/disallowed regions)	98.51/0.00
RMSD bond length (Å)	0.012
RMSD bond angles (º)	1.287

Table 7: Crystallographic and refinement statistics of pH1N1 PA_N crystal soaked with laudanosoline.

3.8.4 Endonuclease Activity Assay

The endonuclease assay used a single-stranded DNA probe with the sequence (6-FAM)TGGCAATATCAGCTCCACA(MGBNFQ) (Applied Biosystems). Reaction buffer contained 50 mM Tris pH 7.5, 50 mM sodium chloride, 1 mM DTT, 5 mM magnesium chloride, 0.5 mM manganese sulfate, and 1 mM CHAPS. DNA probe (50 nM), endonuclease (25 nM), and compound were mixed on ice and then placed in a Varioskan Fluorometer (Thermo Scientific) preincubated at 37˚C to detect fluorescence with an excitation of 488 nm and emission of 518 nm. Fluorescence is measured at various time points and activity/inhibition is calculated with GraphPad Prism 6.0b. Ligand efficiency was calculated using the following equation.

$$\text{LE (in kcal/(mol·NHA))} = -RT\ln(IC_{50})/\text{NHA}$$

3.8.5 Modeling and Docking

All of the modeled compounds were geometry optimized using the Universal Force Field (UFF) (Rappe et al., 1992) as implemented in ArgusLab (Pettersen et al.,

2004). Hydrogens were added to the protein structure using Chimera 1.5.2 and further optimized to improve the hydrogen-bonding network and relieve steric clashes. All solvent molecules, excluding chelated water molecules, were removed. The Mn^{2+} ions and chelated water molecules were retained as parts of the receptor during docking simulations. This was done to ensure that the generated poses displayed reasonable coordination geometry at the active site.

The required input files for AutodockVina (Trott and Olson, 2010) were prepared using AutoDockTools (Morris et al., 2009). Gasteiger-Marsili charges were assigned to the ligands and all torsions were treated as being flexible. The receptor file was prepared by merging non-polar hydrogens, adding Kollman United Atom charges, and assigning AD4 atom types. The created pdbqt receptor files were then modified to account for the presence of a metal. A previous study by Chen et al. (2007) revealed the importance of the charge of the metal ions, and its influence on docking. Accordingly, the charges of the metal ions were scaled iteratively, and a docked conformation that best matched the crystal structure was obtained using a charge of 1.52 for each metal. The search space for docking was limited to $(24\ Å)^3$, from the mean coordinates of the bound fragments, so as to traverse the entire active site. The exhaustiveness level was set to 12 to ensure adequate conformational sampling and the weight for the repulsion term was scaled to 0.60. The said docking procedure was validated by self-docking and cross-docking experiments before embarking on fragment elaboration.

Chapter 4: Fragment Screening by X-ray Crystallography of HIV-1 Integrase Catalytic Core Domain

Synopsis:

After the success of fragment screening by X-ray crystallography against HIV-1 reverse transcriptase and the N-terminal domain of the PA subunit of pH1N1 influenza, a similar screening campaign was conducted for the HIV-1 integrase (IN) catalytic core domain (CCD). Specifically, a fragment-screening campaign was conducted to identify promising small molecule scaffolds targeting allosteric sites, such as the lens epithelium derived growth factor (LEDGF/p75) binding site within the CCD. Surprisingly, only one fragment hit was identified from the screening of 971 compound library. This chapter describes the result of fragment screening against the HIV-1 IN CCD using X-ray crystallography and the lead development that is currently underway. Additionally, several factors that may account for the poor hit rate of this screening campaign is also discussed in detail.

*The work described in this chapter would not have been possible without valuable contributions from members of Kavartskhelia laboratory at Ohio State University (OSU). Specifically, Dr. Lei Feng generously provided purified protein used for crystallographic fragment screening. Dr. Jacques Kessl and Alison Slaughter determined the enzymatic activity of the fragment hits. Pratiq Patel from Fuchs laboratory at OSU is currently performing the chemical synthesis of fragment derivatives.

The importance of integration of viral DNA into that of host cell DNA for HIV viral replication makes HIV-1 integrase (IN) an attractive target for antiviral therapy. However, integrase has proven to be a challenging clinical target with the first integrase inhibitor, raltegravir, gaining FDA approval in 2007. Treatment with raltegravir, an integrase strand-transfer inhibitor (INSTI), results in significant reduction in the viral load of HIV infected patients. Emerging clinical resistance to raltegravir (Quashie et al., 2013) fueled the development of INSTIs that retain activity against wild-type and raltegravir-resistant strains. This effort produced the recently FDA-approved drug, dolutegravir, and formulation of elvitegravir, an INSTI, with cobicistat and two NRTIs, tenofovir and ritonavir. Despite the availability of more efficacious INSTIs, drug resistance continues to be a concern. Recent identification of allosteric sites within IN has fueled the search of allosteric integrase inhibitors (ALLINIs), the focus of this study.

4.1 HIV-1 Integrase

HIV-1 IN is a 288 amino acid protein that is produced via proteolytic cleavage of the Gag-Pol polyprotein precursor. Biochemical studies have shown that HIV-1 IN consists of three distinct structural and functional domains—the N-terminal domain (NTD, residues 1-46), the catalytic core domain (CCD, residues 56-186), and the C-terminal domain (CTD, residues 195-288)—that either directly or indirectly contribute to IN-DNA interactions (Kessl et al., 2009). Poor protein solubility combined with the highly flexible nature of IN has made obtaining the full-length structure a formidable challenge for structural biologists (Yin and Craigie, 2010). Point mutations to enhance protein solubility have facilitated the structural characterization the individual domains

and two-domain protein constructs using X-ray crystallography and solution NMR (Dyda et al., 1994; Eijkelenboom et al., 1995, 1999; Cai et al., 1997).

4.2 Biochemistry of Integration

Integration is initiated by the formation of a nucleoprotein complex, termed the preintegration complex (PIC), derived from the core of the infecting virion (Bowerman et al., 1989). The PIC consists of viral DNA (vDNA) stably associated with viral and host proteins. Viral proteins involved in the formation of the PIC are limited to integrase, nucleocapsid, matrix, reverse transcriptase, and Vpr. On the other hand, more than 20 host cellular proteins have been identified to form part of the PIC, including the lens epithelium derived growth factor/p75 (LEDGF) (Cherepanov et al., 2003), nucleosome assembly protein 1-like 1 protein (NP1L1) (Raghavendra et al., 2010), importin 7 (Ao et al., 2007), barrier-to-autointegration factor (BAF) (Lin and Engelman, 2003), and gemin 2 (Hamamoto et al., 2006). Biophysical characterization of the PIC has been hindered by the relatively low concentration of PIC within the infected cells. Thus, it is not known which components of PIC, with the exception of integrase, remain stably associated or are jettisoned as the PIC is transported from the cytoplasm into the nucleus.

The mechanism of integration remains conserved among all reteroviral integrases. Integration entails first the 3'-end processing of the viral DNA followed by DNA strand transfer for incorporation into the host DNA. Upon completion of viral DNA synthesis, a tetramer of HIV-1 IN binds to the LTR DNA ends to form the stable synaptic complex (SSC). During the 3'-end processing, a water molecule acts as a nucleophile to remove the terminal dinucleotide from the blunt-ended HIV-1 cDNA of the SSC. This produces a recessed CA at the 3' end of the viral DNA and the resulting complex is referred to as the

cleaved-donor complex (CDC). The CDC is transported into the nucleus where it interacts with the target DNA to form the target-capture complex (TCC). Upon binding to the host DNA, the 3'-OH from the newly generated recessed nucleotides initiate a trans-esterification reaction by attacking the phosphodiester bonds on the opposite strands, which are separated by four to six bases in the 5' direction. This allows for joining of the viral DNA to occur on the same face of the double helix and results in the formation of the strand-transfer complex (STC). A gap-repair mechanism using the host DNA repair machinery completes the integration process.

4.3 Domains of HIV-1 Integrase

Comprised of the first 46 residues of HIV-1 IN, the NTD has an HTH fold conserved in all retroviral and retrotransposon integrases (Cai et al., 1997). Residues His12, His16, Cys40, and Cys43 of the NTD stabilize the binding of a zinc atom and contribute to the oligomerization of IN (Figure 41). Mutating the residues involved in zinc binding prevents the functional oligomerization of IN and inhibits the 3'-processing and strand transfer reactions (Lee and Han, 1996; Lee et al., 1997; Zheng et al., 1996).

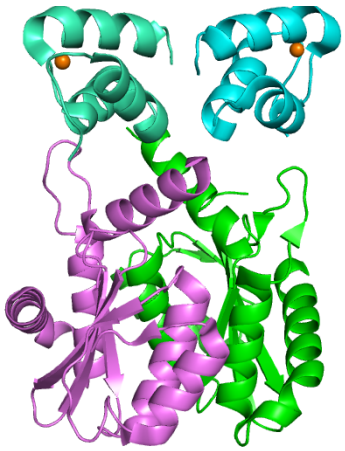


Figure 41: Crystal structure of two domain constructs of HIV-1 IN NTD-CCD (PDB code: 1K6Y). The CCD dimer is depicted as green and violet with the NTD dimer depicted as cyan and teal. Unless otherwise note, all figures are made using MacPyMol (Schrödinger, New York, NY).

A flexible linker (residues 47-55) connects the NTD to the CCD domain of IN, which is responsible for the 3'-processing and strand-transfer activities. The CCD contains a highly conserved active site characteristic of the polynucleotidyl transferases, including bacterial RNAse H (Engelman and Craigie, 1992; Kulkosky et al., 1992). The catalytic active site is comprised of three acidic residues Asp64, Asp116, and Glu152 that are critical for 3' processing, DNA strand transfer, and disintegration activities of IN (Li et al., 2011). Integrase activity is dependent on the presence of Mg^{2+} or Mn^{2+} ions (Engelman and Craigie, 1995; Delelis et al., 2008). Based on concentrations of these ions in the cell, Mg^{2+} is believed to be the physiological co-factor. Structural studies revealed the binding of a single of Mg^{2+} ion. However, based on structurally and functionally similar transferases, it is believed that a two-metal ion mechanism is responsible for the observed integrase activity and the binding of DNA stabilizes the second metal ion (Goldgur et al., 1998). The CCD is also involved in the functional tetramerization of IN. Mutating residues involved in the formation of the CCD dimer interface is associated with adverse effects on IN activity.

The CCD is connected to the CTD through a flexible loop formed by residues 187-194. The CTD is rich in basic amino acids and adopts an Src homology 3-like fold (Eijkelenboom et al., 1995). Based on structurally similar proteins, CTD is believed to bind to the minor groove of DNA in a non-specific manner. This stabilizes the DNA binding and enables the integrase catalytic activity. Dimerization was observed upon structural characterization of individual and the two domain CTD constructs (Figure 42). Comparison of the individual and two domain structures for the NTD and CTD, respectively, revealed that the dimerization of NTD and CCD varied based on the domain

construct used for crystallization. However, the CCD dimer was found to be consistent regardless of the protein construct (CCD by itself or as a part of a two domain construct) used for structural characterization (Wang et al., 2001).

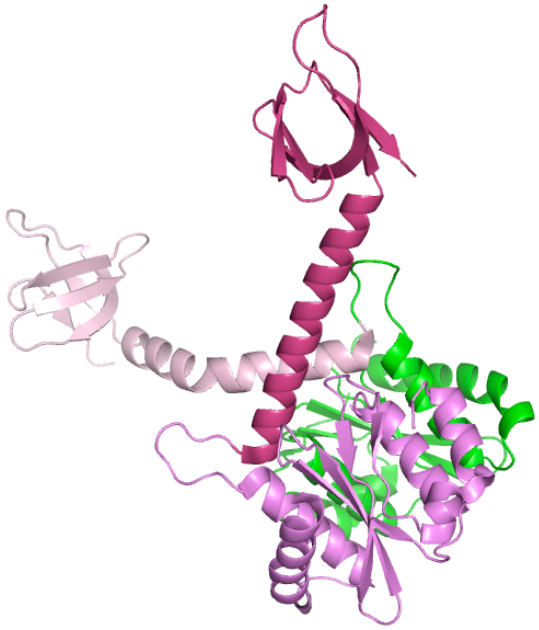


Figure 42: Cartoon depiction of HIV-1 IN CCD-CTD crystal structure (PDB code: 1EX4). The CCD dimer is depicted as green and violet with the CTD dimer depicted as maroon and light pink.

4.4 HIV-1 Integrase Multimerization

Integrase activity is highly dependent upon multimerization. Complementation studies revealed IN activity was maintained after mixing IN with individually inactive domains (Engelman et al., 1993; van Gent et al., 1993). Factors that enhance multimerization, such as Zn^{2+}, also increased the enzymatic activity of IN (Lee et al., 1997). SAXS has also shown that the dimeric form of IN is responsible for the 3'-processing activity (Baranova et al., 2007; Bojja et al., 2013). A tetramer or a dimer of dimers has been found to be necessary for concerted integration. Due to poor protein solubility, the exact structure of the IN tetramer remains unknown. However, Hare et al. (2010) successfully used the more soluble prototype foamy virus (PFV) IN to structurally

characterize the PFV intasomes, which consists of a tetrameric PFV IN bound to a pre-cleaved 19-mer vDNA substrate.

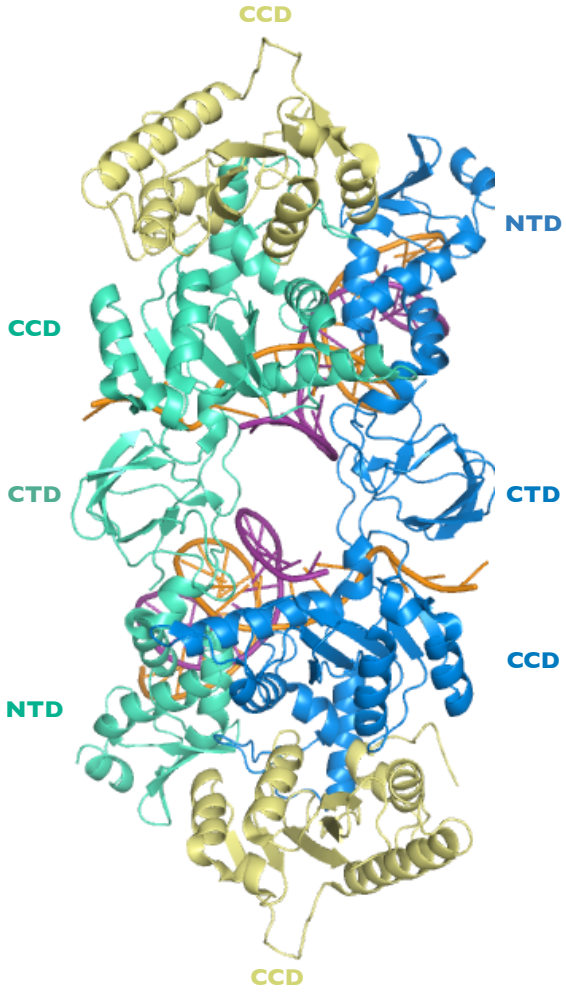


Figure 43: Architecture of the PFV intasome. The inner subunits of the tetramer bound to the viral DNA are shown in blue and cyan and the outer subunits are shown in yellow. The reactive and non-transferred DNA strands are shown in purple and orange, respectively.

Initially, assembled in the presence of Zn^{2+}, PFV intasomes have also been structurally characterized in the presence of physiological co-factors, Mn^{2+} and Mg^{2+} ions, target DNA (tDNA), and INSTIs. Regardless of the substrate, an IN dimer binds to the vDNA, with only one monomer from the dimer making contact with the DNA (Figure 43). Strong electron density for this monomer was observed with the exception of 9

residues from the NTD and 18 residues from the CTD. This was not the case for the second monomer, for which only the CCD could be successfully modeled.

The full PFV intasome was found to be a tetramer, formed by a pair of the symmetrically related PFV IN-vDNA assemblies. The inner PFV IN monomers of the tetramer were found to make extensive contacts with nucleic acid. The catalytic activity of PFV IN is associated with the inner monomer due to the proximity of the catalytic triad of these monomers to the 3'-hydroxyl of the vDNA (Hare et al., 2012). Interestingly, soaking of PFV intasome crystals with Mn^{2+} revealed two-metal binding within the active site, whereas only one metal ion was found upon soaking with Mg^{2+}. However, two-metal binding for Mg^{2+} was observed only in the presence of raltegavir.

4.5 LEDGF/p75 and HIV-1 Integrase

Co-immunoprecipitation from cells overexpressing IN identified the interaction between IN and LEDGF/p75, a member of the hepatoma-derived growth factor family (HDGF) (Cherepanov et al., 2003). The gene encoding for PC4- and SFRS-interacting protein 1 (PSIP1) was used to express the two splice variants, LEDGF/p52 and LEDGF/p75 proteins (Singh et al., 2000). Both LEDGF/p52 and LEDGF/p75 possess the PWWP domain (Izumoto et al. 2000, Stec et al., 2000, Dietz et al., 2002), the nuclear localization signal, the A/T hook-like elements (Aravind and Landsman, 1998), and three charged regions (CR1, CR2, and CR3) (Llano et al., 2006), which allow for nuclear localization and chromatin binding. Unlike LEDGF/p52, the C-terminus of LEDGF/p75 extends for an additional 205 residues and contains the IN-binding domain (IBD, residues 347-429), which has been shown to interact with the catalytic core domain of IN based on both co-localization and structural studies.

Multiple studies entailing mutagenesis, RNAi-mediated depletion, knockdown/knockout studies of cells expressing LEDGF/p75 establish its role as a co-factor for viral replication. IN was concentrated in the cytoplasm upon depletion of LEDGF/p75 by RNAi (Maertens et al., 2003; Llano et al., 2004). Additionally, a reduction in HIV infectivity was observed when LEDGF/p75 expression was reduced by either RNAi or knockout mutations (Vandekerckhove et al., 2006). However, PIC isolated from the LEDGF/p75 knockout cells was found to retain strand-transfer activity (Shun et al., 2007). Thus, binding of LEDGF/p75 is believed to be responsible of the nuclear transport of PIC. Additionally, loss in chromatin association was observed for IN mutants that are unable to interact with LEDGF/p75, indicating that LEDGF/p75 is also responsible for tethering the IN to chromatin. Chromatin linkage of IN is facilitated by the specific domains of LEDGF/p75. The 83-amino acid integrase binding domain (IBD) interacts with IN while the PWWP and the AT-hook DNA binding domains tethers this complex to chromatin (Engelman and Cherepanov et al., 2008; Poeschla, 2008).

Cherepanov et al. (2005) determined the crystal structure of the LEDGF/p75 IBD bound to the Phe185Lys variant of CCD. The IBD formed a right-handed alpha-helix bundle composed of two helix-hairpin-helix repeats, similar to the HEAT domains from other proteins (Figure 44). A less structured inter-helical loop of the IBD extended into a binding pocket at the interface of the dimer formed by the CCD domains. This pocket, referred to as the LEDGF/p75-binding pocket, is formed by key residues 166-171 of the first IN monomer and a hydrophobic patch of residues from alpha helices 1 and 3 of the second monomer (Figure 44). The side chain of Asp366 from the IBD forms bidentate

hydrogen bonds with the primary amino groups of Glu170 and His171 from one monomer IN.

