APPROVAL SHEET

Title of Dissertation: Allosteric Inhibition in Viral Polymerases: Understanding the Impact of Multiple Inhibitors and Predicting Novel Binding Sites using Computational Chemistry Approaches

Name of Candidate: Jodian Allison Brown
Doctor of Philosophy, 2016

Dissertation and Abstract Approved: _____
Dr. Ian F. Thorpe
Assistant Professor
Department of Chemistry & Biochemistry

Date Approved: _____

ABSTRACT

Title of Document: **ALLOSTERIC INHIBITION IN VIRAL POLYMERASES: UNDEERSTANDING THE IMPACT OF MULTIPLE INHIBITORS AND PREDICTING NOVEL BINDING SITES USING COMPUTATIONAL CHEMISTRY APPROACHES**.

Jodian Allison Brown, Ph.D., 2016

Directed By: Assistant Professor, Ian F. Thorpe, Department of Chemistry & Biochemistry

Viral pathogens are responsible for a large number of human infections. For example, the Hepatitis C virus (HCV) affects close to 3% of the world's population. There has been significant progress in the therapeutic options available for the treatment of HCV and most of the highly effective existing therapies are a combination of small molecule inhibitors targeting key viral enzymes. One such target is the polymerase, which is a critical component of the viral life cycle. Our work focuses on understanding inhibition of the polymerase by small molecules that bind to pockets outside the active site (i.e. allosteric inhibitors). The first objective of this study is to determine how allosteric inhibitors alter the structural, dynamic and thermodynamic properties of the HCV polymerase so that these small molecules can synergistically inhibit the enzyme. Our second goal is to identify putative allosteric sites in the polymerases of the Dengue (DENV), West Nile (WNV) and Foot-and-mouth Disease (FMDV) viruses. One of the existing challenges in current HCV treatment is that low fidelity of NS5B increases the rate of mutations in the virus. This results in the

generation of multiple enzyme variants and makes it difficult to target the enzyme with single inhibitors. In contrast, for DENV, WNV and FMDV one of the major hurdles is lack of biochemical and structural data that can help steer more structure-based drug discovery efforts. Results from our studies of the HCV polymerase suggest that allosteric inhibitors from nonoverlapping sites can bind simultaneously and synergistically modulate the enzyme conformation and free energy landscape. This work provides the first molecular descriptions of mechanisms underlying enhanced inhibition of the HCV polymerase. We used validated HCV allosteric sites as targets to evaluate several ligand binding site predictor tools and used the best tool to identify novel putative allosteric pockets in the DENV, WNV and FMDV polymerases. The identified sites displayed structural and/or chemical similarities with two of the NS5B validated allosteric pockets. Our approach provides an economic and faster means to facilitate an essential step of drug discovery (i.e. binding site identification), which is critical for infections with limited therapeutic options.

ALLOSTERIC INHIBITION IN VIRAL POLYMERASES: UNDERSTANDING THE IMPACT OF MULTIPLE INHIBITORS AND PREDICTING NOVEL BINDING SITES USING COMPUTATIONAL CHEMISTRY APPROACHES.

By

Jodian Allison Brown.

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, Baltimore County, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Acknowledgements

This dissertation is a possibility because of the support of many individuals who have walked through this journey with me.

First, I want to thank my research advisor and mentor Dr. Ian F. Thorpe. His ability to simultaneously provide the guidance and space for independence has facilitated my growth as a far more confident and independent researcher.

I want to also express my gratitude to members of my dissertation committee: Drs. Alexander Drohat, Elsa Garcin, Alexander MacKerell and Michael Summers for their expertise and support during this process. My interactions with you all have allowed be to become a more conscientious scientist.

I am also grateful for the continuous support from various lab members of the Thorpe laboratory. Specifically, I have truly enjoyed the journey with lab mates Brittny Davis, Ph.D., Ester Sesmero and Nicole Carbonaro who have become like family. You ladies have been on many occasions selfless (e.g. providing food, encouraging me or taking me to prenatal doctor visits) and I am incredibly grateful to have had you during this process! I also want to thank three undergraduates Joel Abraham, Marie Espiritu and Kendra Johnson who have assisted in several aspects of the research projects carried out.

I also received funding from various sources that allowed me to maximize the doctoral process. I have received financial support from an NIH/NIGMS F31 predoctoral fellowship, Meyerhoff Graduate Fellowship and from the Department of Chemistry & Biochemistry.

There are also many programs and personnel who are a part of my community of support and that I am truly grateful for. The Meyerhoff Graduate Fellows (MGF) and PROMISE (Maryland's AGEP) programs have been instrumental in keeping connecting me to other graduate students and resources needed to navigate the graduate process. Mrs. Justine Johnson (from MGF) and Drs. Wendy Carter-Veale and Renetta Tull (from PROMISE) have all served as professional and personal mentors. I am deeply appreciative of Dr. Wendy Carter-Veale's support that has been especially crucial to my success during this dissertation completion process.

In addition, I appreciate all the administrators and staff in the Department of Chemistry & Biochemistry for their encouragement, assistance and guidance. Patricia Gagne, Cynthia Finch, Ann Gefert, Sandy Tabler, Jane Henderson, William Devilbiss, Michele Mullins, Creighton Smith and Dennis Cuddy for all the support that only you could provide.

Also, I am truly indebted to all of my friends for their unconditional support. Kavelle Whyte, Shimeka Bayley, Shannadora Barnes, Takyiah Stevenson, Kendra Jones, Tamara Appalsammy, Olawunmi Isijola, Triscia Thorpe, Ashley Christie, Dane

Christie, Jayea Mote, Cecilia Komey, Taeyjuana Curry and DeLauren McCauley you all rock!!! Thanks for all the early morning, lunchtime or late evening venting sessions, for watching my baby so I could do research and/or write my dissertation, or for talking me off the ledge when I got too close to the edge.

I am beyond grateful for the continued support, patience and love of my family. I thank my aunt Marlene Abad who for 3 consecutive years of my graduate career sent me encouraging text messages every single day…and continues to encourage me! Also grateful for my cousin Annakay McFarlane-Walker and grandmother Edna Stennett who prayed and encouraged me on a continuous basis. To my sister Rosica Brown, I appreciate all the conversations and videos you made that kept me going when times got difficult. I am forever grateful to my mom (Marcia Grant) and my dad (Franklin Brown) who at a very young age taught me the value of education and hard work. Mom, words are insufficient to express all the ways you have and continue to support me. Specifically, I thank you for being my personal alarm clock every morning for the 6-7 months leading up to my candidacy exam and for being one of the leaders of my cheering squad.