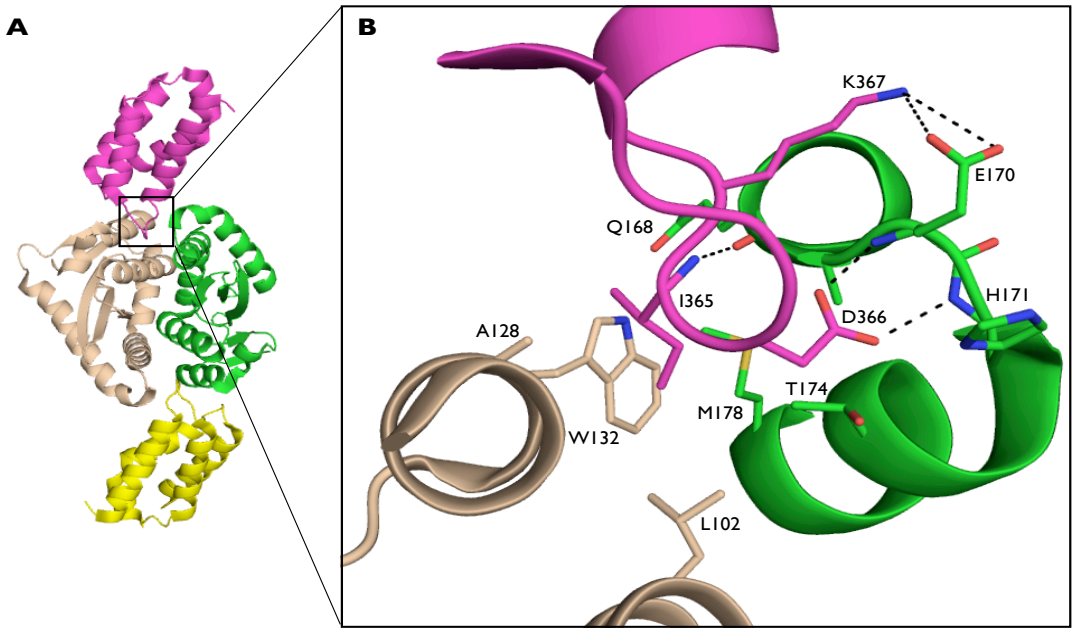


Figure 44: A.) Crystal structure of IBD of LEDGF/p75 interaction with the HIV-1 IN CCD dimer (PDB code: 2B4J). The two copies of IBD are depicted in magenta and yellow and the CCD dimer in beige and green. **B.)** Detailed view of the CCD-IBD interface.

Loss of binding was observed when Asp366 from LEDGF/p75 was mutated to either alanine or asparagine, indicating this hydrogen-bonding interaction is crucial for binding. The side chain of Ile365 from the IBD binds to a hydrophobic patch formed by Leu102, Ala128, and Trp132 from the second monomer of IN and Thr174 and Met178 from the first monomer. Additionally, the backbone amide of Ile 365 forms a hydrogen bond with the backbone carbonyl of Gln168 from the first IN monomer. Side chains of Phe406 and Val408 from the IBD also interact with the exposed side chain of Trp131 from second CCD monomer. Three water molecules are buried within the IBD and CCD interface; however, only one water molecule forms bridging hydrogen bonds with the backbone

carbonyl groups of Ile365 from the IBD and Thr125 from the second monomer of the CCD.

The NTD of integrase has also been found to interact with LEDGF/p75. Introducing mutations, such as His12Asn, or the deleting the entire zinc-binding domain within the NTD significantly reduces the interaction between IN and LEDGF/p75. Hare et al. (2009) provided the first structural insights regarding the exact nature of this interaction using the crystal structure of HIV-2 IN NTD-CCD in complex with the IBD of LEDGF/p75. Despite only ~60% sequence similarity between the CCD of HIV-1 and HIV-2 IN and a significant change to the dimer interface due to the replacement of Ala at positions 128 and 129 of HIV-1 CCD with Met and Val, respectively, in that of HIV-2, interactions similar to the previously reported HIV-1 IN CCD-IBD complex were observed. Even though the CCD of HIV-2 IN has different residues at positions 170 and 171 (Asn170 and Thr171), the formation of bidentate hydrogen bonds between Asp366 of the IBD to the backbone amides was maintained. Hydrophobic contacts between the side chains of Leu368, Phe406, and Val408 from the IBD and the hydrophobic patch of residues from the CCD are retained. The acidic residues from the α1 helix of the NTD form extensive charge-charge interactions with the highly basic residues of the α4 helix from the IBD of LEDGF/p75. A significant reduction in binding was observed upon reverse charge engineering of the interacting residues of the IBD (Hare et al., 2009).

4.6 Allosteric Integrase Inhibitors

Structural characterization of the LEDGF/p75-binding site within the CCD dimer has fueled the discovery of inhibitors targeting allosteric sites within HIV-1 IN, referred to as allosteric integrase inhibitors (ALLINIs). The druggability of the LEDGF/p75

binding pocket was explored using peptides harboring key interacting residues, such as Ile365 and Asp366. These peptides were found to enhance tetramerization of IN in the absence of viral DNA (Hayouka et al., 2007) as well as inhibit the binding of vDNA (Hayouka et al., 2007; Al-Mawsawi et al., 2008) and LEDGF/p75 (Al-Mawsawi et al., 2008) to IN. Based on these results, it was inferred that the mechanism of action for LEDGF/p75 inhibitors is likely due to their ability to affect multimerization of IN.

The promising inhibitory activity of peptides mimicking the IBD of LEDGF/p75 fueled the discovery of small molecule inhibitors of LEDGF/p75 binding. Two alternate approaches were undertaken to identify quinoline-based inhibitors of LEDGF/p75 binding (Figure 45). Christ and colleagues (2010) used an *in silico* pharmacophore model derived from the structural information from the IN-CCD-IBD crystal structure to screen a commercial library of 200,000 compounds. On the other hand, a high throughput screen for inhibitors of IN 3'-processing activity led to the discovery of a quinoline-based scaffold, BI-A, by Fenwick and colleagues (2011). Although these hits did not have antiviral activity, both had promising *in vitro* activity with LEDGIN-1 inhibiting LEDGF/p75-IN interaction by 36% at a concentration of 100 μM whereas BI-A inhibits 50% of IN 3'-processing activity at 9 μM. Additionally, both of these compounds were found to inhibit LEDGF/p75 binding as well as LEDGF/p75-independent 3'-processing and DNA strand-transfer activities of IN *in vitro* (Kessl et al., 2012). In the absence of vDNA, these inhibitors were found enhance the multimerization of HIV-1 IN and inhibit the *in vitro* formation of the SSC. Crystal structures of BI-A and LEDGIN-1 revealed compound binding to the LEDGF/p75-binding pocket, thereby establishing them as true allosteric IN inhibitors (ALLINIs). The key pharmacophore was found to be the

carboxylic acid functionality, which mimics the bidentate hydrogen bonds formed by the side chain of Asp366 in IBD. Despite their lack of antiviral activity, BI-A and LEDGIN-1 have been developed into potent inhibitors of HIV-1 viral replication with EC_{50} values of 10-100 nM.

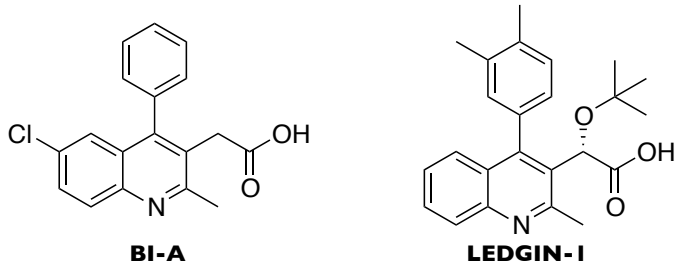


Figure 45: Chemical structures of the first reported inhibitors of LEDGF/p75 binding.

The promising activity of BI-A and LEDGIN-1 combined with the previous difficulties experienced in developing raltegravir and new generation INSTIs has fueled the search for allosteric inhibitors of LEDGF. SPR and NMR were used to identify benzodioxole-4-carboxylic acid as a promising inhibitor from a fragment-screening campaign of 500 compounds (Rhodes et al., 2011). Subsequent, crystallographic characterization verified that this fragment to be an ALLINI bound at the LEDGF/p75 binding pocket. Derivatives of the original hit were found to inhibit LEDGF/p75 binding and the 3'-processing activity of IN as well HIV-1 infectivity in the cell (Peat et al., 2012).

4.7 Fragment Screening

A fragment-screening campaign by X-ray crystallography using an in-house library of 971 compounds was conducted to identify promising new chemical scaffolds targeting the HIV-1 IN CCD. Unlike HIV-1 RT and pH1N1 PA_N, which gave hit rates of 4.0% and 1.0%, respectively; only one hit (0.1%) was obtained from crystallographic fragment screening of the HIV-1 IN CCD. This fragment hit and subsequent work that is

currently underway will be discussed in full detail. Additionally, several hypotheses will be discussed to account for the poor hit rate of this screening campaign.

One of the by-products of screening was the significant improvement in the resolution quality of the IN CCD crystals from 1.80 to 1.40 Å upon soaking into one cocktail, PD1d (see Material and Methods for refinement statistics). Although no fragment binding was observed, the improvement in resolution allowed for a portion of a highly disordered internal loop (residues 141-153), which includes the catalytic residue, Glu152, to be modeled. Based on the electron density, the CCD is a catalytically inactive conformation, with the side chain of Glu152 pointing away from the active site.

4.7.1 Crystal Optimization

The original crystallization condition for the Phe185Lys construct of IN CCD provided by Lei Feng produced crystals that diffracted X-rays to 1.80 Å within three hours of setup. However, the majority of these crystals were extremely small with only one out of 24 crystallization drops producing a usable crystal. Thus, optimization of the crystallization conditions was necessary to produce a larger quantity of high-quality crystals amenable for fragment screening by X-ray crystallography. No improvement in crystal size and quality was observed upon altering the protein concentration, the pH, or the concentration of the precipitant (PEG 8000) in the crystallization solution. After numerous crystallization trials, a combination of pre-seeding (see Materials and Methods) and reducing the volume of the well solution from 500 μL to 50 μL produced larger quantities of highly robust crystals. With this optimized condition, the number of harvestable crystals improved from one crystal out of 24 crystallization drops to three

crystals per drop. The availability of suitable crystals allowed for rapid fragment screening using the high-throughput screening protocol described in Chapter 2.

4.7.2 Fragment Hit

Surprisingly, screening of the fragment cocktails at concentrations of 20 mM revealed binding of only one fragment to the CCD. This particular hit was not exciting as the fragment bound to a non-biologically relevant pocket formed by crystal contacts with a crystallographically related CCD monomer. However, fragment binding was validated by soaking of the individual components of the cocktail and the cocktail itself, which served as a positive control, at a concentration of 20 mM. Fragment binding was observed for the cocktail; however no electron density for the fragment was observed from the individual fragment soak. Subsequently, soaking was repeated at concentrations of 20, 50, and 100 mM. Once again, fragment binding was not observed for the 20 mM soak of the individual fragments. However, the 50 mM soak revealed fragment binding for **55** (referred internally as MB36-3) to the site originally identified from the cocktail soak. Weak electron density was also observed at the LEDGF/p75 binding site. The 100 mM soak of **55** revealed strong electron density for fragment bound to the original site as well as the LEDGF/p75 site (Figure 46).

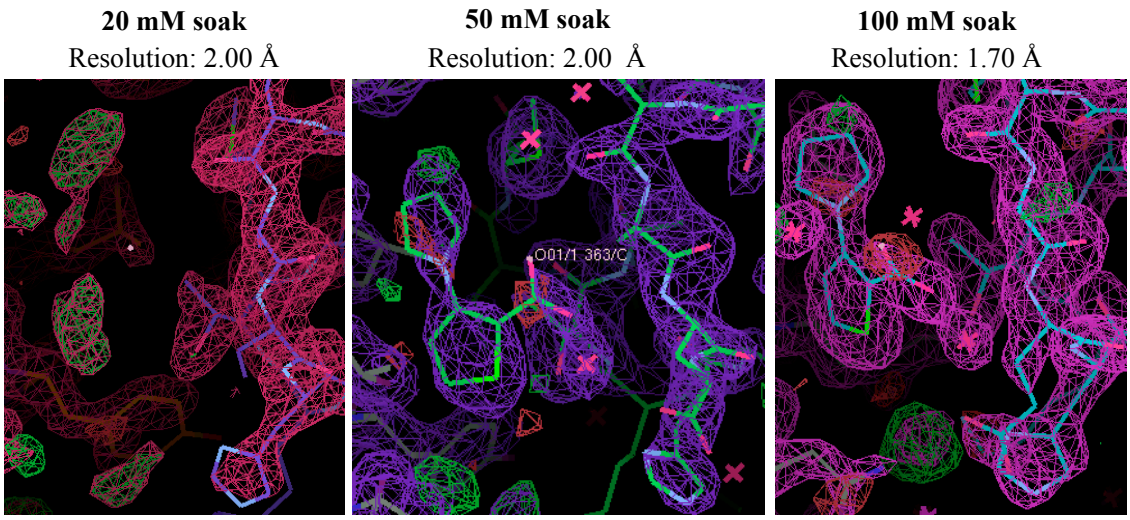


Figure 46: Higher concentration soaks reveal binding of **55** to the LEDGF/p75-binding site at the CCD dimer interface. Electron density from a $2F_o$-F_c map (contoured at 1.0 σ) for **55** bound at the LEDGF/p75 site at the CCD dimer interface increases with increasing concentration.

4.7.3 Crystal Structure of 55 Bound to the CCD dimer

55 bound to two sites within the CCD located approximately 27 Å apart within the individual CCD monomer. The binding site for the first molecule of **55** is created by crystallographic contact with a symmetry-related molecule of CCD. Although this pocket does not exist within the biologically relevant CCD dimer, binding of **55** to this region is interesting. As shown in Figure 47, the carboxylic acid moiety of **55** forms a hydrogen-bond network with the solvent water molecules and the backbone amide of Gly94. Additionally, it forms hydrophobic interactions with the side chains of Lys186, Arg187, and Lys188 from the symmetry-related molecule. This interaction with the KRK region of CCD is of interest because Lys186 and Lys188 form vital salt bridges with the side chains of Glu11 and Asp25 and hydrophobic interactions with Tyr15 from the NTD. Mutagenesis studies revealed that Lys186 and Lys188 are essential for the functional tetramerization of IN and a reductions in the 3'-processing and strand-transfer activities of IN are observed (Berthoux et al., 2007).

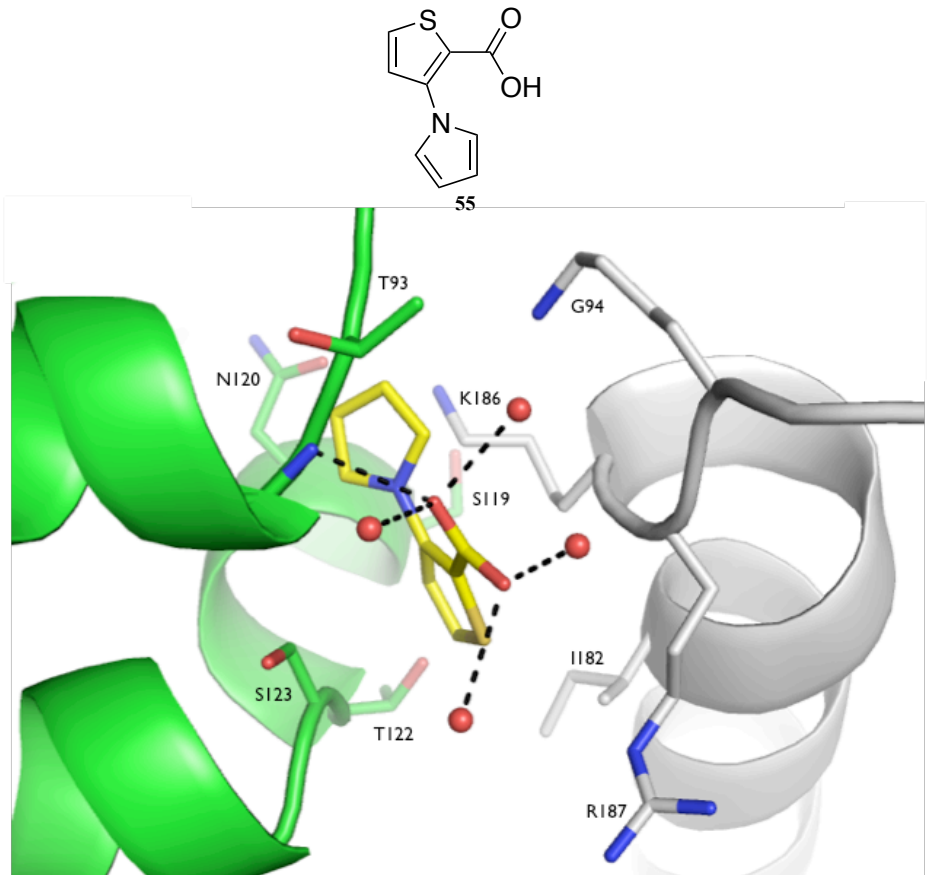


Figure 47: 55 bound to a pocket formed by the CCD monomer (green) and symmetry-related molecule (gray).

Superposition generated by aligning the symmetry-related molecule from the CCD bound to **55** to the IN NTD-CCD structure suggests that **55** can potentially disrupt the stability of the NTD-CCD domains (Figure 48). However, fragment binding would require the side chain of Ser17 from the NTD to shift, thereby disrupting the interaction between the NTD and CCD domains.

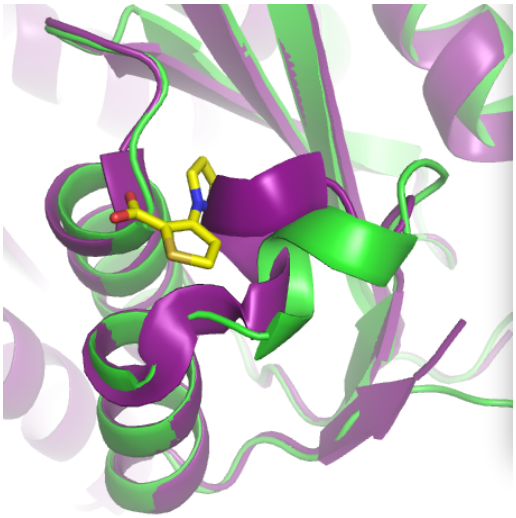


Figure 48: Superposition of the crystal structure of HIV-1 IN CCD with **55** bound (in green) to the published HIV-1 IN NTD-CCD (PDB ID: 1KY6) structure (in purple).

It is important to emphasize that **55** does not interact directly with the KRK region of the CCD monomer itself. Instead, it interacts with the KRK region of the symmetry-related molecule of CCD located at the interface of a crystal contact. A superposition generated by aligning the CCD monomer bound to **55** to the IN NTD-CCD structure reveals that this site is occluded by protein, thus binding of **55** may not be possible (Figure 48). Nonetheless, structural characterization of the binding of **55** to NTD-CCD variant of IN is currently being pursued since the extensive interactions between the NTD and the IBD of LEDGF/p75 are necessary for functional IN activity. Additionally, the interaction between inhibitors of LEDGF/p75 binding with respect to the NTD has yet to be structurally characterized.

4.7.4 Fragment Binding to the LEDGF/p75 Site

A 100 mM soak of **55** revealed strong electron density for fragment binding to the LEDGF/p75 binding site at a resolution of 1.70 Å. The refinement statistics for this dataset is described in the Materials and Methods. As shown in Figure 49, binding of **55** induces significant conformational changes within the pocket. The carboxylate of **55**

forms hydrogen-bonding interactions with the backbone of Glu170 and three additional hydrogen bonds with water molecules. Fragment binding clearly stabilize these three water molecules since only weak electron density is observed in the apo structure. The side chain of Gln95 flips into the cavity to form a weak hydrogen bonding interaction with the sulfur from **55**. The side chain of Leu102 flips and that of Trp132 rotates ~30° into the cavity to form hydrophobic interactions with the pyrrole ring of **55**. Additional hydrophobic interactions are formed between **55** and the side chains of Met178 and Thr174 of the CCD monomer and Ala128 from the symmetry-related CCD molecule.

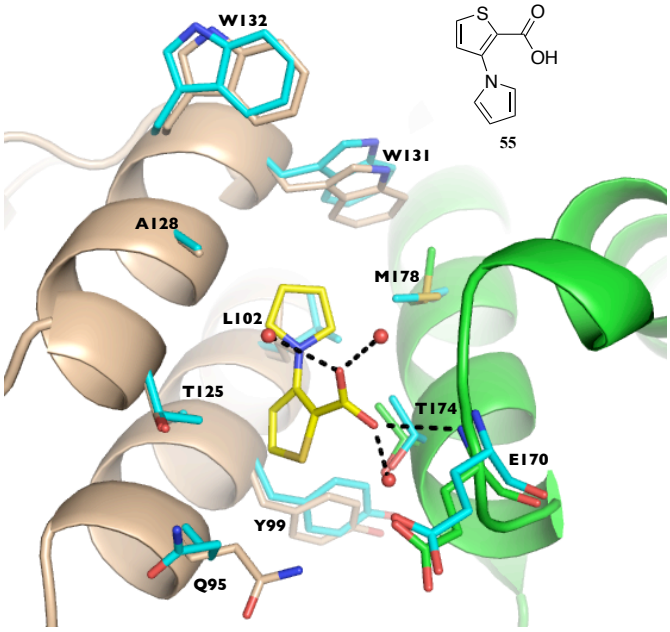


Figure 49: Crystal structure of **55** (yellow) bound to HIV-1 IN CCD dimer (green and beige). Residues in the apo structure are depicted in cyan. Dashed lines in black represent hydrogen-bonding interactions.

Superposition of the crystal structures of **55** and BI-B-II, a potent LEDGF/p75 inhibitor derived from BI-A, bound to the CCD dimer suggests that **55** can readily be expanded into a lead candidate using a structure-based drug design approach (Figure 50).

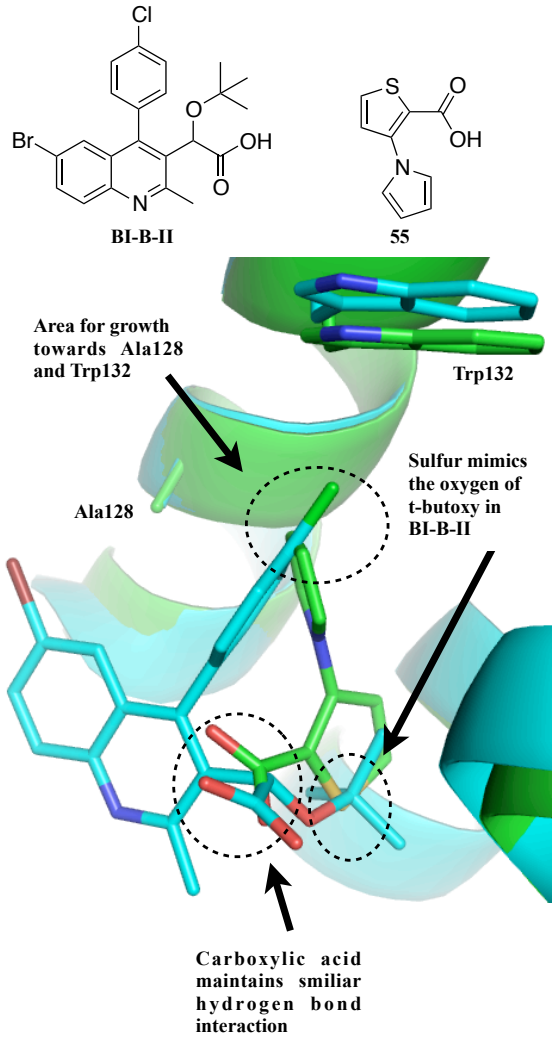


Figure 50: Possible paths for extension of **55** based on a superposition of **55** (green) and BI-B-II (cyan) bound to the CCD dimer.

The overlay indicates room for growth towards Trp131. Previously successful modifications within the BI series in combination with a fragment expansion-based approach can be explored to rapidly improve potency of **55**. Additionally, the two-linked rings based pharmacophore of **55** would allow for development of LEDGF/p75 inhibitors without conflict with existing intellectual property.

4.7.5 Enzymatic Activity

Jacques Kessl and Alison Slaughter from the Kvaratskhelia laboratory at Ohio State University assayed **55** and 15 chemically similar fragments (Figure 49) manually

identified from the in-house fragment library for inhibitory activity. In addition, a new lot of **55** purchased from Maybridge and assayed for activity to ensure the activity of the original sample of **55** was not due to contamination or chemical degradation.

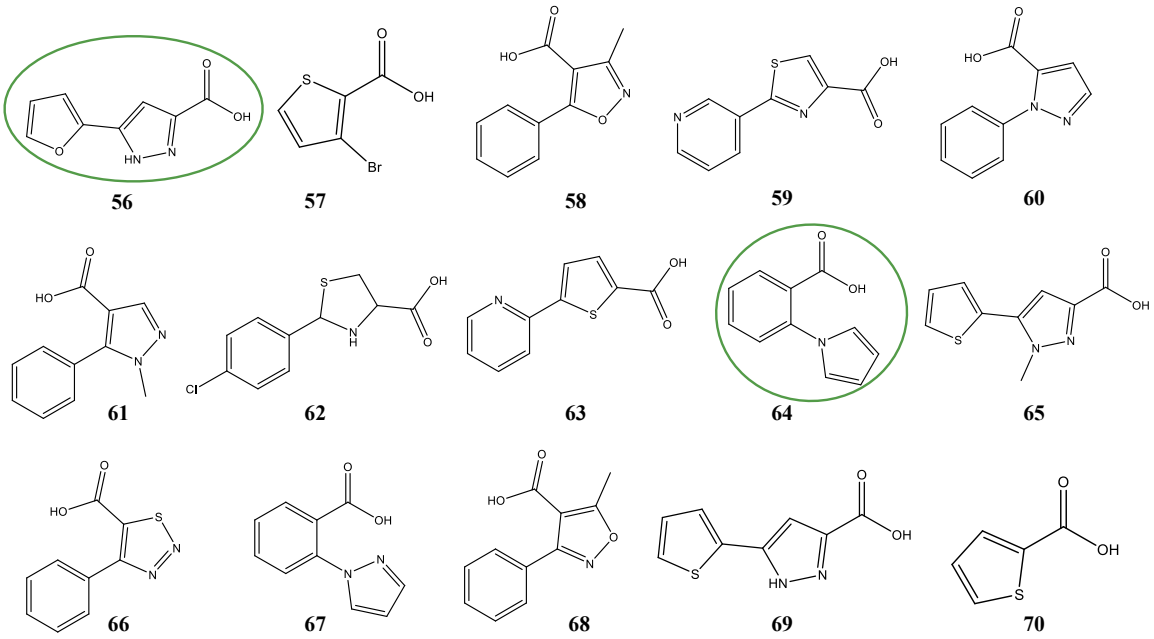


Figure 51: Fragments selected for enzymatic assay based on the initial crystallographic hit, **55**. Compounds circled in green were found to have some inhibitory activity.

The compounds were initially assayed for inhibitory activity in a high-throughput fluorescent-based LEDGF/p75-dependent integration assay at concentrations of 200 and 400 μM with DMSO and BIB-II, a potent ALLINI, as positive controls. Of the 17 compounds assayed, only two fragments (**56** and **64**) were found to be inhibitory at the concentrations tested. A 10-point titration (20 – 200 μM) experiment was subsequently conducted to determine the IC_{50} for these fragments. **56** was found to have an IC_{50} of 85 μM in the LEDGF/p75-dependent integration assay. While the IC_{50} of **64** exceeded the highest concentration tested, nonetheless evidence for activity was observed. In parallel, both fragments were also assayed in a proprietary fluorescent-based inhibition assay for LEDGF/p75 binding at a concentration range of 20 to 200 μM. No inhibition of

LEDGF/p75 binding was observed for **64** at the tested concentrations. However, **56**, was found to have an IC_{50} of roughly 100–200 μM. The effect of **56** and **64** on the multimerization of IN was also assessed using their in-house fluorescent-based assay. Surprisingly, no inhibitory activity was observed for either fragment in the multimerization assay. These initial results suggest that these compounds, specifically **56**, may inhibit IN activity through a new mechanism of action. A series of enzymatic assays developed by the Kvaratskhelia laboratory is being used to characterize the exact mechanisms of action for **56** and **64** and to ensure that these compounds are not false positives.

Based on these results, crystallography was undertaken to obtain the crystal structures of **56** and **64** bound to CCD, respectively. Pre-formed crystals of CCD were soaked overnight into 20, 50, 100, and 500 mM solutions of **56** and **64** dissolved in DMSO, water, and methanol, respectively. Although each of these crystals diffracted X-rays to 1.6–2.4 Å, subsequent refinement of the datasets revealed no fragment binding. Co-crystallization experiments of **56** and **4** in complex with Phe185Lys CCD, respectively, produced microcrystals (<<100 μm) with a bipyramidal morphology different from the tetragonal crystals or the apo CCD. No significant improvement in crystal quality resulted from extensive optimization of the crystallization conditions. The two harvestable co-crystals of **56** diffracted X-rays to only 9–10 Å and diffraction quality quickly decreased during the course of data collection.

4.7.5 Molecular Modeling

Although **55** has promising biological activity, structure-based drug design is challenging due to the lack of a crystal structure for the bound fragment. Thus, fragment

expansion of **55** is the current focus despite its lack of initial inhibitory activity. Surface depiction for the binding region of **55** indicates that possible extension from the pyrrole ring would be favorable. Extensions from both the 2- and 3-positions of the pyrrole ring are possible. However, as shown in Figure 52, extension from the 2-position may be limited sterically, whereas extension from the 3-position would allow for interactions with Trp132.

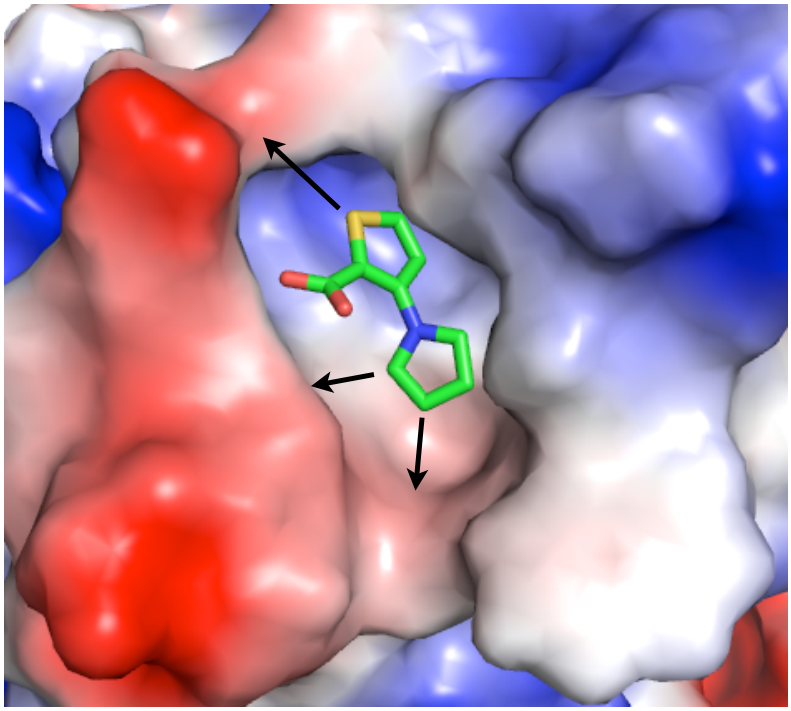


Figure 52: Surface depiction of the binding cavity for **55** (shown in green). The black arrows indicate areas within the pocket for potential fragment expansion of **55**.

Alternatively, **55** can be extended to interact with Gln91 by replacing the thiophene ring with another five-membered aromatic ring. Although the space here is limited, crystal structures of other known LEDGF/p75 inhibitors indicate that the Gln91 side chain is highly flexible and can adopt an "open" conformation, which can allow for further expansion.

The availability of a crystal structure for **55** allows for rapid exploration of fragment growth space by docking using AutoDock Vina (Trott and Olson, 2010). Docked poses were manually ranked based on the ability of the modified derivative to retain the original binding mode of **55** and the hydrogen bond with backbone amide of Glu170 as well as to gain favorable interactions with the protein. Initial docking studies were conducted to explore substitution possibilities at the 2- and 3-positions of the pyrrole ring. As predicted, extension from the 2-position of the pyrrole was limited to small substituents whereas larger substituents, such as rings, were tolerated at the 3-position.

Based on the initial docked poses, it was apparent that the limited conformational flexibility of the designed molecules did not allow for significant interactions with the protein. Thus, molecules with single-atom linkers (CH_2, NH, and O) between the thiophene and pyrrole rings of **55** were designed *in silico* and docked against the **55** bound protein structure. Although these compounds docked successfully into the LEDGF/p75-binding pocket, most were not promising due to the loss of the hydrogen bond between the backbone amide of Glu170 and the carboxylate functionality of the bound fragments. Engineering molecular flexibility through a single-atom linker (CH_2, NH, and O) for substituents at the 3-position proved to be more promising. Regardless of the linker location, the amino linker was found to be the most favored out of the three linkers tested due to its ability to form favorable hydrogen bonds with the backbone carbonyls of nearby residues. Of the compounds screened *in silico*, four compounds (Figure 53) have been prioritized for chemical synthesis, which is currently underway by Pratiq Patel from the Fuchs laboratory at Ohio State University.

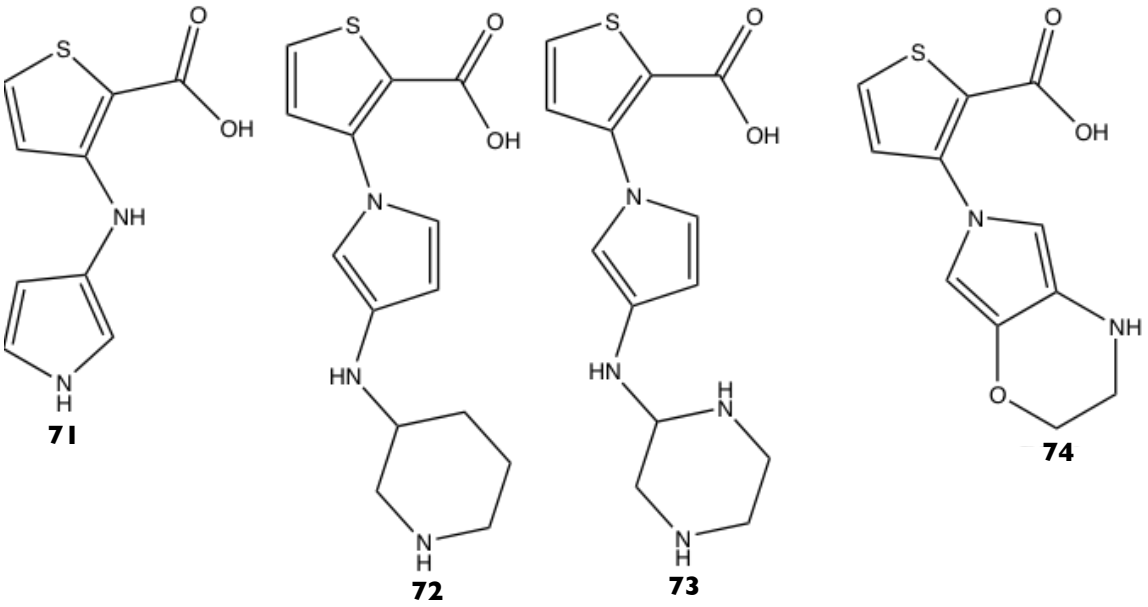


Figure 53: Promising compounds for lead development as predicted by AutoDock Vina.

4.8 HIV-1 Integrase CCD and the Poor Hit Rate in Crystallographic Fragment Screening

HIV-1 IN CCD proved to be a difficult target for fragment screening especially when using X-ray crystallography as the primary screening tool. A lower hit rate, similar to that of pH1N1 PA_N, was expected since screening against an individual domain can limit the potential for binding pockets. However, the identification of only one fragment hit from the primary screen was surprising and we felt it was important to explore and consider the factors that can affect the success of a fragment-screening campaign performed using X-ray crystallography.

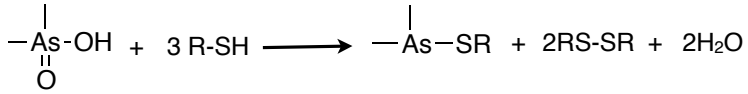
4.8.1 Crystal Packing

The suitability of a protein crystal for screening is typically determined based on its resolution and robustness of the crystal for soaking. Thus, the optimized CCD crystals that can survive overnight soaks in 20% (v/v) DMSO and diffract X-rays to an average resolution of 1.8 Å were believed to be ideal for a crystallographic screening campaign.

However, a recent review by Davies and Tickle (2011) suggests additional factors, specifically crystal packing and solvent channels, should be taken into consideration. Crystal packing can trap the protein in a conformational state that is not favorable for ligand binding. Alternatively, symmetry-related molecules surrounding the asymmetric unit of the crystal can occlude binding sites. However, this is not necessarily the case for the CCD. Here, the biologically relevant CCD dimer and more importantly, the LEDGF/p75-binding site, could not be formed without the crystallographic symmetry-related CCD molecule.

4.8.2 The Crystallization Effect

Although crystal packing did not affect accessibility of either the active site or the allosteric LEDGF/p75 binding sites for fragment binding, it is possible that the crystallization condition could affect accessibility to the catalytic active site. The original crystal structure for the Phe185Lys CCD construct reported by Dyda et al. (1994) as well as those generated during the course of the screening campaign revealed that covalently bonded cacodylate from the crystallization condition to Cys65 and Cys134 of the CCD. Goldgur et al. (1998) proposed the cysteine residues are modified as shown in the reaction below with two of the required thiol groups provided by DTT, another component for crystallization. Covalent binding to Cys65 is of particular concern as it is located near the catalytic active site residues (Asp64, Asp116, and Glu152) (Figure 54).



Comparison to the arsenic-free structure of the Phe185His CCD (Bujacz et al., 1996) revealed no conformational changes for the catalytic residues Asp64 and Asp116, thus the covalent binding of cacodylate was not believed to have an effect on the active

site. However, this cannot be said regarding the third catalytic residue Glu152. The location of Glu152, located within the highly disordered internal loop (residues 141-153) could not be modeled within the published crystal structure for the apo Phe185Lys CCD construct (Dyda et al., 1994). This was not the case for the 1.40 Å crystal structure, which allowed residues – to be modeled. The active site residue, Glu152, was found to be in an "inactive" conformation, pointing away from Asp64 and Asp116. Similarly, Glu152 in the Phe185His construct is also in an inactive conformation (Bujacz et al., 1996). However, this conformation is stabilized by crystal contacts with a symmetry-related molecule, which is not part of the CCD dimer (Goldgur et al., 1998).

Structural comparison to the Cys56Ser, Phe185Lys and Trp131Glu, Phe185Lys double mutant CCD constructs (Figure 52), which crystallize in the presence and absence of cacodylate, respectively, demonstrated that the loop connecting the β4 strand to α2 helix reorients to accommodate the modified cysteine residues (Goldgur et al., 1998). Furthermore, Mg^{2+} binding could be characterized within the cacodylate-free structure of Trp131Glu CCD. Thus, it is possible that the covalently-modified cysteines may affect ligand binding to the active site due to changes in the chemical and conformational landscape of the pocket.

To validate this hypothesis, robotic crystallization screening is currently being explored to identify an alternate, non-reactive crystallization condition for the Phe185Lys CCD construct that can produce crystals amenable for fragment screening and allow access to the active site. The Phe185His CCD construct can also be crystallized and used for fragment screening. However, further optimization is necessary as crystals of Phe185His CCD only diffract to 2.6 Å resolution (Bujacz et al., 1996). The Trp131Glu,

Phe185Lys double mutant CCD construct, which is cacodylate free and diffracts to 2.0 Å resolution, can be used for screening. However, fragment binding to the LEDGF/p75 pocket will be greatly affected.