Finally, I just thank GOD for orchestrating all the people and events that played a part in the successful writing of this chapter in my life. Also, I am just so grateful to my son Jude who was equipped with the temperament to adjust seamlessly with the various throes of life that we were and are constantly bombarded with. Jude, my sincerest hope is that this dissertation will serve as a constant reminder that nothing can prevent you from your divine destiny!

Table of Contents

List of Tables

List of Figures

Chapter 1: Background & Significance – Introduction to Protein Allostery, Allosteric Inhibition & Viral Polymerases

This Chapter is reproduced in part from: ***Brown, J.B.*** *and Thorpe, I.F., Biochemistry, 2015, 108: 1785-1795;* ***Brown, J.B****., Espiritu, M.V., Abraham, J. and Thorpe, I.F., Virus Research, Submitted.*

1.1 Protein Allostery and Allosteric Inhibition

Allosteric sites are pockets on a protein that are outside of the active site and can accommodate small molecules (termed allosteric effectors), which can activate or inhibit the protein's function. There are numerous examples of proteins that have endogenous allosteric effectors that can positively or negatively impact function, such as pyruvate carboxylase that can be activated by acetyl CoA or fructose-1,6-bisphosphatase that is inhibited by adenosine monophosphate.[1, 2] Because of the role of these and many other endogenous effectors in regulating biological function, researchers have used this principle to design drugs that mimic naturally occurring allosteric inhibitors or that target allosteric pockets to treat human diseases. To date, there are many human diseases that can be effectively treated with allosteric inhibitors. For example, there are currently five United States Food and Drug Administration approved allosteric inhibitors (also known as non-nucleoside reverse transcriptase inhibitors) of the polymerase of the human immunodeficiency virus (HIV). Efavirenz, one of these allosteric inhibitors, is one of the most potent and widely used drugs in HIV antiviral therapies.[3-5] As such, allosteric inhibitors represent an attractive group of drugs especially in viral infections because i) to be active *in vivo,* they do not require any phosphorylation events by cellular kinases like active site inhibitors such as nucleoside or nucleotide analogs and ii) since they are targeting a pocket outside of the active site, their

binding is often more specific resulting in less off-target interactions and potentially reducing cellular toxicities. An additional benefit of allosteric inhibitors is that they can be combined with each other (provided that they bind to non-overlapping sites) and/or with active site inhibitors to increase the inhibitory response against the viral protein target, which is often needed in the treatment of many viral infections. Although numerous proteins display allostery, the mechanisms by which these allosteric regulators impact protein function remain poorly identified. Consequently, the potential benefits of allosteric inhibitors warrant a deeper understanding of allosteric inhibition. In addition, there are many viral systems that are not as well studied as HCV, which may benefit from the insights garnered on the mechanisms of allosteric inhibition in the HCV polymerase. However, these less studied systems such as Dengue, West Nile and Foot-and-mouth Disease lack data on allosteric pockets in key enzymes (specifically the viral polymerases) and thus more investigations are needed to identify potential allosteric binding sites.

This dissertation serves to provide molecular insights into the mechanisms of allosteric inhibition of the HCV polymerase when two allosteric ligands are present as well as to identify possible allosteric pockets in the polymerases of several viruses. There are numerous studies that demonstrate an interplay between structure and function in the context of allosteric inhibition.[6-8] However, when the interplay between protein structure and function does not sufficiently describe the mechanism of allosteric inhibition, it has more recently been suggested that dynamics can help explain this type of protein behavior.[6, 9-12] Currently, there are two primary paradigms that describe a protein's behavior in the context of allosteric inhibition: i) induced fit and ii) conformational selection.[7, 10, 12] The induced fit model suggests that inhibitor binding causes the protein to undergo a conformational change to adopt a structural state that is unavailable to the free protein. In contrast, the conformational selection model of allostery suggests that there is a pre-existing population of all possible protein conformations and that inhibitor binding causes a shift in the equilibrium of sampled

conformations. Our results suggest that the HCV polymerase exhibits a mixed mode of allostery in the presence of two allosteric inhibitors, which is discussed in detail in Chapter 2.

1.2 Brief Overview of the Impacts of the Hepatitis C (HCV), Dengue (DENV), West Nile (WNV), Foot-and-mouth Disease (FMDV) Viruses on Human Health

1.2.1 HCV

HCV is a global health challenge, affecting approximately 200 million people worldwide, of which 4 million are Americans.[13, 14] In the United States, this viral infection results in cirrhosis of the liver and is the principal cause of liver transplantation.[13] There are six distinct genotypes of HCV with various subtypes. Of particular interest is genotype 1 (subtypes a and b), the most prevalent strain in North America. Infections by this genotype are particularly difficult to treat, with the current United States Food and Drug Administration approved standard of care for HCV (ribavirin and interferon alpha) being markedly less effective against this genotype.[15] Unfortunately, current therapies are ineffective in treating up to half of all HCV affected patients, especially those infected with genotype 1.[15, 16] Recently, Gilead Sciences introduced Sovaldi, a new direct acting nucleoside inhibitor (NI) targeting the HCV polymerase that exhibits reduced side effects and decreased treatment times. However, this drug is very expensive and most effective when used in combination with another active site inhibitor targeting a different enzyme or in combination with ribavirin and interferon alpha. However, treatment regimens involving ribavirin and interferon alpha are not ideal, as these therapeutics are known to induce adverse side effects. Consequently, there is still a need for new and less expensive polymerase inhibitors that could serve as therapeutics, as well as a need to understand the mechanisms of action of such molecules.