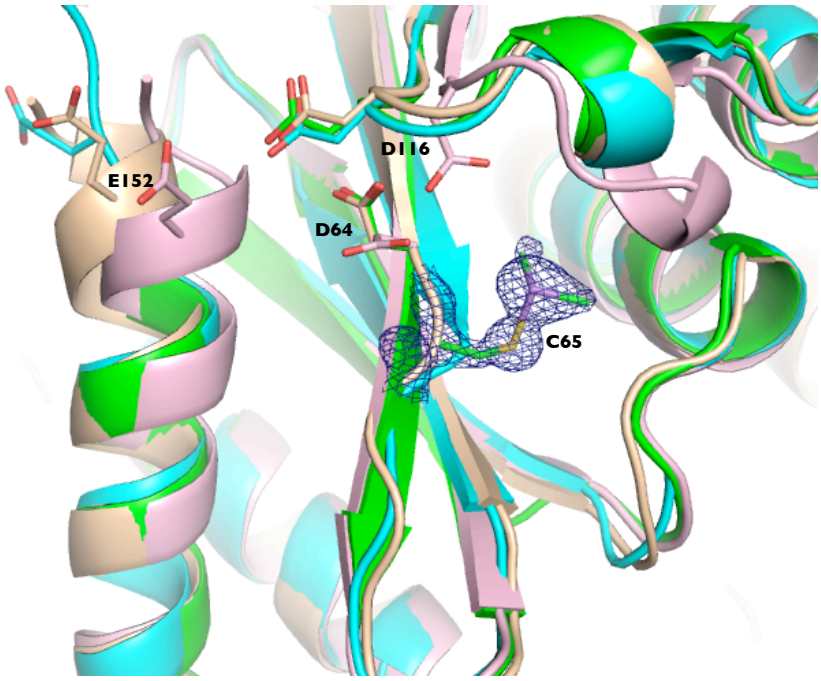


Figure 54: Covalent addition of cacodylate to Cys65 and its effect on the conformation of the catalytic residues of the HIV-1 IN CCD. Shown in green is the 1.4 Å crystal structure of Phe185Lys CCD with the electron density from a $2F_o$-F_c map (contoured at 1.5σ) for the covalently modified Cys65. No significant conformational change is observed for the Asp65 and Asp152 compared to the apo crystal structures for the Phe185His (cyan; PDB code: 2ITG) and the double mutant Cys56Ser, Phe185Lys (shown in beige; PDB code: 1BIZ). Comparison to the Trp131Glu, Phe185Lys CCD structure with unmodified Cys65 (shown in pink; PDB code: 1BIU) reveals the covalent addition of cacodylate significantly alters the active site of the CCD.

4.8.3 Construct Selection

The CCD construct selected for fragment screening may also affect the hit rate of a fragment-screening campaign. The apo crystal structure for the Phe185Lys CCD construct used for fragment screening, which includes residues 50 to 212, has a highly disordered internal loop located between residues 141 and 153. As a result, this region cannot be properly modeled due to a lack of clear electron density for the protein

backbone within the apo crystal structure. However, several cocktail soaks stabilized this region enough to reveal electron density for this region, allowing for most of the missing protein backbone and the side chains of few residues to be modeled. This suggests that fragment binding may stabilize this highly flexible loop to a certain extent by either indirectly stabilizing the crystal lattice or by directly interacting with the residues. In the latter case, hit identification is difficult since this region is not fully modeled within the reference structure and the order of and occupancy for the fragment may be poor.

Presence of disordered regions within protein crystal structures is common. Crystal structures of both pH1N1 PA_N and the RT-rilpivirine complex have disordered residues at the C-terminus, and this factor did not have a dramatic effect on fragment binding. This may not necessarily hold true for disordered regions within the protein (residues 141-153), which happens to be the case for the Phe185Lys CCD construct. It is conceivable that another construct of CCD with electron density for the entire protein backbone may be more amenable for fragment screening, although perhaps somewhat unlikely since most of the CCD in this crystal form is well ordered.

Wielens et al. (2011) conducted a fragment-screening campaign using STD NMR to identify a total of 67 compounds from a 500-compound Maybridge Ro3 library against the HIV-1 IN CCD construct. Fragment binding of each hit was subsequently characterized using ^{15}N HSQC NMR and X-ray crystallography. Since the crystallization condition for these constructs consisted of sodium acetate buffer and ammonium sulfate, covalent attachment of a buffer molecule such as cacodylate to the cysteine residues was not a concern. A concentration of 20 mM was found to be sufficient for fragment soaking. However, subsequent follow-up studies for larger compounds, required dry compound to

be “sprinkled” into crystallization drops containing pre-formed crystals of an alternate CCD construct (residues 50-212 with Cys56Ser, Phe139Asp, and Phe185His mutations) and the cryosolution (Rhodes et al., 2011; Peat et al., 2012). This suggests that IN CCD simply may not be ideal for crystallographic fragment screening, resulting in a higher percentage of false negatives. The need for high concentration soaks also indicates that the intrinsic binding ability of the CCD is quite poor.

4.8.4 Solvent Channels

Soaking requires fragments to diffuse through solvent channels without destroying the crystal lattice. Fragment binding can be hindered if these channels are not sufficiently large enough to permit fragments to travel. The diameter of the narrowest solvent channel was determined using MacPyMol (Schrödinger, New York, NY) for the three crystal forms screened to explore the correlation, if any, with suitability of a crystal form for fragment screening. Symmetry mates were generated within 50 Å for each crystal form and a pseudoatom was generated at the center of the solvent channel. The diameter was estimated by altering the van der Waals radius until the channel was filled.

No definite correlation could be established between the diameter of the solvent channel and suitability of a crystal form for screening. Drastically different results were obtained from fragment screening against both PA_N and CCD, which have similar diameters (Table 8). At the same time, fragment binding was readily observed within the crystal form for the RT-rilpivirine complex, which has the narrowest channel. Thus, the effect of the solvent channel diameter on the success of a fragment-screening campaign remains ambiguous. Davies et al. (2011) also reached a similar conclusion using a dataset of 18 different targets that they had found to be either productive or non-productive for

crystallographic fragment screening. Majority of the protein targets from their datasets were found to have solvent channels of 20 $Å^2$ or higher. It is possible that a much larger dataset may reduce the ambiguity.

Protein	Space Group	Estimated Diameter of Solvent Channel (Å)
RT-TMC278	C2	24
pH1N1 PA_N	$C222_1$	32
IN CCD	$P312_1$	32

Table 8: Estimated diameter of the solvent channels for the crystal forms used for fragment screening by X-ray crystallography.

4.8.5 Solvent Woes

In addition to solvent channels, solvent displacement could also be yet another factor to consider. Water molecules must be displaced to allow for the diffusion of a fragment through the crystal, resulting in a thermodynamic penalty from the desolvation. Presumably, the increase in entropy from the displacement of water or solvent molecules should compensate for the enthalpic penalty acquired from the loss of interactions with other solvent molecules and protein. However, it is possible that favorable entropic contribution may not be sufficient to counter the enthalpic penalty incurred by the displacement of highly conserved water molecules. Difficulties displacing these conserved water molecules can reduce the size of a solvent channel, and thereby restrict the diffusion of fragments based on their size.

Highly conserved water molecules, crystallization additives, and/or solvent used to solubilize fragments can bind to promising druggable sites and pose as an additional challenge for fragment binding. Higher fragment concentrations are often used to “flood” the sites and displace these molecules. However, this may not be entirely successful as the binding energy may not be sufficient enough to compensate for the thermodynamic

penalty of solvent displacement. Using a higher fragment concentration for soaking proved to be beneficial for identifying the binding mode of **55** to a site made by crystallographic contacts and, more importantly, the allosteric LEDGF/p75-binding site. However, this approach was not successful for **56** and **63**, which were found to have μM inhibitory activity in the enzymatic assays.

4.9 Conclusion

Only one fragment was identified from a library of 971 compounds screened against HIV-1 IN CCD using X-ray crystallography. Although a 20 mM cocktail soak revealed fragment binding to a site formed by crystal contacts, a high concentration soak of the individual fragment (**55**) revealed binding to a secondary, biologically relevant allosteric site commonly referred to as the LEDGF/p75-binding site. **55** and a selection of chemically-related compounds within our fragment library were assayed for inhibitory activity. While **55** was not found to be inhibitory at the tested concentrations of 200 and 400 μM, **56** and **56**, however, were found to have promising inhibitory activity. Subsequent crystallography using both soaking and co-crystallization with these two fragments has proven to be challenging. Biological characterization is underway to ensure that these two fragments are not false positives. In parallel, fragment expansion of **55** is currently underway using docking with AutoDockVina and synthesis of promising derivatives and analogs.

4.9 Materials and Methods

4.9.1 Crystallography

HIV-1 IN Phe185Lys CCD (residues 55-208) at 5 mg/mL was mixed in equal volumes protein and a seed stock, consisting of microseeds of HIV-1 IN CCD diluted in

the crystallization solution containing 1 mM manganese chloride, 100 mM MES pH 6.7, 10% (w/v) PEG 8000, and 5 mM DTT. Crystallization trays were prepared at room temperature then transferred to 4 °C for storage.

For fragment screening, cocktails of 4-8 compounds were used for the initial screening with single compound soaks to verify the bound fragment identity and to rule out cooperative binding. Fragments, dissolved in DMSO, were soaked into crystals (final concentration 20 mM and 120% (v/v) DMSO) for 2 hours before flash freezing in liquid nitrogen. X-ray diffraction data collection was performed at the Cornell High Energy Synchrotron Source (CHESS) F1 beamline, and the National Synchrotron Light Source (NSLS) beamline X25.The diffraction data were indexed, processed, scaled, and merged using *HKL2000* (Otwinowski et al., 2011). Structure refinement was carried out using PHENIX (1.8.2-1309) (Adams et al., 2010) and COOT (0.7-rev4459) (Emsley et al., 2010) with riding hydrogens present when the X-ray diffraction resolution is better than or equal to 1.9 Å. Current crystallographic and refinement statistics for IN CCD crystals soaked with 20 mM PD1d and 100 mM MB36-3 are described in Table 9 below.

	20 mM PD1d	100 mM 55
X-ray source	CHESS F1	CHESS F1
Wavelength (Å)	0.917	0.917
Space group	$P3_121$	$P3_121$
Cell constants: a, b, c (in Å)	72.43, 72.43, 66.56	71.91,71.91, 67.68
α, β, γ (in °)	90, 90, 120	90, 90, 120
Resolution range (Å)	1.40-50.00	1.70-50.00
Completeness (%)	99.7	98.5
R_{merge}	0.038	0.078
Average I/σ (*I*)	45.67	16.77
Refinement Statistics		
No. of reflections	40,045	42,443
R_{work}	0.1710	0.2020
R_{free}	0.1916	0.2227
Ramachandran statistics (% of residues in favored/disallowed regions)	97.92/0.00	96.60/0.00
RMSD bond length (Å)	0.012	0.008
RMSD bond angles (°)	1.563	1.399

Table 9: Crystallographic and refinement statistics for HIV-1 Phe185Lys IN CCD crystals soaked with PD1d and **55**.

4.9.2 Docking

The crystal structure of **55** bound to IN CCD was used for docking. The biologically relevant CCD dimer was generated by merging the X-ray crystallographic monomer with its symmetrical molecule using COOT (Emsley et al., 2004). All solvent and ligand molecules were removed. Chimera 1.5.2 (Pettersen et al., 2004) was used to add hydrogens to the protein. Then, the AutoDock Tools (ADT) (Morris et al., 2009) suite was used to prepare AutoDock pdbqt files for docking by merging the non-polar hydrogens, adding Kollman United Atom charges, and assigning AD4 atom types.

Phenix Elbow (Adams et al., 2010) was used to generate the 3D coordinates of ligands from the SMILES. Subsequent optimization of the geometry using the Universal Force Field (UFF) was done using Avogadro (Hanwell et al., 2012). The input pdbqt files were generated for the ligands using ADT (Morris et al., 2010) by assigning Gasteiger Marsili charges to the ligands and all torsions were treated flexible.

A search space for docking using AutoDock Vina (Trott and Olson, 2010) was limited to 25 x 25 x 25 Å to encompass the entire site. The exhaustiveness level was set to 12 to allow for enhanced conformational sampling and the repulsion term was scaled to 3.0. A total of 9 binding modes were generated for each ligand. Docking conditions were validated by self-docking experiments of **55** to this site.

Chapter 5: Fragment Screening by X-ray Crystallography—A Reflection

While science may be highly divergent with respect to focus, theory, and technique, the importance of experimental design remains a fundamental common ground for all basic sciences. Similarly, the successful implementation of fragment screening by X-ray crystallography requires careful consideration of all known experimental parameters, such as protein crystals, fragment or cocktail preparation, soaking conditions, and, most importantly, high throughput. Each of these parameters will be discussed in detail in the following sections. Additionally, this chapter will also highlight some of the key advantages for using X-ray crystallographic approach for fragment screening. The unpredictability of this approach with respect to varying hit rates will also be briefly discussed.

5.1 Experimental Considerations

5.1.1 Protein Crystals

The success of a fragment-screening campaign using X-ray crystallography is heavily dependent on the protein crystal used for screening purposes. Based on our screening experiences as well as others in the field, crystallographic fragment screening requires crystals in which the protein is in a biologically relevant conformation. This establishes a correlation with the enzymatic assays that will be used to characterize the inhibitory activity of the fragment hits and allows for rapid evolution of a fragment hit to a lead candidate using structure-based drug design. Additionally, druggable sites within the protein must not be occluded by either protein-protein crystal contacts, natural substrates, and chemicals used for crystallization, soaking, and/or cryoprotection.

Fragment screening by X-ray crystallography requires large numbers of highly reproducible crystals, which enables rapid data collection and screening. Thus, crystal optimization is essential for initiating a fragment-screening campaign. All three screening campaigns discussed in this thesis required some sort of crystal optimization. Crystal engineering was used to obtain suitable protein constructs for both HIV-1 RT and pH1N1 PA_N. With respect to HIV-1 RT, crystallization was enhanced by introducing point mutations and C-terminal truncations to reduce the surface entropy of the protein (Bauman et al., 2006). Similarly, a high-resolution construct of pH1N1 PA_N was obtained by the truncation and mutagenesis of a H5N1 construct (Bauman et al., 2013). On the other hand, a single point mutation, Phe185Lys, had been previously introduced into HIV-1 IN CCD to improve protein solubility and subsequent optimization of crystallization setup was conducted to obtain larger crystals to facilitate high-throughput handling.

In addition to being highly reproducible, the crystals should also be very robust to survive fragment soaking at high concentrations. This allows for rapid screening and data collection and reduces the amount of labor associated with this approach. Either alternative crystal forms or co-crystallization will be necessary if the crystals are either too fragile to handle or behave poorly during soaking. This was the case during the fragment screening campaign against HIV-1 RT. Originally, a high-resolution construct of the apo HIV-1 RT (RT69A) was selected for the screening campaign. However, the RT69A crystals were not able to withstand the soaking conditions, resulting in a decrease in the throughput of the method. The poor stability of these crystals was attributed to fragment binding to the NNRTI-binding pocket,

which requires large conformational changes to form resulting in collapse of the crystal lattice. We were able to bypass this problem by using the NNRTI rilpivirine to block access to the NNRTI-binding pocket and limit protein flexibility; the RT52A construct in complex with rilpivirine routinely diffracts to 1.8-2.0 Å resolution and sometimes as high as 1.5 Å resolution (Kuroda et al., 2013).

5.1.2 Fragment Solubility

Regardless of the screening technique, fragment solubility continues to be a major challenge for fragment screening. Careful solvent selection is needed to ensure the chemical integrity of the fragment is not affected. The majority of the fragments are hydrophobic in nature and favor organic solvents, such as DMSO. However, since crystallographic fragment screening is conducted at high fragment concentrations, it is necessary to consider crystal stability at high solvent concentrations.

In our case, both HIV-1 RT-rilpivirine and IN CCD crystals were stable in 20% (v/v) DMSO concentration, whereas the pH1N1 PA_N crystals were not. Thus, fragment screening for PA_N was conducted with the fragment concentration at 10 mM and 10% (v/v) DMSO concentration. It is also important to note that while HIV-1 RT-rilpivirine and IN CCD crystals were able to tolerate 20% (v/v) DMSO, soaking times for each crystal were significantly different. Crystals of the HIV-1 RT-rilpivirine complex were only able to tolerate 20% (v/v) DMSO for a maximum of 20 minutes, whereas the HIV-1 IN CCD crystals were sturdy enough to handle overnight soaks at the same DMSO concentration.

Fragment stability within the actual soaking solution also needs to be taken into consideration. It is imperative that the fragments do not chemically react with the

various components of the crystallization buffer. Chemical reactivity may often necessitate the use of a less favored crystal form or cause the omission of a particular fragment.

Fragment solubility within the soaking solution proved to be quite challenging. During our screening campaign using the HIV-1 RT-rilpivirine crystals, L-arginine was found to be an excellent solubilizing agent after several cocktails revealed fragment binding only in the presence of L-arginine. Fortuitously, 80 mM L-arginine was sufficient in improving the solubility of the cocktail/fragment within the soaking solution for all three campaigns. However, the concentration of L-arginine was carefully optimized for each screening campaign conducted.

5.1.3 Cocktail Design

To improve the throughput, fragment-screening campaigns with X-ray crystallography use cocktails containing an average of 4-7 compounds. Thus, cocktail design is yet another important aspect that requires careful thought and approach, as these cocktails will be routinely used for screening a wide range of protein targets.

Cocktail design must take into consideration solubility and chemical reactivity between fragments. In the early days of FBDD, the classic “trial and error” approach was used to manually design cocktails. Today, computational programs using a filter approach are available to facilitate rapid cocktail design, however predictions are often poor. Despite careful cocktail design, lack of chemical reactivity can only be determined empirically. In our case, several cocktails had to be redesigned because a color change was noted upon preparation. Routine quality control analysis of the individual fragments and cocktails should be performed to ensure compound stability.

As discussed in Chapter 1, the screening technique needs to be taken into consideration during cocktail design. Using X-ray crystallography as a primary screening technique requires cocktails consisting of structurally diverse fragments to minimize false negatives. The importance of cocktail design was demonstrated by one of the cocktails (Figure 55) from our fragment library. Soaking this cocktail into a HIV-1 RT-rilpivirine crystal revealed electron density for a fragment near the RNase H region. Subsequent soaking with the individual fragments of this cocktail revealed clear electron density for two fragments (**76** and **81**) bound to the same site. Both molecules share not only similar shape but also a similar chemical scaffold. The electron density observed in the cocktail soak was highly fortuitous as both of these fragments could bind competitively, leading to false negatives.

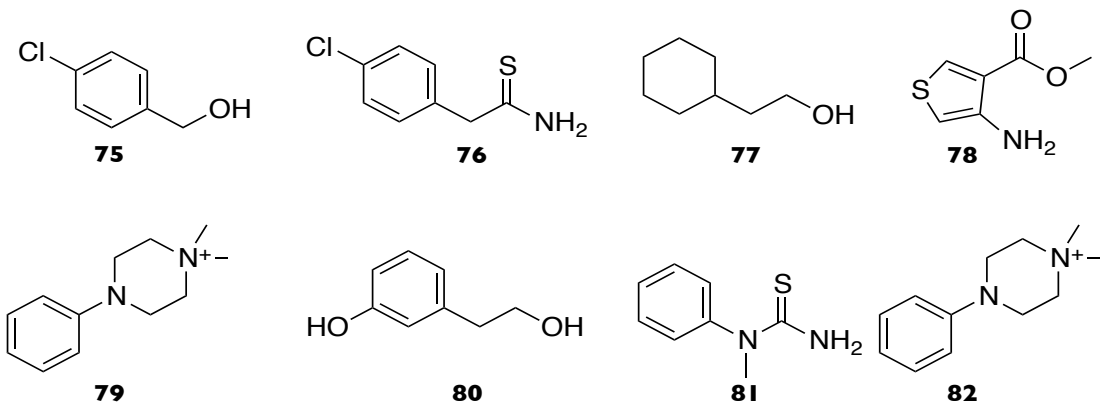


Figure 55: Importance of cocktail design. Individual components of a particular cocktail used for screening.