1.2.2 DENV and WNV

DENV and WNV are both clinically important viruses that, like HCV, are members of the flaviviridae family. Like HCV, both DENV and WNV possess positive-sense single-stranded RNA genomes. They are considered emerging pathogens whose outbreaks may lead to fatal outcomes.[17] About 50% of the world's population across 128 countries is at risk for DENV infection.[18, 19] As of 2010, there were an estimated 390 million DENV infection events globally, of which about 96 million possess some form of clinical severity.[18] WNV has also significantly increased in global prevalence since an outbreak in 1999 in which the virus was introduced to the United States and Europe.[20, 21] WNV epidemics from 1999 to 2006 resulted in 23,925 human disease cases and 946 deaths in the United States alone[17] and this virus continues to induce additional mortality.[20]

There have been efforts to discover and develop vaccines and antiviral therapies against DENV and WNV, but these have thus far been unsuccessful.[22-25] Currently, there are several vaccine candidates in clinical trials for DENV, [26, 27] although an approved vaccine is not yet available. In addition, inhibitors have been identified that target various enzymes critical to the life cycle of DENV.[28-30] An adenosine analog (an active site inhibitor) as well as an N-sulfonylanthranilic acid NNI (non-nucleoside thought to bind at the entrance of the template channel) of the DENV polymerase have been identified,[31] though these small molecules have not progressed to clinical trials. To date, there are at least 5 vaccine candidates for WNV that have entered clinical trials, with a trial for one of the most promising candidates beginning as recently as July 2015.[20, 32-34] There have also been several studies identifying small molecule inhibitors of the WNV polymerase.[25, 35, 36] Unfortunately, these small molecules are primarily broad spectrum active site antivirals that may result in nonspecific targeting of host proteins or fail during *in vivo* experiments.[35, 36]

1.2.3 FMDV

Unlike the other viral systems, FMDV is not a member of the flaviviridae family and, as will be discussed in Chapter 3, serves as an additional test of the robustness of our hypothesis centered on identifying allosteric pockets based on structural similarities (see section 1.5.1). FMDV is a member of the *Aphthovirus* genus in the picornaviridae family and infects cloven-hoofed domestic animals such as cattle in addition to more than 40 wild animal species.[21] The cost of a FMDV outbreak in the UK in 2001 has been valued at 6 billion pounds sterling and it is estimated that an epidemic of similar proportions would cost the United States about 100 billion dollars.[37] Due to an extremely high mutation rate during replication, there are about 65 subtypes of FMDV. This results in high antigenic diversity and makes it difficult to have a universal vaccine.[38, 39] The continued threat of costly outbreaks and lack of a universal vaccine warrants continued drug discovery efforts for FMDV. The FMDV polymerase serves as a challenging test of whether known structural information about the HCV polymerase can be used to garner useful insight into the properties of an enzyme with more significant dissimilarities in genome and structure than DENV or WNV. FMDV possesses a positive sense single strand RNA genome like HCV, DENV and WNV, as well as a polymerase with similar overall architecture to the corresponding enzymes from these viruses. However, there are more etiological and structural differences between FMDV polymerase and the other viruses than among the other three viral proteins, which will be discussed throughout section 1.5 and Chapter 4.

An obstacle limiting clinical success of allosteric inhibitors for the polymerases of FMDV, DENV and WNV is reduced potency with regard to *in vivo* cellular assays relative to that measured for *in vitro* studies.[23] One of the underlying causes for this observation is that polymerase regions outside of the active site and palm domains are subject to reduced evolutionary constraints to maintain their amino acid composition, as they are not directly involved in catalysis. Consequently, there is a lower barrier for the emergence of resistance as a result of amino acid substitutions. This situation is exacerbated by the fact that these

polymerases do not possess proofreading ability. Thus, during infection the virus typically exists as a collection of quasi-species, each with slightly different genetic composition, providing a fertile platform for the development of drug resistance. Nonetheless, it has been shown that applying multiple allosteric inhibitors in combination may diminish this problem.[23] Studies described in Chapter 2 of this document have begun to elucidate the molecular mechanisms underlying the enhanced inhibition obtained when multiple NNIs are presented to the HCV polymerase,[40] which may facilitate future efforts to employ similar strategies in other viral polymerases.

1.2.4 Significance

Although there have been recent advances in the treatment of HCV with the introduction of Sovaldi to the pharmaceutical market, the low fidelity of the HCV polymerase results in quasi-species of the virus. This leads to multiple NS5B variants and makes it difficult to target the enzyme with single inhibitors. Consequently, there is an urgent need to optimize combinations of ligands that can bind and produce an enhanced inhibitory effect against the HCV polymerase. Determining the interactions and thermodynamic properties that mediate allosteric effects in the presence of multiple ligands will provide a logical framework that will be useful in steering future small molecule optimization, developing inhibition strategies for combating NS5B drug resistance and improving the current understanding of how small molecule inhibitors affect protein function.

The strong structural similarities shared by the RNA-dependent RNA polymerases of HCV and the other viruses (DENV, WNV, FMDV) suggest that these other viral polymerases may exhibit allosteric sites analogous to those observed in the HCV enzyme. If so, it may be possible to identify novel small molecules that could inhibit these enzymes in a similar manner to that achieved for HCV. A challenge that prevents better understanding of allosteric inhibition in the polymerases of DENV, WNV and FMDV is limited biochemical and/or

structural information identifying allosteric sites, especially for the WNV enzyme. This makes the problem well suited for the application of computational tools to predict novel binding sites within these proteins. In doing so, we have placed emphasis on allosteric sites for two main reasons: i) active sites are well conserved across many viral families, thus predicting active site binding is not anticipated to be difficult and ii) in drug discovery for viral polymerases, allosteric sites tend to be unique to the virus. Thus, inhibitors that bind at these sites may reduce nonspecific side effects that can lead to host cellular toxicities. Our hypothesis is that we can use existing information available for the HCV polymerase to make meaningful inferences about the locations of allosteric sites in the DENV, WNV and FMDV polymerases due to the strong structural and functional similarities among these enzymes.

1.3 Structure of the HCV, DENV, WNV & FMDV RNA- dependent RNA Polymerases

1.3.1 HCV Polymerase

The HCV genome encodes several structural and non-structural proteins. The non-structural protein NS5B is an RNA-dependent RNA polymerase critical for viral replication [41] and is at the center of many biochemical and drug design studies. NS5B has three canonical polymerase domains (the palm, thumb and fingers regions) that encircle the active site (**figure 1**).[42, 43] Thus far, crystallographic data shows at least four distinct allosteric sites on NS5B to which non-nucleoside inhibitors (NNIs) bind, with two sites each in the palm and thumb domains (**figure 1**).[13, 44] Thumb sites I and II (known as NNI-1 and NNI-2) are located at the top and base of the thumb domain, respectively (see **figure 1**). The palm sites partially overlap and are differentiated based on palm site I (NNI-3) being located closer to the interface between the palm and thumb domains, while palm site II (NNI-4) extends into the Arginine 200 hinge region that is closer to the active site.[13, 44]

NNIs span a range of chemical scaffolds that can bind to different regions within the known binding sites. However, most fail after entry into clinical trials due to the development of unforeseen toxicities.[13, 45] Many studies have identified and optimized inhibitors specific to the active site as well as allosteric pockets of NS5B. Active site inhibitors have been more successful in the clinic, but have a higher risk of targeting host polymerases compared to allosteric inhibitors, the latter being more specific to viral polymerases.[13, 46, 47] Thus, by targeting the allosteric pockets of NS5B one may reduce the non-specific interactions that are problematic for active site inhibitors. Furthermore, NS5B contains multiple allosteric sites, which provides the possibility of using several NNIs in combination.

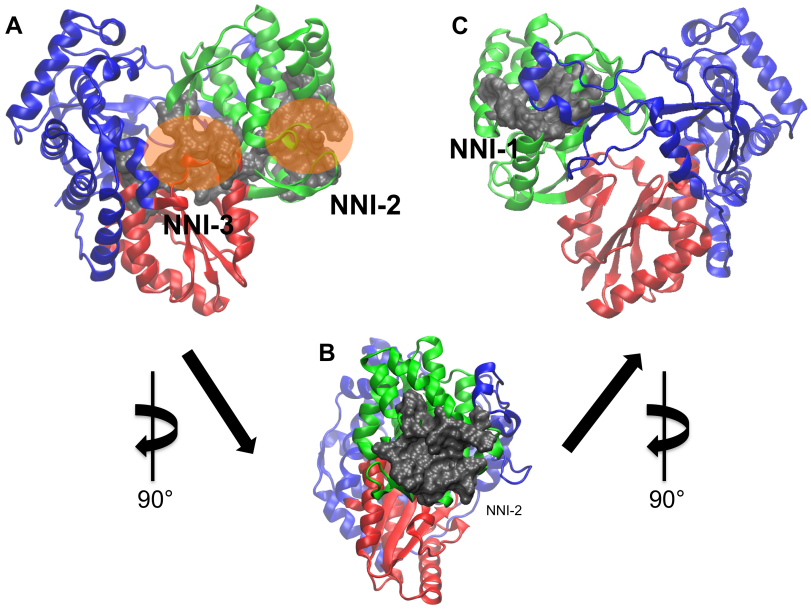


Figure 1.1: Allosteric Sites in the HCV polymerase. Protein residues are shown in ribbon representation while residues lining the binding pockets are shown in grey surface representation. Panel A shows a frontal view of the NNI-2 and NNI-3 sites, which are differentiated by the positions of orange translucent spheres. Panel B is a magnified lateral view of the NNI-2 site rotated 90° about the y-axis. Panel C is a rear view of the protein depicting the NNI-1 pocket after an additional 90° rotation about the y-axis. The thumb, palm and fingers domains are shown in green, red and blue respectively.

1.3.2 Resistance

Despite the positive features of NNIs noted in section 1.3.1, there are several challenges impeding the use of such inhibitors in the clinic. One such challenge is that genetic mutations can arise in NS5B and cause it to become resistant to NNIs. This problem

is exacerbated by the lack of proofreading activity in NS5B during replication, which results in low fidelity and an increased risk for mutations that makes it more likely for resistance to emerge.[13, 41, 45] Resistance occurs when an inhibitor has a significantly depreciated or lack of inhibitory effect against the enzyme after sustained usage. As a result, there is still an urgent need to circumvent HCV resistance to NS5B inhibitors such as NNIs. Insights from studies mitigating this need could guide us on future applications of combining NNIs with each other or in concert with an NI.

It is possible that viral resistance to NNIs can be overcome by employing multiple NNIs in concert. The simultaneous application of NNIs targeting non-overlapping binding sites has been shown to enhance inhibition of NS5B. For instance, studies involving the use of both benzothiazine (NNI-3) and benzimidazole (NNI-1) inhibitors have demonstrated enhanced inhibition relative to application of either compound individually (see **figure 1** for site locations).[47, 48] There have also been studies combining pyridine carboxamide (NNI-4) and dihydropyranone (NNI-2) inhibitors resulting in enhanced inhibition.[49] However, the molecular mechanisms that govern allosteric inhibition are still largely unknown, particularly when multiple allosteric inhibitors are applied in combination. In this work we employ molecular simulations and binding free energy calculations to understand the molecular mechanisms that govern enhanced inhibitory effects when allosteric ligands simultaneously bind to the NNI-2 and NNI-3 sites on the enzyme. Our results suggest that simultaneous binding of inhibitors at both sites is possible and that the presence of dual inhibitors induces synergistic changes to enzyme conformation and dynamics.

1.3.3 DENV, WNV and FMDV Polymerases

The polymerases of DENV, WNV and HCV are alike in their global protein folds[50] (**figures 1 and 2**). However, the full-length polymerases of DENV and WNV are larger than that of HCV, possessing a nuclear localization sequence and methyltransferase (MTase)

domain in addition to the RNA-dependent RNA polymerase domain for a total of about 900 residues.[22, 23] The FMDV polymerase is the smallest of the four enzymes, possessing less than 500 residues, and has not been reported to contain any nuclear localization sequences or MTase domain.[37] Unlike DENV and WNV, there has been an experimental study identifying and confirming the presence of an allosteric inhibitor binding site located in the fingers domain of the FMDV polymerase, proximal to, but not overlapping with the active site.[39] While there is no crystal structure containing a ligand bound to this site, mutagenesis has revealed the residues critical for inhibitor binding. Thus, this allosteric site provided another opportunity to test the robustness of the ligand binding site predictor (LBSP) selected in Chapter 3.