5.1.4 Quality Control

In the early days of FBDD, quality control was an afterthought as the purchased fragments are of nominally high purity. However, after several painful experiences of wasted time and resources pursuing a contaminant, rigorous quality control of fragments and their cocktails has become of upmost importance. Like

many of our counterparts in industry and academia, we have learned the importance of quality control the hard way. In our case, electron density for a fragment, **83**, at the RNase H region was found upon soaking into crystals of HIV-1 RT-rilpivirine. However, despite the strong electron density, modeling the compound into this density proved to be challenging. Enzymatic assays by collaborator Michael Parniak revealed that **83** had an IC_{50} value of 14 μM. The promising inhibitory data fueled extensive efforts to clearly characterize the structure of the bound ligand by both soaking and co-crystallization with **83**. An additional quantity of **83** was purchased from the original vendor to support this crystallographic effort. Surprisingly, with the new lot of **83**, only weak fragment binding was observed upon soaking and a ten-fold loss in inhibitory activity (Figure 56). HPLC analysis of both the original and new lots of **83** revealed a major peak for the compound and several minor peaks representing contaminants. Soaking purified **83** into crystals of RT52A-rilpivirine revealed poor electron density indicating the electron density originally observed was most likely due to a contaminant. Further characterization of the purified contaminants was not pursued due to a limited supply of the original fragment. The amount of time and resources spent pursuing this fragment hit and its derivatives emphasized the importance of quality control.

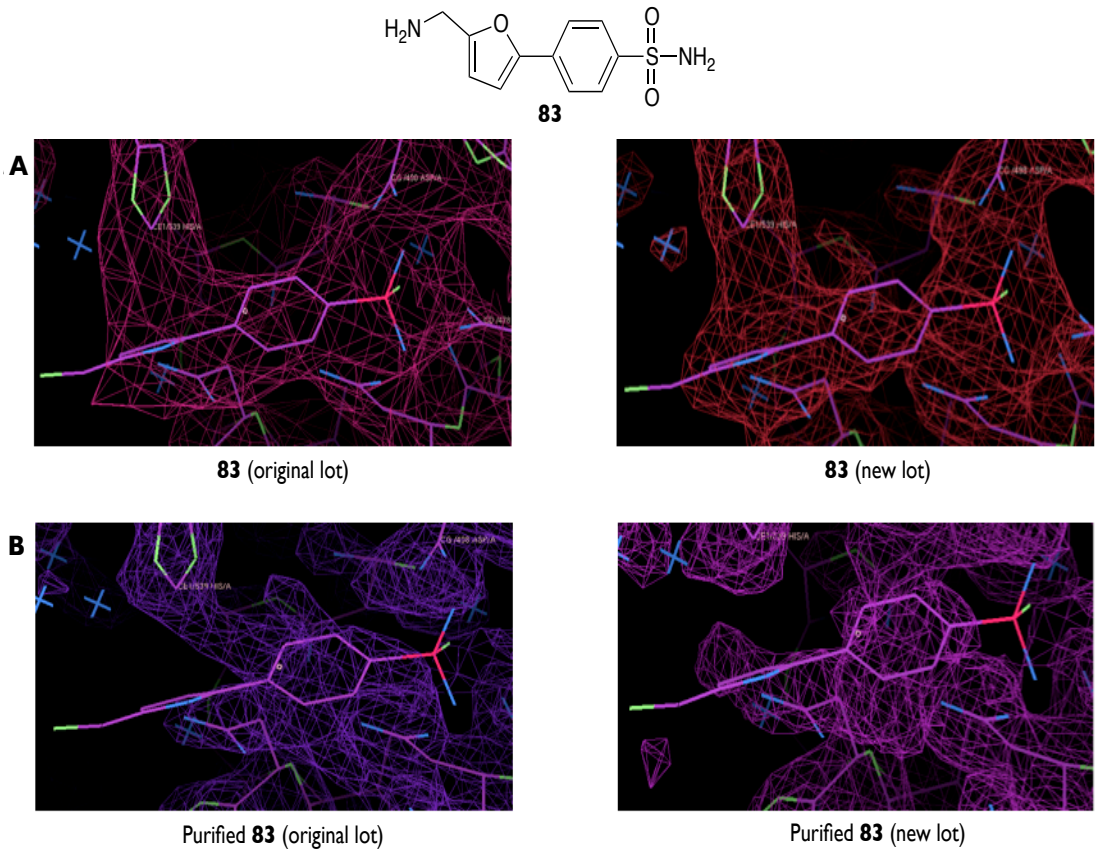


Figure 56: Importance of compound purity. **A.)** Electron density for binding of **83** from the original and new lots. **B.)** Electron density for the binding of purified **83** for the original and new lots.

5.1.5 Development of High-Throughput Screening Methodology

Limitations in high throughput combined with the amount of labor associated with X-ray crystallography has restricted its application as a primary screening technique for fragment screening both in academia as well as in industry. However, recent technological advancements in crystal freezing and synchrotron resources have greatly minimized the amount of labor and improved the throughput associated with X-ray crystallography. A cassette or puck system, which facilitated the transport of larger quantities of crystals to the synchrotron, was used during the course of all three screening campaigns. A total of seven pucks with each puck single capable of holding 16 crystals can be transported in a CP100 Dewar. This allows for the storage

of 112 crystals per dewar instead of the traditional cane transport system, which only allows for 56 crystals to be stored. Use of the puck system significantly reduced ice build-up around the crystals, thus reducing the time spent at the beam rinsing the crystals with liquid nitrogen. Compared to the cane system, the pucks allowed for rapid freezing of the crystals and reduced the handling error that is associated with crystal handling and screening of multiple cocktails in one sitting.

The compatibility of the puck system with the robotic crystal mounting system, commonly referred to as automounters, available at both CHESS F1 and Brookhaven X25 and X29 beamlines enabled rapid data collection by reducing the time, labor, and the risk associated with manual mounting and dismounting of a crystal. The use of the automounter and availability of crystallographic software at the beamlines facilitated the use of the low-resolution data collection strategy for initial screening. As described in Chapter 2, the low-resolution data collected for a crystal was quickly processed using HKL2000 and then CNS was used to generate F_o-F_o maps from the processed dataset to identify fragment binding. Crystals, for which electron density indicated fragment binding, were re-mounted for a high-resolution data collection. This data collection strategy would not have been successful without the use of the automounter, which significantly reduces the risk and time associated with manual handling of the crystals.

Advancements in detector technology also improved the efficiency and high throughput of fragment screening by X-ray crystallography. For instance, the time spent for data collection can be reduced using the Pilatus 6M detector at Brookhaven X25 beamline. The short readout and fast framing times associated with this detector

allows for collection of diffraction data in fine-phi slicing mode with continuous rotation of the crystal without opening and closing of the shutter after each frame. Use of the Pilatus 6M detector reduced the average time for a high-resolution pass for a PA_N endonuclease crystal from six minutes to two minutes. In combination with the automounter, the Pilatus 6M detector allowed for a peak data collection rate of 15 data sets per hour with approximately half of the time used for mounting/unmounting and crystal centering. Thus, complete datasets for a total of 112 PA_N crystals can be collected within a 12-hour X-ray beam allocation.

In short, the technological advancements with regards to crystal transport and resources available at the synchrotrons significantly reduced the amount of labor and time that is typically associated with X-ray crystallography. This combined with an efficient data collection and processing strategy improved the throughput of this approach.

5.2 Advantages of Crystallographic Fragment Screening

5.2.1 Protein Flexibility and Allostery

Lack of protein flexibility has often been cited as a major drawback for using X-ray crystallography as a primary screening technique. It is important to note, while the formation of a crystal requires the protein to adopt a particular conformation, the protein still possesses substantial degrees of flexibility in the crystalline crystal form. This was demonstrated during the by the X-ray crystallographic fragment-screening campaign against HIV-1 RT. Here, not only fragments bound to existing crevices but also successfully trapped previously unseen protein conformations of RT, as

evidenced by the discovery of fragment binding to the Knuckles and NNRTI Adjacent sites.

In particular, the Knuckles pocket, which is normally a buried cavity within previously reported RT structures, was successfully characterized within the RT-TMC278 crystal form despite a 1.2 Å shift in the fingers subdomain (Figure 57). Furthermore, Knuckles binders bound readily to this pocket even after a 10-second fragment soak during a time-based series of soaking experiments with intervals ranging from 10 seconds to 30 minutes. This and many other experiments have indicated that proteins can be surprisingly dynamic within crystals. Furthermore, fragment binding can stabilize the conformation of one of these dynamically accessible states and when coupled with X-ray crystallography, these alternative conformations can also successfully be characterized.

Figure 57: Superposition of RT52A-ripilvirine structures with (in orange) and without (in green) **2** (shown in spheres) bound to the Knuckles pocket.

5.2.2 X-ray Crystallography and Compounds Binding to Multiple Sites

Typical fragment-screening campaigns using biophysical screening techniques, such as NMR and SPR, often discard molecules identified to bind non-specifically to a protein target in favor for a specific binder. Frequently it is believed these promiscuous binders are a result of compound aggregation rather than actual potency. Thus, it is possible that viable, more attractive fragment hits, particularly those that act via an allosteric mechanism, are believed to be false positives and discarded. To a certain degree, using X-ray crystallography as the primary screening technique can reduce occurrence of this type of false positive. Structural characterization of the promiscuous fragment bound to the target protein allows for rapid refinement of molecular specificity by making and testing derivatives guided by the knowledge of the individual binding modes even when multiple copies of the initial fragment are bound. This has successfully been demonstrated in the case of influenza endonuclease.

Of the three proteins crystallographically screened, the H1N1 PA_N was found to have the most non-specifically bound fragments. In particular, a promising fragment, hydroxypyridinone, was pursued for lead development despite the multiple binding modes observed. As described in Chapter 3, three molecules of hydroxypyridinone occupied the active site in a manner that allowed for interaction with key residues and chelation to the active-site metals, including a previously unseen third metal ion. Fragment expansion guided by docking against the available protein structure allowed for efficient exploration of the chemical modification space of the initial fragment hit to enhance both potency and specificity. Structure-based

drug design efforts ultimately produced **54**, an inhibitor of influenza endonuclease with significant antiviral activity.

5.3 Biological Insight and Fragment Screening

Fragment screening has always been associated with drug discovery and the follow-up chemistry and drug effort. This proved to be true in our case, specifically when discussing the fragment screening work conducted for H1N1 PA_N. More surprising, however, was the insight gained about the biology of a particular protein and more specifically its interaction with a drug molecule through fragment screening, specifically with respect to HIV-1 RT and pH1N1 PA_N.

5.3.1 Importance of Water Molecules

During the course of the fragment screening campaign against the complex of a crystal-engineered variant of HIV-1 RT in complex with rilpivirine, soaking of several fragments enhanced the resolution of the crystals. In particular, soaking with 6% (w/v) trimethylamine-N-oxide (TMAO) improved the resolution of the RT52A-rilpivirine crystals from 1.80 Å to 1.51 Å (Figure 58).

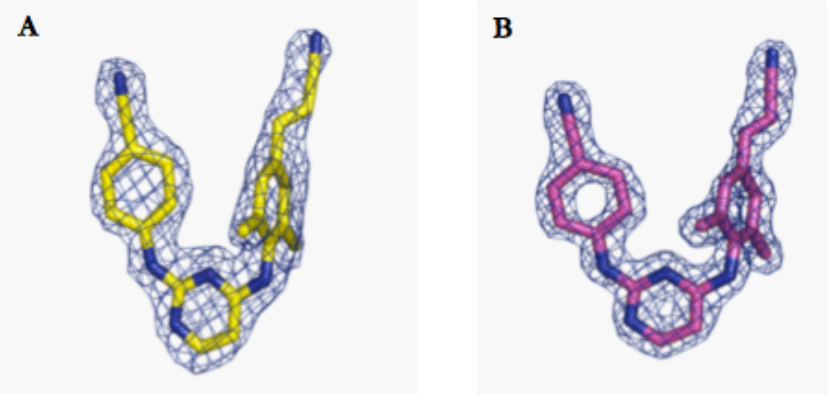


Figure 58: Comparison of electron density for rilpivirine at resolutions of 1.80 Å (panel A) and 1.51 Å (panel B).

Several studies have characterized the effect of TMAO on protein folding and stability (Yancey and Somero, 1979; Lin et al., 1994; Jaravine et al., 2000). Yancey et al. (1982) showed that TMAO not only promotes protein folding into more compact forms but also promotes folding into more biologically relevant forms. This was later attributed to the destabilization of the denatured state due to the unfavorable interactions with TMAO and the protein backbone (Wang et al., 1997). Molecular dynamics simulations conducted by Zou et al. (2002) revealed that TMAO acts as a protein stabilizer by enhancing the water structure. Thus, it is believed that the improvement in the resolution quality of the HIV-1 RT52A-ripilvirine crystals is the by-product of an enhancement in the water structure within the crystal by TMAO.

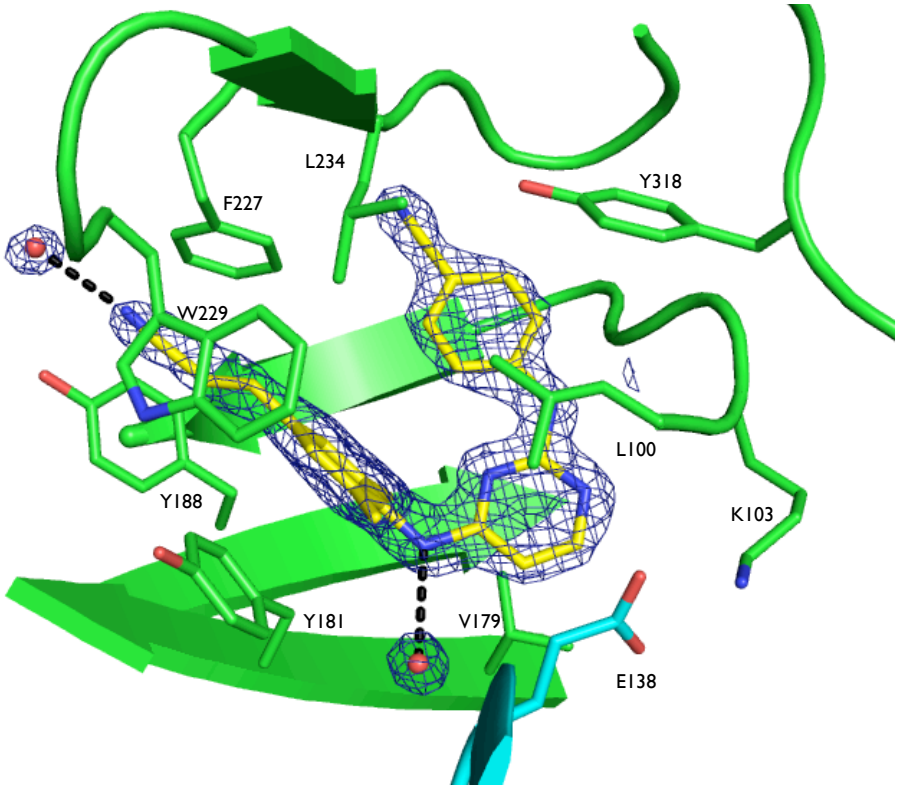


Figure 59: High-resolution crystal structure of RT52A-rilpivirine with the previously undetected water molecule (upper left corner) crucial for the high inhibitory potency of rilpivirine (in yellow). The p66 subunit is depicted as green and the p51 subunit as cyan.

Although no fragment binding was observed from the 6% (v/v) TMAO soak, the higher-resolution structure revealed the presence of a water molecule near the cinnamonitrile of rilpivirine (Figure 59). A combination of two-dimensional vibrational echo spectroscopy and molecular dynamics separately identified this water molecule to be an important contributing factor to the observed inhibitory activity of rilpivirine against wild-type and mutant HIV-1 RT (Kuroda et al., 2012). Here, the two-nitrile substituents, one on each arm of rilpivirine, were used as vibrational probes to explore the dynamic nature of the NNRTI-binding pocket with respect to the "wild-type" RT52A and double-mutant variants, RT51A (Leu100Ile/Lys103Asn) and RT55A (Lys103Asn/Tyr181Cys).

A combination of 2D-IR and molecular dynamics experiments revealed that the cinnamonitrile functionality of rilpivirine formed a hydrogen bond with a mobile water molecule, previously unseen in the 1.8 Å HIV-1 RT52A-rilpivirine structure. At the time, the experimental results from the 2D-IR experiments and molecular-dynamics simulations and the structural results for this complex were not concordant. However, improvement in the resolution of the HIV-1 RT52A-rilpivirine crystals after a short soak in 6% (w/v) TMAO revealed that this water molecule was present within the structure and formed a crucial hydrogen bond with the cinnamonitrile functionality as predicted by the 2D-IR experiments.

Identification of TMAO as a useful crystallographic additive has some precedent. TMAO has been proven to be a useful cryoprotectant at high concentrations, specifically 4 M (Mueller-Dieckmann et al., 2011). Our soaking experiment revealed that lower concentrations of TMAO could improve the

resolution quality of crystals. This beneficial property was not limited to the RT52A-ripilvirine crystals. Enhancement in diffraction resolution was also observed for crystals for RT-nucleic acid complexes (Steven Tuske, personal communication) and pH1N1 PA_N.

Despite its favorable properties, fragment soaking in the presence of TMAO proved to be deleterious. Fragments previously identified to bind to HIV-1 RT, no longer bound with the addition of TMAO in the soaking solution due to poor fragment solubility. Presumably, this is due to the incompatible nature of the hydrophobic nature of the fragments and the charged nature of TMAO. This was substantiated by visual precipitation of fragment/cocktail soaking solutions with TMAO.

5.3.2 Functional Studies Using Fragments: Third Metal Binding

Crystal structures for four out of the eight fragment hits identified from a screening campaign against H1N1 influenza endonuclease revealed the presence of a previously unseen third metal ion within the active site. This was surprising, as previous structural and calorimetric studies have established a two-metal ion binding model. Subsequently, metal ions (Mn^{2+}, Zn^{2+}, Ca^{2+}, and Ni^{+2}) at high concentrations were respectively soaked into EDTA-treated PA_N crystals. Although soaking of Mn^{+2}, Zn^{+2}, and Ni^{+2} ions proved to be inconclusive due to loss of X-ray diffraction resolution, the Ca^{+2} soak revealed clear electron density for M2 and M3 ions coordinating with water molecules and residues, as seen in the fragment bound structure. Thus, the presence of the third metal ion is not purely a by-product of fragment binding and may have functional relevance.

Three-metal ion coordination in related nucleases is not entirely a novel concept. Other endonucleases, such as T5 flap endonuclease IV, and *Eco*RV, are believed to have three metal ions involved in the catalytic mechanism. A combination of structural and molecular modeling suggests that, in case of *Eco*RV, that the third metal ion stabilizes the negative charge of the transition state and is also involved in the recruitment of a water molecule to protonate the leaving group during catalysis. Although this does not necessarily have to hold true for influenza endonuclease, presence of the third metal in our structures and similar endonucleases has stimulated further exploration regarding the importance of the third metal ion on the enzymatic activity of PA_N.

5.4 Conclusion

We have successfully demonstrated X-ray crystallography to be a powerful technique for identifying promising chemical scaffolds using the fragment-based approach. Compared to other screening techniques, such as NMR and SPR, the amount of time, high throughput, and labor associated with X-ray crystallography has limited its application as a primary screening method. However, by taking advantage of the technological improvements at synchrotrons and crystal handling combined with an efficient strategy for rapid hit identification, we were able to improve the throughput and significantly reduce the labor associated with this screening method.

Crystallography proved to be highly advantageous as the structural characterization of the bound fragment greatly facilitated rapid SBDD. With respect to pH1N1 endonuclease, the availability of the crystal structure proved to be invaluable for the development of a non-specifically bound fragment into a new class

of endonuclease inhibitors with antiviral activity. In addition, crystallography permitted detection and characterization of potentially druggable allosteric sites in HIV-1 RT and IN CCD. Crystallographic fragment screening also provided novel biological insights with regard to binding of the third metal ion within the active site of PA_N and the important contribution of a water molecule to the HIV-1 RT inhibitory activity of rilpivirine.

References

2009 HIV sequence database, Los Alamos National Laboratory [Online]. Available: www.hiv.lanl.gov (accessed April 2011).

Abbondanzieri, E. A., Bokinsky, G., Rausch, J. W., Zhang, J. X., Le Grice, S. F. J., Zhuang, X. (2008) Dynamic binding orientations direct activity of HIV reverse transcriptase. *Nature*, **453**, 184-189.

Adams P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallography D. Biological Crystallography*, **66**, 213–221.

Aiken, C., Konner, J., Landau, N. R., Lenburg, M. E., Trono, D. (1994) Nef induces CD4 endocytosis: requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain. *Cell*, **76**, 853-864.

Aiken, C. (2006) Viral and cellular factors that regulate HIV-1 uncoating. *Current Opinion HIV AIDS*, **1**, 194-199.

Al-Mawsawi, L. Q., Christ, F., Dayam, R., Debyser, Z., Neamati, N. (2008) Inhibitory profile of a LEDGF/p75 peptide against HIV-1 integrase: insight into integrase-DNA complex formation and catalysis. *FEBS* Letters, **582**, 1425-1430.

Andrews, P. R., Craik, D. J., Martin, J. L. (1984) Functional group contributions to drug-receptor interactions. *Journal of Medicinal Chemistry*, **27**, 1648–1657.

Ao, Z. Huang, G., Yao, H., Xu, Z., Labine, M., Cochrane, A. W., Yao, X. (2007) Interaction of human immunodeficiency virus type 1 integrase with cellular nuclear import receptor importin 7 and its impact on viral replication. *Journal of Biological Chemistry*, **282**, 13456-13467.

Aravind, L., Landsman, D. (1998) AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Research*, **26**, 4413–4421.

Auewarakul, P., Wacharapornin, P., Srichatrapimuk, S., Chutipongtanate, S., Puthavathana, P. (2005) Uncoating of HIV-1 requires cellular activation. *Virology*, **337**, 93-101.

Babaoglu, K., Shoichet, B. K. (2006) Deconstructing fragment-based inhibitor discovery, *Nature Chemical Biology*, **2**, 720-723.

Baell, J. B. (2013) Broad Coverage of Commercially Available Lead-like Screening Space with Fewer than 350,000 Compounds. *Journal of Chemical Informatics and Modeling*, **53**, 39-55.

Baell, J. B., Holloway, G.A. (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *Journal of Medicinal Chemistry*, **53**, 2719–2740.

Baker, S. F., Guo, H., Albrecht, R. A., Garcia-Sastre, A., Topham, D. J., Martinez-Sobrido, L., (2013) Protection against lethal influenza with a viral mimic. *Journal of Virology*, **87**, 8591-8605.

Balzarini, J. (2004) Current status of non-nucleoside reverse transcriptase inhibitors of human immunodeficiency virus type 1. *Current Topics in Medicinal Chemistry*, **4**, 921-944.

Bell, N. M., Lever, A. M. L. (2012) HIV Gag polyprotein: processing and early viral particle assembly. *Trends in Microbiology*, **21**, 136-144.

Baranova, S., Tuzikov, F. V., Zakharova, O. D., Tuzikova, N. A., Calmels, C., Litvak, S., Tarrago-Litvak, L., Parissi, V., Nevinsky, G. A. (2007) Small-angle X-ray characterization of the nucleoprotein complexes resulting from DNA-induced oligomerization of HIV-1 integrase. *Nucelic Acids Research*, **3**, 975-987.

Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., xler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W., Montagnier, L. (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, **220**, 868-871.

Baughman, B. M., Slavish, P. J., DuBois, R. M., Boyd, V. A., White, S. W., Webb, T. R. (2012) Identification of Influenza Endonuclease Inhibitors Using a Novel Fluorescence Polarization Assay. *ACS Chemical Biology*, **7**, 526–534.

Bauman, J. D., Das, K., Ho, W. C, Baweja, M., Himmel, D. M., Clark, A. D., Oren, D. A., Boyer, P. L., Hughes, S. H., Shatkin, A. J., Arnold, E. (2008) Crystal engineering of HIV-1 reverse transcriptase for structure-based drug design. *Nucleic Acids Research*, **36**, 5083-5092.

Bauman, J. D., Patel, D., Baker, S., Vijayan, R. S. K., Xiang, A., Parhi, A., Martinez-Sobrido, L., LaVoie, E. J., Das, K., Arnold, E. (2013) Crystallographic Fragment Screening and Structure-Based Optimization Yields a New Class of Influenza Endonuclease Inhibitors. *ACS Chemical Biology*, in press.

Beaton, A. R., Krug, R. M. (1981) Selected host cell capped RNA fragments prime influenza viral RNA transcription in vivo. *Nucleic Acids Res*. **9**, 4423–4436.

Berthoux, L., Sebastain, S., Muesing, M. A., Luban, J. (2007) The role of lysine 186 in HIV-1 integrase mutlimerization. *Virology*, **364**, 227-236.

Blaney J., Nienaber V., Burley S. K. Fragment-based lead discovery and optimisation using X-ray crystallography, computational chemistry, and high-throughput organic synthesis. W. Jahnke, D. A. Erlanson (Eds.), Fragment-based Approaches in Drug Discovery. Methods and Principles in Medicinal Chemistry, (2006) Wiley–VCH: Weinheim, Germany, 215-248.

Bodewes, R., Morick, D., de Mutsert, G., Osinga, N., Bestebroer, T., van der Vliet, S., Smits, S. L., Kuiken, T., Rimmelzwaan, G. F., Fouchier, R. A., Osterhaus, A. D. (2013) Recurring influenza B virus infections in seals. *Emerging Infectious Disease*, **19**, 511-522.

Boodor, K., Boyapati, V., Gopu, V., Boisdore, M., Allam, K., Miller, J., Treleaven, W. D., Weldeghiorghis, T., Aboul-ela, F. (2009) Design and implementation of an ribonucleic acid (RNA) directed fragment library. *Journal of Medicinal Chemistry*, **52**, 3753-3761.

Bohacek, R.S., McMartin, C., Guida, W.C. (1996) The art and practice of structure-based drug design: A molecular modeling perspective. *Medicinal Research Reviews*, **16**, 3-50.

Bojja, R. S., Andrake, M. D., Merkel, G., Weigand, S., Dubrack, R. L., Skalka, A. M. (2013) Architecture and assembly of HIV integrase multimers in the absence of DNA substrates. *Journal of Biological Chemistry*, **288**, 7373-7386.

Bour, S., Strebel, K. (2003) The HIV-1 vpu protein: a multifunctional enhancer of viral particle release. *Microbes and Infection*, **5**, 1029-1039.

Bouvier, N. M., Palese,P. (2008) The biology of influenza viruses. *Vaccine*, **26S**, D49-D53.

Bowerman, B., Brown, P. O., Bishop, J. M., Varmus, H. E. (1989) A nucleoprotein complex mediates the integration of retroviral DNA. *Genes & Development*, **3**, 469-478.

Brough, P. A., Aherne, W., Barril, X., Borgognoni, J., Boxall, K., Cansfield, J. E., Cheung, K. M., Collins, I., Davies, N. G., Drysdale, M. J., Dymock, B., Eccles, S. A., Finch, H., Fink, A., Hayes, A., Howes, R., Hubbard, R. E., James, K., Jordan, A. M., Lockie, A., Martins, V., Massey, A., Matthews, T. P., McDonald, E., Northfield, C. J., Pearl, L. H., Prodromou, C., Ray, S., Raynaud, F. I., Roughley, S. D., Sharp, S. Y., Surgenor, A., Walmsley, D. L., Webb, P., Wood, M., Workman, P., Wright, L. (2009) 4,5-diarylisoxazole Hsp90 chaperone inhibitors: potential therapeutic agents for the treatment of cancer. *Journal of Medicinal Chemistry*, **51**, 196-218.

Brunger, A. T., Adams, P. D., Clore, G. M., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., Warren, G. L. (1998) Crystallography & NMR system (CNS): a new software suite for macromolecular structure determination. *Acta Crystallography D. Biological Crystallography*, **54**, 905-921.

Bujacz, G., Alexandratos, J., Zhou-Liu, Q., Clement-Mella, C., Wlodawer, A. (1996) The catalytic domain of human immunodeficiency virus integrase: ordered active site in the F185H mutant. *FEBS Letters*, **398**, 175-178.

Buonaguro, L., Tornesello M. L., Buonaguro F. M. (2007) Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *Journal of Virology*, **81**, 10209-10219.

Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G.M., Gronenborn, A.M. (1997) Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nature Structure Biology*, **4**, 567–577.

Checkley, M. A., Luttge, B. G., Freed, E. O. (2011) HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. *Journal of Molecular Biology*, **410**, 582-608.

Chen, D., Menche, G., Power, T. D., Sower, L., Peterson, J. W., Schein, C. H. (2007) Accounting for ligand-bound metal ions in docking small molecules on adenylyl cyclase toxins. *Proteins*. **67**, 593-605.

Chen, J. C., Krucinski, J., Miercke, L. J., Finer-Moore, J. S., Tang, A. H., Leavitt, A. D., Stroud, R. M. (2000) Crystal structure of the HIV-1 integrase catalytic core and C-terminal

domains: a model for viral DNA binding. *Proceedings in National Academy of Sciences USA*, **97**, 8233-8238.

Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, E., De Clercq, E., Debyser, Z. (2003) HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *Journal of Biological Chemistry*, **278**, 372–381.

Cherepanov, P., Ambrosio, A. L., Rahman, S., Ellenberger, T., Engelman, A. (2005) Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proceedings in National Academy of Sciences USA*, **102**, 17308–17313.

Christ, F., Voet, A., Marchand, A., Nicolet, S., Desimmie, B. A., Marchand, D., Bardiot, D., Van der Veken, N. J., Van Remoortel, B., Strelkov, S. V., De Maeyer, M., chaltin, P., Debyser, Z. (2010) Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. *Nature Chemical Biology*, **6**, 442-448.

Cianci, C., Chung, T. D. Y., Meanwell, N., Putz, H., Hagen, M., Colonno, R. J., Krystal, M. (1996) Identification of N-hydroxamic acid and N-hydroxyimide compounds that inhibit the influenza virus polymerase. *Antiviral Chemistry Chemotherapy*, **7**, 353-60.

Cihlara, T., Rayb, A. S. (2010) Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine. *Antiviral Research*, **85**, 39–58.

Clancy, S. (2008) Genetics of the influenza virus. *Nature Education* 1, http://www.nature.com/scitable/topicpage/genetics-of-the-influenza-virus-716.

Congreve, M., Carr, R., Murray, C., Jhoti, H. (2003) A rule of three for fragment-based lead discovery? *Drug Discovery Today*, **8**, 876-877.

Congreve, M., Murray, C. W., Blundell, T. L. (2005) Structural Biology and Drug Discovery, *Drug Disc. Today*, **10**, 895-907.

Congreve, M., Chessari, G., Tisi, D., Woodhead, A. J. (2008) Recent developments in fragment-based drug discovery. *Journal of Medicinal Chemistry*, **51**, 3661-3680.

Crepín, T., Dias, A., Palencia, A., Swale, C., Cusack, S., Ruigrok, R. W. (2010) Mutational and metal binding analysis of the endonuclease domain of the influenza virus polymerase PA subunit. *J. Virol*. **84**, 9096–9104.

Dalvit, C., Flocco, M., Veronesi, M., Stockman, B. J. (2002) Fluorine-NMR competition binding experiments for high-throughput screening of large compound mixtures. *Combinatorial Chemistry High Throughput Screening*, 5, 605-611.

Dalvit, C. (2009) NMR methods in fragment screening: theory and a comparison with other biophysical techniques. *Drug Discovery Today*, **14**, 1051-1057.

Das, K., Clark, A. D., Lewi, P. J., Heeres, J., de Jonge, M. R., Koymans, L. M. H., Vinkers, H. M., Daeyaert, F., Ludovici, D. W., Kukla, M. J., De Corte, B., Kavash, R. W., Ho, C. Y., Ye, H., Lichtenstein, M. A., Andries, K., Pauwels, M., de Bethune, M. P., Boyer, P. L.,

Clark, P., Hughes, S. H., Janssen, P. A. J., Arnold, E. (2004) Roles of conformational and positional adaptability in structure-based design of TMC125-R165335 (etravirine) and related non-nucleoside reverse transcriptase inhibitors that are highly potent and effective against wild-type and drug-resistant HIV-1 variants. *Journal of Medicinal Chemistry*, **47**, 2550-2560.

Das, K., Lewi, P. J., Hughes, S. H., Arnold, E. (2005) Crystallography and design of anti-AIDS drugs: conformational flexibility and positional adaptability are important in the design of non-nucleoside HIV-1 reverse transcriptase inhibitors. *Progress in Biophysical Molecular Biology*, **88**, 209-231.

Das, K., Aramini, J. M., Ma, L., Krug, R. M., Arnold, E. (2010) Structures of influenza A proteins and insights into antiviral drug target. *Nature Structural & Molecular Biology*, **17**, 530-538.

Das, K. (2012) Antivirals targeting influenza A virus. *Journal of Medicinal Chemistry*, **55**, 6263-6277.

Davies T. G., Tickle I. J. (2012) Fragment screening using X-ray crystallography. *Topics in Current Chemistry*, **317**, 33-59.

Davies, D. R., Begley, D. W., Hartley, R. C., Staker, B. L., Stewart, L. J. (2011) Predicting the success of fragment screening by X-ray crystallography. *Methods in Enzymology*, **493**, 91-114.

Davis, B. J., Erlanson, D. A. (2013) Learning from our mistakes: the 'unknown knowns' in fragment screening, *Bioorganic & Medicinal Chemistry Letters*, **23**, 2844-2852.

de Bethune, M. (2010) Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery, development, and use in the treatment of HIV-1 infection: a review of the last 20 years (1989-2009). *Antiviral Research*, **85**, 75-90.

De Clerq, E. (2010) Antiretroviral drugs. *Current Opinion in Pharmacology*, **10**, 507-515.

Delelis, O., Carayon, K., Saib, A., Deprez, E., Mouscadet, J. (2008) Integrase and integration: biochemical activities of HIV-1 integrase. *Retrovirology*, **5**, 114.

di Marzo-Veronese, F., Copeland, T. D., DeVicol, A. L., Rahman, R., Oroszlan, S., Gallo, R. C., Sarngadharan, M. G. (1986) Characterization of highly immunogenic p66/p51 as the reverse transcriptase of HTLV-III/LAV. *Science*, **231**, 1289-1291.

Dias, A., Bouvier, D., Crepin, T., McCarthy, A. A., Hart, D. J., Baudin, F., Cusack, S., and Ruigrok, R. W. (2009) The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature*, **458**, 914–918.

Dietz F, Franken S, Yoshida K, Nakamura H, Kappler J, Gieselmann V. (2002) The family of hepatoma-derived growth factor proteins: characterization of a new member HRP-4 and classification of its subfamilies. *Biochemistry Journal*, **366**, 491-500.

Ding, J., Das, K., Tantillo, C., Zhang, W., Clark, A. D., Jr., Jessen, S., Lu, X., Hsiou, Y., Jacobo-Molina, A., Adries, K., Pauwels, R., Moereels, H., Koymans, L., Janssen, P. A. J., Smirth,

Jr., R. H., Koepke, M. K., Michejda, C. J., Hughes, S. H., Arnold, E. (1995a) Structure of HIV-1 reverse transcriptase in a complex with the non-nucleoside inhibitor alpha-APA R 95845 at 2.8 Å resolution. *Structure*, **3**, 365-379.

Ding, J., Das, K., Moereels, H., Koymans, L., Andries, K., Janssen, P. A., Hughes, S. H., Arnold, E. (1995b) Structure of HIV-1 RT/TIBO R 86183 complex reveals similarity in the binding of diverse nonnucleoside inhibitors. *Nature Structure Biology*, **2**, 407-415.

Ding, J., Hughes, S.H., Arnold, E. (1997) Protein–nucleic acid interactions and DNA conformation in a complex of HIV-1 RT with a double-stranded DNA template-primer. *Biopolymers*, **44**, 125–138.

Ding, J., Das, K., Hsiou, Y., Sarafianos, S. G., Clark, A. D. Jr., Jacobo-Molina, A., Tantillo, C., Hughes, S. H., Arnold, E. (1998) Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with double-stranded DNA and an antibody Fab fragment at 2.8 Å resolution. *Journal of Molecular Biology*, **284**, 1095–1111.

Doan, L., Handa, B., Roberts, N. A., Klumpp, K. (1999) Metal ion catalysis of RNA cleavage by the influenza virus endonuclease. *Biochemistry*, **38**, 5612–5619.

Dubois, R. M., Slavish, P. J., Baughman, B. M., Yun, M. K., Bao, J., Webby, R. J., Webb, T. R., White, S. W. (2012) Structural and biochemical basis for development of influenza virus inhibitors targeting the PA endonuclease. *PLOS Pathog.* **8**, e1002830.

Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R., Davies, D. R. (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science*, **266**, 1981–1986.

Eijkelenboom, A.P., Puras Lutzke, R.A., Boelens, R., Plasterk, R.H., Kaptein, R., Hard, K. (1995) The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nature Structure Biology*, **2**, 807–810.

Eijkelenboom, A.P., Sprangers, R., Hard, K., Puras Lutzke, R.A., Plasterk, R.H., Boelens, R., Kaptein, R. (1999) Refined solution structure of the C-terminal DNA-binding domain of human immunovirus-1 integrase. *Proteins*, **36**, 556–564.

Emsley, P., Lohkamp, B., Scott, W. G., Cowtan, K. (2010) Features and development of Coot. *Acta Crystallography D. Biological Crystallography*, **66**, 486-501.

Engelman, A., Craigie, R. (1992) Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. *Journal of Virology*, **66**, 6361–6369.

Engelman, A., Bushman, F. D., Craigie, R. (1993) Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *EMBO Journal*, **12**, 3269-3275.

Engelman, A., Craigie, R. (1995) Efficient magnesium-dependent human immunodeficiency virus type 1 integrase activity. *Journal of Virology*, **69**, 5908-5911.

Engelman, A., Cherepanov, P. (2008) The lentiviral integrase binding protein LEDGF/p75 and HIV-1 replication. *PLOS Pathogens*, **4**, e1000046.

English, A. C., Done, S. H., Caves, L. S., Groom, C. R., Hubbard, R. E. (1999) Locating interaction sites on proteins: the crystal structure of thermolysin soaked in 2% to 100% isopropanol. *Proteins: Structure, Function, and Genetics*, **37**, 628-640.

Enousf, R., Ren, J., Ross, C., Jones, Y., Stammers, D., Stuart, D. (1995) Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. *Nature Structural Biology*, **2**, 303-308.

Erlanson D. A., Wells J. A., Braisted A. C. (2004) Tethering: fragment-based drug discovery. *Annual Review of Biophysics Biomolecular Structure*, **33**, 199–223.

FDA Influenza antiviral drugs and related information, 2013

Feinberg, M. B., Baltimore, D., Frankel, A. D. (1991) The role of Tat in the human immunodeficiency virus life cycle indicates a primary effect on transcriptional elongation. *Proceeding in National Academy of Sciences USA*, **88**, 4045-4049.

Fenwick, C. W., Tremblay, S., Wardrop, E., Bethell, R., Coulomb, R., Elston, R., Faucher, A. M., Mason, S., Simoneau, B., Tsantrizos, Y. (2011) Resistance studies with HIV-1 non-catalytic site integrase inhibitors. *Antiviral Therapy*, **16**, A9.

Flocco, M. M., Mowbray, S. L. (1994) Planar stacking interactions of arginine and aromatic-side chains in proteins. *Journal of Molecular Biology*, **235**, 709-717.

Fodor, E. (2013) The RNA polymerase of influenza A virus: mechanisms of viral transcription and replication. *Acta Virologica*, **57**, 113-122.

Food and Drug Assosication (FDA) (2013) Antiretroviral drugs used in the treatment of HIV infection. http://www.fda.gov/ForConsumers/byAudience/ForPatientAdvocates/HIVandAIDSActivities/ucm118915.htm.

Fouchier, R. A., Munster, V., Wallensten, A., Bestebroer, T. M., Herfst, S., Smith, D., Rimmelzwaan, G. F., Olsen, B., Osterhaus, A. D. (2005) Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls. *Journal of Virology*, **79**, 2814-2822.