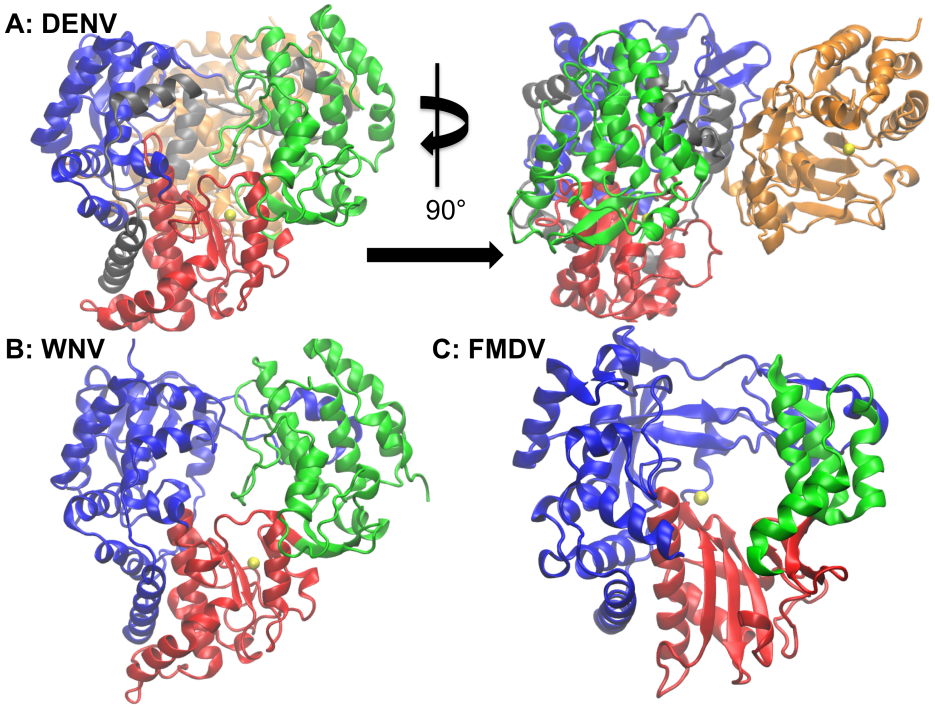


Figure 1.2: Polymerases of DENV, WNV and FMDV. The enzyme subdomains are colored as in Figure 1: this color scheme is maintained throughout the remaining figures. The figure image in panel A shows the frontal view of the DENV polymerase with a magnesium ion (yellow) present in the active site while the second image depicts the DENV enzyme rotated 90° about the y-axis to allow for the methyltransferase (MTase) domain (colored in gold) to be visible. Panel C is the WNV polymerase with a magnesium ion (yellow) in the active site. While WNV also possesses an MTase domain, a full-length structure containing both the polymerase and MTase domains has not yet been solved. Finally, panel C shows the FMDV with a catalytic manganese ion (yellow).

1.4 Molecular Dynamics Simulations

As mentioned in section 1.1, dynamics may play an integral role in our understanding of allosteric inhibition of proteins. To this end, molecular dynamics (MD) simulations provide a way to relate protein motions to functional properties.[51-54] Using approximations of Newton's equations of motion, MD simulations allow us to compute time-dependent properties of a target protein. As a result, we are able to calculate dynamic and thermodynamic properties of a given protein system.

A typical MD simulation of an enzyme requires placing the protein in a three-dimensional volume that is symmetric (such as an octahedron). Solvent molecules such as water in addition to ions such as chloride are also placed inside the volume and used to generate a periodic lattice. Consequently, the molecules are able to experience some effects of being realistically solvated due to the periodicity of the system.[55] The first step of most MD simulation protocols is energy minimization, which serves to reduce the energy of the system via modifying the atomic positions. After minimizing the total energy, MD simulations of biomolecules are typically executed under constant temperature in concert with either constant volume or constant pressure until the protein conformational ensemble fluctuates about a well-defined average. An assumption is that protein motions at this point will be equivalent to equilibrium motions of an actual solvated system. MD simulations provide several outputs. Of particular interest is the trajectory file, which consists of the atomic coordinates as a function of time for a given system and allow for visualization of the system. The trajectory output can also be employed in several analyses,[55, 56] some of which were used to answer one of the key research questions at hand (see Chapter 2).

MD simulation is an attractive tool because it provides information on both structure and dynamics. Although X-ray crystallography has elucidated a large amount of structural information regarding the polymerases of NS5B, DENV, WNV and FMDV, this technique is limited to a static snapshot of protein structure and provides very limited dynamic insights. Accordingly, since dynamics may be the missing link between structure and function in our

understanding of allosteric inhibition, x-ray crystallography is not ideal. There are techniques such as nuclear magnetic resonance (NMR) that can provide dynamic information for proteins. However, for large biomolecules such as the HCV polymerase (approximately 66 kilodaltons), it is difficult to employ NMR as the signals from different parts of the protein become hard to differentiate.[57] Thus, MD represents an apt technique to study both dynamics and structure at a molecularly detailed level.[55, 56] However, MD simulations can be limited in their ability to capture physiological timescales; in an effort to have more representative ensembles one can either run simulations in replicates or for longer time periods. For the MD simulations discussed in Chapter 2, we used long trajectories, on the order of 1 microsecond.

1.5 Dissertation outline

We hypothesize that the simultaneous binding of two allosteric inhibitors will induce structural and dynamic changes that are responsible for their enhanced inhibitory impact on the HCV polymerase and will provide insights into the mechanisms mediating enhanced inhibition. Thus, comparisons of ligand-free and ligand-bound MD simulations will provide insights into any structural or dynamic changes that may mediate enhanced inhibition. Our second hypothesis is that we can use HCV polymerase as a model to determine if putative allosteric sites are present in the polymerases of the less studied DENV, WNV and FMDV in a first step to understanding enhanced inhibition in those systems.

1.5.1 Specific Aims:

We will address our hypotheses through the following specific aims:

i) Evaluate the structural, dynamic and thermodynamic changes mediating enhanced inhibition in the HCV polymerase

ii) Determine if the structural similarities that exist between the polymerases of HCV and DENV/WNV/FMDV correspond to presence of allosteric sites in the enzymes of the less studied viral systems.