Freed, E. O, Martin, M. A. (1995) The role of human immunodeficiency virus type 1 envelope glycoproteins in virus infection. *Journal of Biological Chemistry*, **270**, 23883–23886.

Ganser-Pornillos, B.K., Yeager, M., and Sundquist, W.I. (2008) The structural biology of HIV assembly. *Current Opinions in Structural Biology*, **18**, 203–217.

Goldgur, Y., Dyda, F., Hickman, A. B.,Jenkins, T. M.,Craigie, R., Davies, D. R. (1998). Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proceedings in National Academy of Sciences USA*, **95**, 9150-9154.

Gomez, C., Hope, T. J. (2005) The ins and outs of HIV replication. *Cell Microbiology*, **7**, 621-626.

Gomez, L., Lorenzo, M. M., Fenton, M. J. (2013) Immunobiology of influenza vaccines. *Recent Advances in Chest Medicine*, **143**, 502-510.

Goody, R. S., Muller, B., Restle, T. (1991) Factors contributing to the inhibition of HIV reverse transcriptase by chain-terminating nucleotides in vitro and in vivo. *FEBS Letters*, **291**, 1–5.

Guo, Y. J., Jin, F. G., Wang, P., Wang, M., Zhu, J. M. (1983) Isolation of influenza C virus from pigs and experimental infections of pigs with influenza C virus. *Journal of General Virology*, **64**, 177-182.

Hadjuk P.J., Greer J. (2007) A decade of fragment-based drug design: strategic advances and lessons learned. *Nature Reviews Drug Discovery*, **6**, 211–219

Halford, B. (2012) Bosutinib buyer beware. *Chemical & Engingeering News*, **90**, 34.

Hamamoto, S., Nishitsuji, H., Amagasa, T., Kannagi, M., Masuda, T. (2006) Identification of a novel human immunodeficiency virus type 1 integrase interactor, Gemin2, that facilitates efficient viral cDNA synthesis in vivo. *Journal of Virology*, 80, 5670-5677.

Hann, M.M., Leach, A.R., Harper, G. (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *Journal of Chemical Information and Computer Sciences*, **41**, 856-864.

Hannah, V. V., Atmanene, C. Zeyer, D., van Dorsselaer, A. (2010) Native MS: an ESI, way to support structure- and fragment-based drug discovery. *Future Medicinal Chemistry*, **2**, 35-50.

Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., Hutchison, G. R. (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, **4**, 17.

Hare, S., S. S. Gupta, E. Valkov, A. Engelman, and P. Cherepanov. (2010) Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature*, **464**, 232-236.

Hare, S., Maertens, G.N., and Cherepanov, P. (2012) 3'-Processing and strand transfer catalysed by retroviral integrase in crystallo. *EMBO Journal*, **31**, 3020-3028.

Hastings, J. C., Selnick, H., Wolanski, B., Tomassini, J. E. (1996) Anti-influenza virus activities of 4-substituted 2,4-dioxobutanoic acid inhibitors. *Antimicrob. Agents Chemother.* **40**, 1304–1307.

Havlir, D., Cheeseman, S. H., Mclaughlin, M., Murphy, R., Erice, A., Spector, S. A., Greenough, T. C., Sullivan, J. L., Hall, D., Myers, M., Lamson, M., Richman, D. D. (1995) High-dose nevirapine: safety, pharmacokinetics, and antiviral effect in patients with human immunodeficiency virus infection. *Journal of Infectious Diseases*, **171**, 537-541.

Hayouka, Z., Rosenbluh, J., Levin, A., Loya, S., Lebendiker, M., Veprintsev, D., Kotler, M., Hizi, A., Loyter, A., Friedler, A. (2007) Inhibiting HIV-1 integrase by shifting its oligomerization equilibrium. *Proceedings in National Academy of Sciences*, **104**, 8316-8321.

Hermann, J. C., Chen, Y., Wartchow, C., Menke, J., Gao, L., Gleason, S. K., Haynes, N. E., Scott, N., Petersen, A., Gabriel, S., Vu, B. T., George, K. M., Narayanan, A., Li, S. H., Qian, H., Beatini, N., Niu, L., Gan, Q.F. (2013) Metal Impurities Cause False Positives in High-Throughput Screening Campaigns. *ACS Medicinal Chemistry Letters*, **4**, 197-200.

Hesterkamp T., Whittaker M. (2008) Fragment-based activity space: smaller is better. *Current Opinion in Chemical Biology,* **12**, 260–268

Hinshaw, V. S., Air, G. M., Gibbs, A. J., Graves, L., Prescott, B., Karunakaran, D. (1982) Antigenic and genetic characterization of a novel hemagglutinin subtype of influenza A viruses from gulls. *Journal of Virology*, **42**, 865-872.

Hirsch, V. M., R. A. Olmsted, M. Murphey-Corb, R. H. Purcell, and P. R. Johnson. (1989) An African primate lentivirus (SIVsm) closely related to HIV- 2. *Nature*, **339**, 389-392.

Honda, A., Uéda, K., Nagata, K., Ishihama, A. (1987) Identification of the RNA polymerase-binding site on genome RNA of influenza virus. *J. Biochem.* **102**, 1241–1249.

Hopkins, A.L., Groom, C.R., Alex, A. (2004) Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today*, **9**, 430-431.

Horton, N. C., Newberry, K. J., Perona, J. J. (1998) Metal ion-mediated substrate assisted catalysis in type II restriction endonucleases. *Proceedings in National Academy of Science U.S.A.*, **95**, 13489–13494.

Horton, N. C., Perona, J. J. (2004) DNA cleavage by EcoRV endonuclease: two metal ions in three metal ion binding sites. *Biochemistry*, **43**, 6841–6857.

Howard, N., Abell, C., Blakemore, W., Chessari, G., Congreve, M., Howard, S., Jhoti, H., Murray, C. W., Seavers, L. C. A., van Montfort, R. L. M. (2006) Application of fragment screening and fragment linking to the discovery of novel thrombin inhibitors. *Journal of Medicinal Chemistry*, **46**, 1346-1355.

Hsiou, Y., Ding, J., Das, K., Clark, A. D. Jr., Hughes, S. H., Arnold, E. (1996) Structure of unliganded HIV-1 reverse transcriptase at 2.7 A resolution: implications of conformational changes for polymerization and inhibition mechanisms. *Structure*, **4**, 853-860.

Hu, W., Hughes, S. H. (2012) HIV-1 reverse transcription. *Cold Spring Harbor Perspectives in Medicine*, **2**, a006882,

Huang, T. S., Palese, P., Krystal, M. (1990) Determination of influenza virus proteins required for genome replication. *J.Virol.* **64**, 5669–5673.

Huang, H., Chopra, R., Verdine, G. L., Harrison, S. C. (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science*, **282**, 1669-1675.

Hunter, E., Swanstrom R. (1990) Retrovirus envelope glycoproteins. *Current Topic in Microbiology and Immunology*, **157**, 187–253.

Ivanov, I., Tainer, J. A., McCammon, J. A. (2007) Unraveling the three-metal-ion catalytic mechanism of the DNA repair enzyme endonuclease IV. *Proceedings in National Academy of Science U.S.A.*, **104**, 1465–1470.

Iwai, Y., Takahashi, H., Hatakeyama, D., Motoshima, K., Ishikawa, M., Sugita, K., Hashimoto, Y., Harada, Y., Itamura, S., Odagiri, T., Tashiro, M., Sei, Y. Yamaguchi, K., Kuzuhara, T. (2010) Anti-influenza activity of phenethylphenylphthalimide analogs derived from thalidomide. *Bioorganic Medicinal Chemistry*, 18, 5379-53790.

Iwai, Y., Murakami, K., Gomi, Y., Hashimoto, T., Asakawa, Y., Okuno, Y., Ishikawa, T., Hatakeyma, D., Echigo, N., Kuzuhara, T. (2011) Anti-influenza activity of marchantins, macrocyclic bisbibenzyls contained in liverworts. *PLOS One*, **6**, e19825.

Izumoto Y., Kuroda T., Harada, H., Kishimoto, T., Nakamura, H. (1997). Hepatoma- derived growth factor belongs to a gene family in mice showing significant homology in the amino terminus. *Biochemical Biophysical Research Communication*, **238**, 26-32.

Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D. Jr., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., Hizi, A., Hughes, S., Arnold, E. (1993) Crystal structure of HIV-1 RT complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proceeding in National Academy of Sciences USA*, **90**, 6320–6324.

Janssen, P. A. J., Lewi, P. J., Arnold, E., Daeyaert, F., de Jonge, M., Heeres, J., Koymans, L., Vinkers, M., Guillemont, J., Pasquier, E., Kukla, M., Ludovici, D., Andries, K., de Bethune, M. P., Pauwels, R., Das, K., Clark, A. D. Jr., Frenkel, Y. V., Hughes, S. H., Medaer, B., De Knaep, F., Bohets, H., De Clerck, F., Lampo, A., Williams, P., Stoffels, P. (2005) In search of a novel anti-HIV drug: multidisciplinary coordination in the discovery of 4-[[4-[[4-[(1E)-2-cyanoethenyl]-2,6-dimethylphenyl]amino]-2-pyrimidinyl]amino]benzonitrile (R278474, rilpivirine). *Journal of Medicinal Chemistry*, **48**, 1901-1909.

Jencks, W.P. (1981) On the attribution and additivity of binding-energies. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, **78**, 4046-4050.

Jochmans, D., Vingerhoets, J., Arnoult, E., Geeraert, L., Guillemont, J. Human immunodeficiency virus type 1 non-nucleoside reverse transcriptase inhibitors. R. L. LaFemina (Ed.), Antiviral Research: Strategies in Antiviral Drug Disocvery, ASM Press, Washinton (2009), 33-50.

Jordan, J. B., Poppe, L., Xia, X., Cheng, A. C., Sun, Y., Michelsen, K., Eastwood, H., Schnier, P. D., Nixey, T., Zhong, W. (2012) Fragment based drug discovery: practical implementation baed on 19F NMR spectroscopy. *Journal of Medicinal Chemistry*, **55**, 678-687.

Kao, S.Y., Calman, A. F., Luciw, P. A., Peterlin, B. M. (1987) Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature*, **330**, 489-493.

Kawaoka, Y., Yamnikova, S., Chambers, T. M., Lvov, D. K., Webster, R. G. (1990) Molecular characterization of a new hemagglutinin, subtype H14, of influenza A virus. *Virology*, **179**, 759-767.

Kessl, J. J., McKee, C. J., Eidahl, J. O., Shriabai, N., Katz, A., Kvaratskhelia, M. (2009) HIV-1 integrase-DNA recognition mechanism. *Viruses*, **1**, 713-736.

Kessl, J., Jena, N., Koh, Y., Taskent-Sezgin, H., Slaughter, A., Feng, L., de Silva, S., Wu, L., Le Grice, S. F. J., Engelman, A., Fuchs, J. R., Kvaratskhelia, M. (2012) Multimode, Cooperative Mechanism of Action of Allosteric HIV-1 Integrase Inhibitors. *Journal of Biological Chemistry*, **287**, 16801-16811.

Klages, J., Coles, M., Kessler, H. (2006) NMR-based screening: a powerful tool in fragment–based drug discovery, *Molecular Biosystems*, **2**, 318-332.

Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A., and Steitz, T.A. (1992). Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science*, **256**, 1783-1790.

Koster, H., Crann, T., Brass, S., Herhaus, C., Zentgraf, M., Neumann, L., Heine, A., Klebe, G. (2011) A small non-rule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes. *Journal of Medicinal Chemistry,* **54**, 7784–7796.

Kovall, R. A., Matthews, B. W. Type II restriction endonucleases: structural, functional, and evolutionary relationships. *Current Opinions in Chemical Biology*, **3**, 578–583.

Kowalinski, E., Zubieta, C., Wolkerstorfer, A., Szolar, O. H. J., Ruigrok, R. W. H., Cusack, S. (2012) Structural analysis of specific metal chelating inhibitor binding to the endonuclease domain of influenza pH1N1 (2009) polymerase. *PLOS Pathogens* **8**, e10002831.

Kozikowski, B. A., Burt, T. M., Tirey, D. A., Williams, L. E., Kuzmak, B. R., Stanton, D. T., Morand, K. L., Nelson, S. L. (2003) The Effect of Room-Temperature Storage on the Stability of Compounds in DMSO. *Journal of Biomolecular Screening*, **8**, 205-209.

Krug, R. M., Alonso-Caplen, F. V., Julkunen, I., Katze, M. G. (1989) Expression and replication of the influenza virus genome. In *The Influenza Viruses* (Krug, R. M. Ed.), Plenum Publishing Corp., New York.

Krug, R. M., Aramini, J. M. (2009) Emerging antiviral targets for influenza A virus. *Trends in Pharmacological Sciences*, **30**, 269-277.

Kulkosky, J., Jones, K.S., Katz, R.A., Mack, J.P., Skalka, A.M. (1992) Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transpo- sases. *Molecular Cell Biology*, **12**, 2331–2338.

Kuntz, I.D., Chen, K., Sharp, K.A., Kollman, P.A. (1999) The maximal affinity of ligands. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 9997-10002.

Kuroda, D. G., Bauman, J. D., Challa, R., Patel, D., Troxler, T., Das, K., Arnold, E., Hochstrasser, R. M. (2013) Snapshot of the equilibrium dynamics of a drug bound to HIV-1 reverse transcriptase. *Nature*, **5**, 174-181.

Kuzuhara, T., Iwai, Y., Takahashi, H., Hatakeyama, D., Echigo, N. (2009) Green tea catechins inhibit the endonuclease activity of influenza A virus RNA polymerase. *PLOS Currents*, **1**, RRN1052.

Kwong, P.D., Wyatt, R., Sattentau, Q. J., Sodroski, J., Hendrickson, W. A. (2000) Oligomeric modeling and electrostatic analysis of the gp120 envelope glycoprotein of human immunodeficiency virus. *Journal of Virology*, **74**, 1961-1972.

Ladbury, J.E. (2010) Calorimetry as a tool for understanding biomolecular interactions and an aid to drug design. *Biochemical Society Transactions*, **38**, 888-893.

Leach, A. R., Hann, M. M., Burrows, J. N., Griffen, E. J. (2006) Fragment screening: an introduction. *Molecular Biosystems*, **2**, 430-446.

Leach, A. R., Hann, M. M. (2011) Molecular complexity and fragment-based drug discovery: ten years on. *Current Opinion in Chemical Biology*, **15**, 489-496.

Lee, S.P., Han, M.K. (1996) Zinc stimulates Mg2+-dependent 3′-processing activity of human immunodeficiency virus type 1 integrase in vitro. *Biochemistry*, **35**, 3837–3844.

Lee, S. P., Xiao, J., Knutson, J. R., Lewis, M. S., Han, M. K. (1997) Zn+2 promotes the self-association of human immunodeficiency virus type-1 integrase in vitro. *Biochemistry*, **36**, 173-180.

Li, M. Z., Elledge, S. J. (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nature Methods*, **4**, 251-256.

Li, X., Krishnan, L., Cherepanov, P., Engelman, A. (2011) Structural biology of retroviral DNA integration. *Virology*, **411**, 194-205.

Lin, C. W., Engelman, A. (2003) The barrier-to-autointegration factor is a component of functional human immunodeficiency virus type 1 preintegration complexes. *Journal of Virology*, **77**, 5030-5036.

Lin, T. Y., Timasheff, S. N. (1994) Why do some organisms use a urea-methylamine mixture as osmolyte? Thermodynamic compensation of urea and trimethylamine N-oxide interactions with protein. *Biochemistry*, **33**, 12695-12701.

Liu, S., Abbondanzieri, E. A., Rausch, J. W., Le Grice, S. F. J., Zhuang, X. (2008) Slide into action: dynamic shuttling of HIV reverse transcriptase on nucleic acid substrates. *Science*, **322**, 1092-1097.

Llano, M., Vanegas, M., Fregoso, O., Saenz, D. T., Chung, S., Peretz, M., Poeschla, E. M. (2004) LEDGF/p75 deter- mines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral pre-integration complexes. *Journal of Virology*, **78**, 9524 – 9537.

Llano, M., Vanegas, M., Hutchins, N., Thompson, D., Delgado, S., Poeschla, E. M. (2006) Identification and characterization of the chromatin-binding domains of the HIV-1 integrase interactor LEDGF/p75. *Journal of Molecular Biology*, **360**, 760-773.

Luria, S., Chambers, I., Berg, P. (1991) Expression of the type 1 human immunodeficiency virus Nef protein in T cells prevents antigen receptor-mediated induction of interleukin 2 mRNA. *Proceedings in National Academy of Sciences USA*, **88**, 5326-5330.

Maertens, G., Cherepanov, P., Pluymers, W., Busschots, K., De Clercq, E., Debyser, Z., Engelborghs, Y. (2003) LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *Journal of Biological Chemistry*, **278**, 33528 – 33539.

Malim, M. H., Emerman, M. (2008) HIV-1 accessory proteins---ensuring viral survival in a hostile environment. *Cell Host and Microbe*, **3**, 388-398.

Manuguerra, J. C., Hannoun, C. (1992) Natural infection of dogs by influenza C virus. *Research in Virology*, **143**, 199-204.

Maple, H. J., garlish, R. A., Rigauu-Roca, L., Porter, J., Whitcombe, I., Prosser, C. E., Kennedy, J., Henry, A. J., Taylor, R. J., Crump, M. P., Crosby, J. (2012) Automated protein-ligand interaction screening by mass spectrometry. *Journal of Medicinal Chemistry*, **55**, 837-851.

Martinez-Sobrido, L., Cadagan, R., Steel, J., Basler, C. F., Palese, P., Moran, T. M., Garcia-Sastre, A. (2010) Hemagglutinin-pseudotyped green fluorescent protein-expressing influenza viruses for the detection of influenza virus neutralizing antibodies. *Journal of Virology*, **84**, 2157−2163.

Mattos, C., Ringe, D. (1996) Locating and characterizing binding sites on proteins. *Nature Biotechnology*, **14**, 595-599.

Mattos, C., Bellamacina, C. R., Peisach, E., Pereira, A., Vitkup, D., Petsko, G. A., Ringe, D. (2006) Multiple solvent crystal structures: probing binding sites, plasticity, and hydration. *Journal of Molecular Biology*, **57**, 1471-1482.

Mayer, M., Meyer, B. (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angewandte Chemie International Edition*, **38**, 1784-1788.

McDonald, E. Biosynthesis of isoquniolines. G. Grethe (Ed.), The chemistry of hetrocyclic compounds. (1981) John Wiley & Sons, Inc. West Sussex, 275-379.

McSharry, J. J., McDonough, A. C., Olson, B. A., Drusano, G. L. (2004) Phenotypic drug susceptibility assay for influenza virus neuraminidase inhibitors. *Clinical Diagnostic Laboratory Immunology*, **11**, 21–28.

Miller, M. D., Warmerdam, M. T., Gaston, I., Greene, W.C., Feinberg, M.B. (1994) The human immunodeficiency virus-1 nef gene product: a positive factor for viral infection and replication in primary lymphocytes and macrophages. *Journal of Experimental Medicine*, **179**, 101-113.

Misumi, S., Inoue, M., Dochi, T., Kishimoto, N., Hasegawa, N., Takamune, N. Shoji, S. (2010) Uncoating of human immunodeficiency virus type 1 requires prolyl isomerase Pin1. *Journal of Biological Chemistry*, **285**, 25185-25195.

Monto, A. S. (2005) The threat of an avian influenza pandemic. *New England Journal of Medicine*, **352**, 323-325.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., Olson, A. J. (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. *Journal of Computational Chemistry*, **16**, 2785-91.

Muraki, Y., Hongo, S. (2010) The molecular virology and reverse genetics of influenza C virus. *Japanese Journal of Infectious Disease*, **63**, 157-165.

Nakamura, C. E., Abeles, R. H. (1985) Mode of interaction of beta-hydroxy-beta-methylglutaryl coenzyme-a reductase with strong binding inhibitors - compactin and related compounds. *Biochemistry*, **24**, 1364-1376.