1.5.2 Objectives

We accomplish our specific aims via the research objectives listed below:

I. ***Determine how changes in protein dynamics and conformation mediate enhanced inhibition in the HCV polymerase.*** The hypothesis is that enhanced inhibition is achieved through the disruption of correlated motions and changes to the conformational space when multiple allosteric inhibitors are simultaneously bound. (Specific Aim 1, Chapters 2)

II. ***Evaluate how changes in binding free energies mediate enhanced inhibition in the HCV polymerase (NS5B).*** The working hypothesis is that the binding free energy for the doubly bound system will be more favorable than the sum of the binding free energies for the individually bound systems. (Specific Aim 1, Chapters 2)

III. ***Evaluate several ligand binding site predictor (LBSP) tools to determine the algorithm with the best predictive power at identifying putative allosteric pockets in the polymerases of Dengue, West Nile and Foot-and-mouth Disease viruses using NS5B as a model system.*** The HCV polymerase represents a great model system with multiple allosteric sites, which will aid in identifying a LBSP that is apt for predicting allosteric sites in viral polymerases. (Specific Aim 2, Chapter 3)

IV. ***Identify putative allosteric sites in the polymerases of Dengue, West Nile and Foot-and-mouth Disease viruses.*** Our working hypothesis is that because there are structural similarities between the HCV polymerase and the other viral polymerases, they may also possess allosteric sites and these pockets may share structural properties with those of the HCV polymerase. (Specific Aim 2, Chapter 3)

1.6 Chapter Overview

The primary goals of this dissertation are to understand mechanisms of enhanced inhibition in the HCV polymerase and to identify putative allosteric sites within the polymerases of DENV, WNV, FMDV.. In Chapter 2, we employ molecular dynamics (MD) simulations to understand the changes to structure, dynamics and thermodynamics when different combinations of palm- and thumb- allosteric inhibitors are simultaneously bound to the HCV polymerase. Our findings suggest that NNI-3 and NNI-2 inhibitors can bind simultaneously in an energetically favorable manner, jointly impacting the free energy landscape and protein motions. Specifically, in the presence both of these inhibitors we observe a mixed of induced-fit and conformational selection modes of allosteric behavior. Chapter 3 discusses the evaluation of several ligand binding site predictor (LBSP) tools using the HCV polymerase as a model system and subsequent use of the LBSP that was most successful for the HCV polymerase to identify putative allosteric sites in the polymerases of DENV, WNV and FMDV. We observed that all three enzymes had putative allosteric sites that shared structural and/or chemical similarities to two validated allosteric sites in the HCV enzymes. Lastly, Chapter 4 summarizes the implications of our findings and describes future research efforts.

1.7 References

[1] Adina-Zada, A. Z., T. N.; St. Maurice, M.; Jitrapakdee, S.; Cleland, W. W.; Attwood, P. V. (2012) Allosteric regulation of the biotin-dependent enzyme pyruvate carboxylase by acetyl-CoA, *Biochemical Society Transactions 40*, 567-572.

[2] Hines, J. K. K., C. E.; Fromm, H. J.; Honzatko, R. B. (2007) Structure of inhibited fructose-1,6-bisphosphatase from Escherichia coli: distinct allosteric inhibition sites for AMP and glucose 6-phosphate and the characterization of a gluconeogenic switch, *Journal of Biological Chemistry 282*, 24697-24706.

[3] Larru, B. E., J.; Lowenthal, E.D. (2014) Antiretroviral treatment in HIV-1 infected pediatric patients: focus on efavirenz, *Pediatric Health, Medicine and Therapeutics 5*, 29-42.
[4] Rakhmanina, N. Y. v. d. A., J. N. (2010) Efavirenz in the therapy of HIV infection, *Expert Opinion on Drug Metabolism & Toxicology 6*, 95-103.
[5] Bastos, M. M. C., C. P.; Bezerra, T. C.; da Silva, F. C.; Boechat, N. (2016) Efavirenz a nonnucleoside reverse transcriptase inhibitor of first-generation: Approaches based on its medicinal chemistry, *European Journal of Medicinal Chemistry 108*, 455-465.
[6] Bustos-Jaimes, I. (2008) Allostery and co-operativity in protein function, *Advances in Protein Physical Chemistry*, 219-248.
[7] Monod, J., Wyman, J., and Changeux, J. P. (1965) On the Nature of Allosteric Transitions: A Plausible Model, *J Mol Biol 12*, 88-118.
[8] Tsai, C. J., del Sol, A., and Nussinov, R. (2008) Allostery: absence of a change in shape does not imply that allostery is not at play, *J Mol Biol 378*, 1-11.
[9] Clarkson, M. W., Gilmore, S. A., Edgell, M. H., and Lee, A. L. (2006) Dynamic coupling and allosteric behavior
in a nonallosteric protein, *Biochemistry 45*, 7693–7699.
[10] Koshland, D. E., Jr., Nemethy, G., and Filmer, D. (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits, *Biochemistry 5*, 365-385.
[11] Popvych, N. (2006) Dynamically Driven Protein Allostery, *Nature Structural & Molecular Biology 13*, 831838.
[12] Tsai, C.-J., Sol, Antonio del, Nussinov, Ruth. (2008) Allostery: Absence of a Change in Shape Does Not Imply that Allostery Is Not at Play, *Journal of Molecular Biology 378*, 1-11.
[13] Beaulieu, P. L. (2009) Recent advances in the development of NS5B polymerase inhibitors for the treatment of hepatitis C virus infection, *Expert Opin. Ther. Patents 19*, 145-164.
[14] Leonard B. Seff, J. H. H. (2003) National Institutes of Health Consensus Development Conference: Management of Hepatitis C: 2002, *Hepatology 36*, s1 - s2.
[15] Mann, M. P., McHutchison, J. G., Gordon, S. C., Rustgi, V. K., Shiffman, M., Reindollar, R., Goodman, Z. D., Koury, K., Ling, M., and Albrecht, J. K. (2001) Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial, *Lancet 358*, 958-965.
[16] Koch, U. (2006) Allosteric Inhibition of the Hepatitis C Virus NS5B RNA Dependent RNA
Polymerase, *Infectious Disorders - Drug Targets 6*, 31-41.
[17] Malet, H. E., M.-P.; Selisko, B.; Butcher, R. E.; Wright, P. J.; Roberts, M.; Gruez, A.; Sulzenbacher, G.; Vonrhein, C.; Bricogne, G.; Mackenzie, J. M.; Khromykh, A. A.; Davidson, A. D.; Canard, B. . (2007) Crystal Structure of the RNA Polymerase Domain of the West Nile Virus Non-structural Protein 5, *Journal of Biological Chemistry 282*, 10678 - 10689