Nakazawa, M., Kadowaki, S. E., Watanabe, I., Kadowaki, Y., Takei, M., Fukuda, H. (2008) PA subunit of RNA polymerase as a promising target for anti-influenza virus agents. *Antiviral Research*, **78**, 194-201.

Nanni, R. G., Ding, J., Jacobo-Molina, A. J., Hughes, S. H., Arnold, E. (1993). Review of HIV-1 reverse transcriptase three dimensional structure: Impli- cations for drug design. *Perspectives in Drug Discovery and Design,* **1**, 129–150.

Navratilova, I., Hopkins, A.L. (2010) Fragment screening by surface plasmon resonance. *ACS Medicinal Chemistry Letters*, **1**, 44-48.

Nekhai, S., Jeang, K. T. (2006) Transcriptional and post-transcriptional regulation of HIV-1 gene expression: role cellular co-factors for Tat and Rev. *Future Microbiology*, **1**, 417-426.

Neumann, T., Junker, H.D., Schmidt, K., Sekul, R. (2007) SPR-based fragment screening: Advantages and applications. *Current Topics in Medicinal Chemistry*, **7**, 1630-1642.

Oldenburg, K., Pooler, D., Scudder, K., Lipinski, C., Kelly, M. (2005) High throughput sonication: evaluation for compound solubilization. *Combinatorial Chemistry & High Throughput Screening*, **8**, 499-512.

Oprea, T.I., Davis, A.M., Teague, S.J., Leeson, P.D. (2001) Is there a difference between leads and drugs? A historical perspective. *Journal of Chemical Information and Computer Sciences*, **41**, 1308-1315.

Osterhaus, A. D. M. E., Rimmelzwaan, G. F., Martina, B. E. E., Bestebroer, T. M., Fouchier, R. A. M. (2000) Influenza B virus in seals. *Science*, **288**, 1051–1053.

Otwinowski Z., Minor W, Borek, D, Cymborowski, M. DENZO and SCALEPACK. E. Arnold, D. M. Himmel, M. Rossmann (Eds.), International Tables for Crystallography Volume F. Crystallography of Biological Macromolecules, Second Edition. (2011) John Wiley & Sons, Inc., West Sussex, 282-295.

Palese, P., Shaw, M. L. (2007). Orthomyxoviridae: the viruses and their replication. In Fields Virology, 5th edn, pp. 1647–1689. Edited by D. M. Knipe & P. M. Howley. Philadelphia, PA: Lippincott Williams & Wilkins.

Parkes, K. E. B., Ermert, P., Fassler, J., Ives, J., Martin, J. A., Merrett, J. H., Obrecht, D., Williams, G., Klumpp, K., (2003) Use of a pharmacophore model to discover a new class of influenza endonuclease inhibitors. *Journal of Medicinal Chemistry*, **46**, 1153−1164.

Parhi, A., K., Xiang, A., Bauman, J. D., Patel, D., Vijayan, R. S. K., Das, K., Arnold, E., LaVoie, E. J. (2013) Phenyl substituted 3-hydroxypyridin(1H)-2-ones: inhibitors of influenza A endonuclease. *Bioorganic & Medicinal Chemistry*, in press.

Peat, T. S., Rhodes, D. I., Vandegraaff, N., Le, G., Smith, J. A., Clark, L. J., Jones, E. D., Coates, J. A. V., Thienthong, N., Newman, J., Dolezal, O., Mulder, R., Ryan, J. H., Savage, G. P., Francis, C. L., Deadmna, J. J. (2012) Small Molecule Inhibitors of the LEDGF Site of Human Immunodeficiency Virus Integrase Identified by Fragment Screening and Structure Based Design. *PLOS One*, **7**, e40147.

Perryman, A.L., Zhang Q., Soutter, H. H., Rosenfeld, R., McRee, D. E., Olson, A. J., Elder, J. E., Stout, C. D. (2010) Fragment-based screen against HIV protease. *Chemical Biology Drug Design*, **75**, 257–268.

Perspicace, S., Banner, D., Benz, J., Müller, F., Schlatter, D., Huber, W. (2009) Fragment-based screening using surface plasmon resonance technology. *Journal of Biomolecular Screening*, **14**, 337-349.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., Ferrin, T. E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, **25**, 1605–1612.

Plantier, J. C., Leoz, M., Dickerson, J. E., De, O. F., Cordonnier F., Lemee V., Damond F., Robertson D. L., Simon, F. (2009) A new human immunodeficiency virus derived from gorillas. *Nature Medicine*, **15**, 871-872.

Plotch, S. J., Bouloy, M., Ulmanen, I., Krug, R. M. (1981) A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* **23**, 847–858.

Poeschla, E.M. (2008) Integrase, LEDGF/p75 and HIV replication. *Cellular and Molecular Life Science*, **65**, 1403–1424.

Quashie, P. K., Mesplede, T., Wainberg, M. A. (2013) HIV drug resistance and the advent of integrase inhibitors. *Current Infectious Disease Report*, **15**, 85-100.

Raghavendra, N. K., Shkriabi, N., Graham, R. L. J., Hess, S., Kvaratskhelia, M., Wu, L. (2010) Identification of host proteins associated with HIV-1 preintegration complexes isolated from infected $CD4^+$ cells. *Retrovirology*, **7**, 1742-4690-7-66.

Rees, D. C., Congreve, M., Murray, C. W., Carr, R. (2004) Fragment-based lead discovery. *Nature Reviews Drug Discovery*, **3**, 660-672.

Ren, J., Esnouf, R., Hopkins, A., Ross, C., Jones, Y., Stammers, D., Stuart, D. (1995) The structure of HIV-1 reverse transcriptase complexed with 9-chloro-TIBO: lessons for inhibitor design. *Structure*, **3**, 915-926.

Rhodes, D. I., Peat, T. S., Vandegraaff, N., Jeevarajah, D., Le, G., Jones, E. D., Smith, J. A., Coates, J. A. V., Winfield, L. J., Theinthong, N., Newman, J., Lucent, D., Ryan, J. H., Savage, G. P., Francis, C. L., Deadman, J. J. (2011) Structural basis for a new mechanism

of inhibition of HIV-1 integrase identified by fragment screening and structure-based design. *Antiviral Chemistry & Chemotherapy*, **21**, 155-168.

Rishton, G.M. (1997) Reactive compounds and in vitro false positives in HTS. *Drug Discovery Today*, **2**, 382–384

Rishton, G.M. (2003) Nonleadlikeness and leadlikness in biochemical screening. *Drug Discovery Today,* **8**, 86-96.

Rohm, C., Zhou, N., Suss, J., Mackenzie, J., Webster, R. G. (1996) Characterization of a novel influenza hemagglutinin, H15: criteria for determination of influenza A subtypes. *Virology*, **217**, 508-516.

Sahlberg, C., Zhou, X. (2008) Development of non-nucleoside reverse transcriptase inhibitors for anti-HIV therapy. *Anti-Infective Agents in Medicinal Chemistry*, **7**, 101-117.

Sam, M. D., Perona, J. J. (1999) Catalytic roles of divalent metal ions in phosphoryl transfer by EcoRV endonuclease. *Biochemistry* **38,** 6576–6586.

Sagong, H. Y., Parhi, A., Bauman, J. D., Patel, D., Vijayan, R. S. K., Das, K., Arnold, E., LaVoie, E. J. (2013) 3‐Hydroxyquinolin-2(1H)‐ones As Inhibitors of Influenza A Endonuclease. *ACS Medicinal Chemistry Letters*, **4**, 547-550.

Sarafianos, S. G., Das, K., Tantillo, C., Clark, A. D. Jr., Ding, J., Whitcomb, J. M., Boyer, P. L., Hughes, S. H., Arnold, E. (2001) Crystal structure o HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *The EMBO Journal*, **20**, 1449-1461.

Schickli, J. H., Flandorfer, A., Nakaya, T., Martinez-Sobrido, L., Garcia-Sastre, A., Palese, P. (2001) Plasmid-only rescue of influenza A virus vaccine candidates. *Philosophical Transactions of the Royal Society Biological Sciences*, **356**, 1965−1973.

Scott, A. D., Phillips, C., Alex, A., Flocco, M., Bent, A., Randall, A., O'Brien, R., Damian, L., Jones, L. H. (2009) Thermodynamic optimization in drug discovery: a case study using carbonic anhydrase inhibitors. *ChemMedChem*, **4**, 1985-1989.

Scott, D. E., Coyne, A. G., Hudson, S. A., Abell, C. (2012) Fragment based approaches in drug discovery and chemical biology. *Biochemistry*, **51**, 4990-5003.

Scott, L.J., Perry, C.M. (2000) Delavirdine: a review of its use in HIV infection. *Drugs*, **60**, 1411–1444.

Seelamgari, A., Maddukuri, A., Berro, R., de La Fuente, C., Kehn, K., Deng, L., Dadgar, S., Bottazzi, M. E., Ghedin, E., Pumfery, A., Kashanchi, F. (2004) Role of viral regulatory and accessory proteins in HIV-1 replication. *Frontier Bioscience*, **9**, 2388-2413.

Sharp, P.M., Bailes, E., Robertson, D. L., Gao, F., Hahn, B. H. (1999) Origins and evolution of AIDS viruses. *Biological Bulletin*, **196**, 338-42.

Shuker, S.B., Hajduk, P.J., Meadows, R.P., Fesik, S.W. (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science*, **274**, 1531-1534.

Shun, M. C., Raghavendra, N. K., Vandegraaff, N., Daigle, J. E., Hughes, S., Kellam, P., Cherepanov, P., Engelman, A. (2007) LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes & Development*, **21**,1767-1778.

Siegal, G., Ab, E., Schultz, J. (2007) Integration of fragment screening and library design. *Drug Discovery Today*, **12**, 1032-1039.

Singh, D.P., Kimura, A., Chylack Jr., L.T., Shinohara, T., 2000a. Lens epithelium- derived growth factor (LEDGF/p75) and p52 are derived from a single gene by alternative splicing. Gene 242 (1-2), 265–273.

Sluis-Cremer, N., Temiz, N. A., Bahar, I. (2004) Conformational changes in HIV-1 reverse transcriptase induced by nonnucleoside reverse transcriptase inhibitor binding. *Current HIV Research*, **2**, 323-332.

Scott, L. J., Perry, C. M. (2000) Delavirdine: a review of its use in HIV infection. *Drugs*, **60**, 1411-1444.

Song, J. M., Lee, K. H., Seong, B. L. (2005) Antiviral effect of catechins in green tea on influenza virus. *Antiviral Research*, **68**, 66–74.

Stec, I., Nagl, S. B., van Ommen, G. J., den Dunnen, J. T. (2000) The PWWP domain: a potential protein-protein interaction domain in nuclear proteins influencing differentiation? *FEBS Letters*, **473**, 1-5.

Stockman, B. J., Dalvit, C. (2002) NMR screening techniques in drug discovery and drug design. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **41**, 187-231.

Stopak, K., de Noronha, C., Yonemoto, W., Greene, W. C. (2003) HIV-1 vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. *Molecular Cell*, **12**, 591-601.

Swayze, E. E., Jefferson, E. A., Sannes-Lowery, K. A., Blyn, L. B., Risen, L.M., Arakawa, S., Osgood, S. A., Hofstadler, S. A., Griffey, R. H. (2002) SAR by MS: a ligand based technique for drug lead discovery against structured RNA targets. *Journal of Medicinal Chemistry*, **45**, 3816-3819.

Syson, K., Tomlinson, C., Chapados, B. R., Sayers, J. R., Tainer, J. A., Williams, N. H., Grasby, J. A. (2008) Three metal ions participate in the reaction catalyzed by T5 flap endonuclease. *Journal of Biological Chemistry*, **283**, 28741–28746.

Tantillo, C., Ding, J., Jacobo-Molina, A., Nanni, R. G., Boyer, P. L., Hughes, S. H., Pauwels, R., Andries, K., Janssen, P. A., Arnold, E. (1994) Locations of anti-AIDS drug binding sites and resistance mutations in the three-dimensional structure of HIV-1 reverse transcriptase. Implications for mechanisms of drug inhibition and resistance. *Journal of Molecular Biology, 243, 369-387.*

Teague, S.J., Davis, A.M., Leeson, P.D., Oprea, T. (1999) The design of leadlike combinatorial libraries. *Angewandte Chemie International Edition*, **38**, 3743-3748.

Tebit, D. M., Arts, Eric. J. (2011) Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *Lancet Infectious Disease*, **11**, 45-56.

Tomassini, J., Selnick, H., Davies, M. E., Armstrong, M. E., Baldwin, J., Bourgeois, M., Hastings, J., Hazuda, D., Lewis, J., McClements, W., Ponticello, G., Radzilowski, E., Smith, G., Tebben, A., and Wolfe, A. (1994) Inhibition of Cap (m7GpppXm)-dependent endonucelase of influenza-virus by 4-subsituted 2,4-dioxobutanoic acid compounds. *Antimicrobial Agents Chemotherapy*, **38**, 2827−2837.

Tomassini JE, Davies ME, Hastings JC, Lingham R, Mojena M, Raghoobar SL, Singh SB, Tkacz JS, Goetz MA (1996) A novel antiviral agent which inhibits the endonuclease of influenza viruses. *Antimicrobial agents and chemotherapy* 40(5): 1189-1193.

Torres, F.E., Recht, M.I., Coyle, J.E., Bruce, R.H., Williams, G. (2010) Higher throughput calorimetry: opportunities, approaches and challenges. *Current Opinion in Structural Biology*, **20**, 598-605.

Trono, D. (1995) HIV accessory proteins: leading roles for the supporting cast. *Cell*, **82**, 189-192.

Trott, O., Olson, A. J. (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, **31**, 455–461.

Van Gent, D. C., Vink, C., Groeneger, A. A., Plasterk, R. H. (1993) Complementation between HIV integrase proteins mutated in different domains. *EMBO Journal*, **12**, 3261-3267.

Vandekerckhove L, Christ F, Van Maele B, De Rijck J, Gijsbers R, et al. (2006) Transient and stable knockdown of the integrase cofactor LEDGF/p75 reveals its role in the replication cycle of human immunodeficiency virus. *Journal of Virology*, **80**, 1886–1896.

Vanwetswinkel, S., Heetebrij, R.J., van Duynhoven, J., Hollander, J.G., Filippov, D.V., Hajduk, P.J., Siegal, G. (2005) TINS, target immobilized NMR screening: An efficient and sensitive method for ligand discovery. *Chemistry and Biology*, **12**, 207-216.

Verlinde, C. L. M. J., Fan, E., Shibata, S., Zhang, Z., Sun, Z., Deng, W., Ross, J., Kim, J., Xiao, L., Arakaki, T., Bosch, J., Caruthers, J. M., Larson, E.T., LeTrong, I., Napuli A., Kelley, A., Mueller, N., Zucker, F., Van Voorhis, W. C., Buckner, F. S., Merritt, E. A., Hol, W. G. J. (2009) Fragment-based cocktail crystallography by the medical structural genomics of pathogenic protozoa consortium. *Current Topics in Medicinal Chemistry*, **9**, 1678-1687.

Vivat, H. V., Atmanene, C., Zeyer, D., Van Dorsselaer, A., Sanglier-Cianferani, S. (2010) Native MS: and ESI way to support structure and fragment based drug discovery. *Future Medicinal Chemistry*, **2**, 35-50.

Vrouenraets, S. M. E., Wit, F. W. N. M., van Tongeren, J., Lange, J. M. A. (2007) Efavirenz: a review. *Expert Opinion on Pharmacotherapy*, **8**, 851-871.

Vulpetti, A., Hommel, U., Landrum, G., Lewis, R., Dalvit, C. (2009) Design and NMR-based screening of LEF, a library of chemical fragments with different local environment of fluorine. *Journal of American Chemistry Society*, **131**, 12949-12959.

Vulpetti, A., Dalvit, C. (2012) Fluorine local environment: from screening to drug design. Technical and practical aspects of 19F NMR-based screening: toward sensitive high-throughput screening with rapid deconvolution. *Drug Discovery Today*, **17**, 890-897 592-597.

Wang, A., Bolen, D. W. (1997) A naturally occurring protective system in urea-rich cells: Mechanism of osmolyte protection of proteins against urea denaturation. *Biochemistry*, **36**, 9101-9108.

Wang, J.Y. Ling, H., Yang, W., Craigie, R. (2001) Structure of a two-domain fragment of HIV-1 integrase: Implications for domain organization in the intact protein. *EMBO Journal*, **20**, 7333–7343.

Ward, R. A., Brassington, C., Breeze, A. L., Caputo, A., Critchlow, S., Davies, G., Goodwin, L., Hassall, G., Greenwood, R., Holdgate, G. A., Mrosek, M., Norman, R. A., Pearson, S., Tart, J., Tucker, J. A., Vogtherr, M., Whittaker, D., Wingfield,J., Winter, J., Hudson, K. (2012) Design and synthesis of novel lactate dehydrogenase A inhibitors by fragment-based lead generation. *Journal of Medicinal Chemistry*, **55**, 3285-3306.

Wielens, J., Headey, S. J., Deadman, J. J., Rhodes, D. I., Parker, M. W., Chalmers, D. K, Scanion, M. J. (2011) Fragment-based design of ligands targeting a novel site on the integrase enzyme of human immunodeficiency virus 1. *ChemMedChem*, **6**, 258–261.

WHO HIV/AIDS, World Health Organization [online]. Available: http://www.who.int/hiv/en/ (accessed July 2013).

Williams, G. (2011) The Pyramid^{TM} approach to fragment based biophysical screening, in Label-Free Technologies for Drug Discovery (eds M. Cooper and L. M. Mayr), John Wiley & Sons, Ltd., Chichester, UK.

Yamada, K., Koyama, H., Hagiwara, K., Ueda, A., Sasaki, Y., Kanesashi, S., Ueno, R., Nakamura, H. K., Kuwata, K., Shimizu, K., Suzuki, M., Aida, Y. (2012) Identification of a novel compound with antiviral activity against influenza A virus depending on PA subunit of viral RNA polymerase. *Microbes and Infection*, **14**, 740-747.

Yancey, P. H., Somero, G. N. (1979) Counteraction of urea destabilization of protein structure by methylamine osmoregulatory compounds of elasmobranch fishes. *Biochemistry Journal*, **183**, 317-323.

Yancey, P. H., Clark, M. E., Hand, S. C., Bowlus, R. D., Somero, G. N. (1982) Living with water stress: evolution of osmolyte systems. *Science*, **217**, 1214-1222.

Yin, Z., Craigie, R. (2010) Modeling the HIV-1 intasome: a prototype view of the target of integrase inhibitors. *Viruses*, **2**, 2777-2781.

Yuan, P., Bartlam, M., Lou, Z., Chen, S., Zhou, J., He, X., Lv, Z., Ge, R., Li, X., Deng, T., Fodor, E., Rao, Z., and Liu, Y. (2009) Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. *Nature*, **458**, 909−913.

Zambon, Maria C. (1999) Epidemiology and pathogenesis of influenza. *Journal of Antimicrobial Chemotherapy*, **44**, 3-9.

Zhang, J., Chung, T., Oldenburg, K. (1999) A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *Journal of Biomolecular Screening*, **4**, 68-73.

Zhao, C., Lou, Z., Guo, Y., Ma, M., Chen, Y., Liang, S., Zhang, L., Chen, S., Li, X., Liu, Y., Bartlam, M., Rao, Z. (2009) Nucleoside monophosphate complex structures of the endonuclease domain from the influenza virus polymerase PA subunit reveal the substrate bindng site inside the catalytic center. *Journal of Virology*, **83**, 9024-9030.

Zheng, R., Jenkins, T.M., Craigie, R. (1996) Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proceedings in National Academy of Sciences USA*, **93**, 13659–13664.

Zou, Q., Bennion, B. J., Daggett, V., Murphy, K. P. (2002) The molecular mechanism of stabilization of proteins by TMAO and its ability to counteract the effects of urea. *Journal of American Chemical Society*, **124**, 1192-1202.