[18] Bhatt, S. G., P. W.; Brady, O. J.; Messina, J. P.; Farlow, A.W.; Moyes, C. L. et. al. . (2013) The global distribution and burden of dengue, *Nature 496* 504 - 507.
[19] Brady, O. J. G., P. W.; Bhatt, S.; Messina, J. P.; Brownstein, J. S.; Hoen, A. G.; Moyes, C. L.; Farlow, A. W.; Scott, T. W.; Hay, S. I. . (2012) Refining theGlobal Spatial Limits of Dengue Virus Transmission by Evidence-Based Consensus, *PLOS Neglected Tropical Diseases 6*.
[20] Brandler, S. T., F. (2013) Vaccines in Development against West Nile Virus, *Viruses 5*, 2384 - 2409
[21] Grubman, M. J. B., B. (2004) Foot-and-Mouth Disease, *Clinical Microbiology Reviews 17*, 465 - 493
[22] Malet, H. M., N.; Selisko, B.; Romette, J.-L.; Alvarez, K.; Guillemot, J. C.; Tolou, H.; Yap, T. L.; Vasudevan, S.; Lescar, J.; Canard, B. (2008) The flavivirus polymerase as a target for drug discovery, *Antiviral Research 80*, 23 - 35.
[23] Noble, C. G. C., Y.-L.; Dong, H.; Gu, F.; Lim, S. P.; Schul, W.; Wang, Q.-Y.; Shi, P.-Y. (2010) Strategies for development of dengue virus inhibitors *Antiviral Research 85*, 450 - 462
[24] Chappell. (2008) West Nile Virus NS2B/NS3 protease as an antiviral target, *Current Medicinal Chemistry*.
[25] Gu, B. O., S.; Wang, L.; Mason, P.; Bourne, N.; Cuconati, A.; Block, T. M. . (2006) Discovery of small molecule inhibitors of West Nile virus using a high-throughput sub-genomic replicon screen, *Antiviral Research 70* 39 - 50
[26] Blaney, J. E. M., J. M.; Murphy, B. R.; Whitehead, S. S. (2005) Recombinant, Live-Attenuated Tetravalent Dengue Virus Vaccine Formulations Induce a Balanced, Broad, and Protective Neutralizing Antibody Response against Each of the Four Serotypes in Rhesus Monkeys, *Journal of Virology 79*, 5516 - 5528
[27] Thisyakorn, U. T., C. . (2014) Latest developments and future directions in dengue vaccines *Therapeutic Advances in Vaccines 2*, 3 - 9
[28] Smith. (2015) Identifying Initiation and Elongation Inhibitors of Dengue Virus RNA Polymerase in a High-Throughput Lead-Finding Campaign, *Journal of Biomolecular Screening*
[29] Lim, S. W., Q.; Noble, C. et al. . (2013) Ten Years of Dengue Drug Discovery: Progress and Prospects, *Antiviral Research 100*, 500 - 519
[30] Schwartz, L. M. H. M. E. D., A. P.; Longini, I. M. . (2015) The dengue vaccine pipeline: Implications for the future of dengue control, *Vaccine 33*, 3293 - 3298.
[31] Yin, Z. C., Y.-L.; Schul, W.; Wang, Q.-Y; Gu, F.; Duraiswamy, J.; Kondreddi, R. R.; Niyomrattanakit, P.; Lakshminarayana, S. B.; Goh, A.; Xu, H. Y.; Lui, W.; Liu, B.; Lim, J. Y. H.; Ng, C. Y.; Qing, M.; Lim, C. C.; Yip, A.; Wang, G.; Chan, W. L.; Tan, H. P.; Ling, K.; Zhang, B.; Zou, G.; Bernard, K. A.; Garrett, C.; Beltz, K.; Dong, M.; Weaver, M.; He, H.; Pichota, A.; Dartois, V.; Keller, T. H.; Sh, P.-Y. . (2009) An adenosine nucleoside inhibitor of dengue virus, *Proceedings of the National Academy of Sciences*

[32] Pinto, A. K. R., J. M.; Poore, E. A.; Patil, P. P.; Amanna, I. J.; Slifka, M. K.; Diamond, M. S. . (2013) A Hydrogen Peroxide-Inactivated Virus Vaccine Elicits Humoral and Cellular Immunity and Protects against Lethal West Nile Virus Infection in Aged Mice, *Journal of Virology 87*, 1926 - 1936.
[33] Petersen, L. R. (2014) *West Nile Virus: From Africa to Europe, America, and Beyond*, Springer, Netherlands.
[34] Yamshchikov, V. (2015) Development of human live attenuated West Nile infectious DNA vaccine: Conceptual design of the vaccine candidate *Virology Journal 484*, 59 - 68
[35] Furuta, Y. G., B. B.; Shiraki, K. T. K.; Smee, D. F.; Barnard, D. L. (2015) Favipiravir (T-705), a novel viral RNA polymerase inhibitor, *Antiviral Research 100*, 446 - 454
[36] Jordan, I. B., T.; Fischer, N.; Lau J. Y.-N.; Lipkin, W. I. . (2000) Ribavirin Inhibits West Nile Virus Replication and Cytopathic Effect in Neural Cells, *The Journal of Infectious Diseases 182*, 1214 - 1217.
[37] Ferrer-Orta, C. A., A.; Perez-Lugue, R., Escarmís, C.; Domingo, E.; Verdaguer, N. (2004) Structure of Foot-and-Mouth Disease Virus RNA-dependent RNA Polymerase and Its Complex with a Template-Primer RNA, *Journal of Biological Chemistry 279*, 47212 - 47221.
[38] Domingo, E. B., E.; Escarmís, C.; Sobrino, F. (2002) Foot and mouth disease virus, *Comparative Immunology, Microbiology and Infectious Diseases 25*, 297 - 308
[39] Durk, R. C. S., K.; Cornelison, C. A.; Rai, D. K.; Matzek, K. B.; Leslie, M. D.; Schafer, E.; Marchand, B.; Adedeji, A.; Michailidis, E.; Dost, C. A.; Moran, J.; Pautler, C.; Rodriguez, L. L.; McIntosh, M.A.; Rieder, E.; Sarafianos, S. G. (2010) Inhibitors of Foot and Mouth Disase Virus Targeting a Novel Pocket of the RNA-dependent RNA Polymerase, *PLoS ONE 5*, e15049.
[40] Brown, J. A. T., I. F. (2015) Dual Allosteric Inhibitors Jointly Modulate Protein Structure and Dynamics in the Hepatitis C Virus Polymerase *Biochemistry 54*, 4131 - 4141.
[41] Lindenbach, B. D., and Rice, C. M. (2005) Unravelling hepatitis C virus replication from genome to function, *Nature 436*, 933-938.
[42] Lesburg, C. A. (1999) Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site, *Nature Structural Biology 6*, 937-943.
[43] Bressanelli, S. (1999) Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus, *PNAS 96*, 13034-13039.
[44] Li, H. (2009) Discovery of (R)-6-Cyclopentyl-6-(2-(2,6- diethylpyridin-4-yl)ethyl)-3-((5,7-dimethyl- [1,2,4]triazolo[1,5-a]pyrimidin-2-yl)methyl)-4-hydroxy-5,6-dihydropyran-2-one (PF-00868554) as a Potent and Orally Available Hepatitis C Virus Polymerase Inhibitor, *Journal of Medicinal Chemistry 52*, 1255-1258.
[45] Burton, J. R., Jr., and Everson, G. T. (2009) HCV NS5B polymerase inhibitors, *Clin Liver Dis 13*, 453-465.

[46] Beaulieu, P. L. (2007) Non-nucleoside inhibitors of the HCV NS5B polymerase: progress in the discovery and development of novel agents for the treatment of HCV infections, *Curr Opin Investig Drugs 8*, 614-634.
[47] Condon, S. M. (2005) Allosteric Inhibitors of Hepatitis C NS5B RNA-dependent RNA Polym-
erase, *Curr. Med. Chem. - Anti-Infective Agents 4*, 99-110.
[48] Le Pogam, S. (2006) Selection and Characterization of Replicon Variants Dually Resistant to Thumb- and Palm-Binding Nonnucleoside Polymerase Inhibitors of the Hepatitis C Virus, *Journal of Virology 80*, 6146-6154.
[49] Cheng, C. C. H., Xiaohua ; Shipps Jr., Gerald W.; Wang, Yu-Sen; Wyss, Daniel F.; Soucy, Kyle A.; Jiang, Chuan-kui; Agrawal, Sony; Ferrari, Eric; He, Zhiqing; Huang, H.-C. . (2010) Pyridine Carboxamides: Potent Palm Site Inhibitors of HCV NS5B Polymerase, *ACS Medicinal Chemistry Letters 1*, 466 - 471.
[50] Yap, T. L. X., T.; Chen, Y.-L.; Malet, H.; Egloff, M.-P.; Canard, B.; Vasudevan, S. G.; Lescar, J. . (2007) Crystal Structure of the Dengue Virus RNA-Dependent RNA Polymerase Catalytic Domain at 1.85-Angstrom Resolution, *Journal of Virology 81*, 4753 - 4765
[51] Berendsen, H. J. H., S. (2000) Collective protein dynamics in relation to function, *Current Opinion in Structural Biology 10*, 165-169.
[52] Fenwick, R. B. E.-M., S.; Salvatella, X. (2011) Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles, *European Biophysics Journal 40*, 1339-1355.
[53] Lange, O. F. G., H. (2006) Can Principle Components Yield a Dimension Reduced Description of Protein Dynamics on Long Time Scales?, *Journal of Physical Chemistry B 110*, 22842-22852.
[54] Rod, T. H. R., J. L.; Brooks, C. L., III. (2003) Correlated motion and the effect of distal mutations in dihydrofolate reductase, *Proceedings of the National Academy of Sciences 100*, 6980-6985.
[55] Leach, A. R. (2001) *Molecular Modeling: Principles and Applications*, 2nd ed., Pearson Education Limited, Essex, England.
[56] Cramer, C. J. (2004) *Essentials of Computational Chemistry: Theories and Models*, 2nd ed., Wiley, West Sussex, England.
[57] Fernandez, C. W., G. (2003) TROSY in NMR studies of the structure and function of large biological macromolecules, *Current Opinion in Structural Biology 13*, 570-580.

Chapter 2: Dual Inhibitors Jointly Modulate the Protein Structure and Dynamics of the HCV Polymerase

This chapter is in part reproduced from: ***Brown, J. A.*** *and Thorpe, I. F., Biochemistry,* ***2015****, 54, 4131 – 4141*.

2.1 Introduction

As mentioned in Chapter 1.5.1, one of the primary objectives of this dissertation is to evaluate the structural, dynamic and thermodynamic changes mediating enhanced inhibition in the HCV polymerase. We carried out molecular dynamics (MD) simulations of the free polymerase or bound to one or two inhibitors (each system is described in Table 2.1). We observed that both allosteric inhibitors were able to simultaneously modulate the enzyme's conformational space and dynamics in novel ways compared to the free and individually bound systems. In addition, free energy calculations of the dually bound enzyme system indicate the computed binding affinity was more favorable than the sum of that observed for the individually bound systems. This observation suggests that the binding of both ligands is energetically favorable and provides further insights into a model of enhanced inhibition of the HCV polymerase.

In the following discussion we provide evidence supporting the functional significance of the novel changes to the conformational space and dynamics in the dually bound enzyme. Specifically, we observed conformations in the dually bound system that were more closed compared to the free or individually bound systems that

may represent a state that is not functionally competent. We also observed changes in the protein motions of key motifs involved in various steps of the replication process such as RNA template binding. Thus, the findings of this study further our understanding of how the behavior of the HCV polymerase may be altered simultaneously by two allosteric inhibitors, which is useful in defining the determinants of enhanced inhibition. Such insights may help to frame our understanding of allosteric inhibition in the context of two inhibitors for other viral polymerases such as those from DENV and WNV.

System Components [Identifier]	Total Simulation Time (ns)	Truncated Octahedron Unit Cell Length (Å)	Total Number of Water Residues	Total Number of Atoms
1quv [FREE]	1000	91.5244	17360	60391
1quv+vgi [NNI2]	1000	91.6288	17343	60381
1quv+3ms [NNI3]	1000	91.5259	17353	60426
1quv+3ms+vgi [DUAL]	1002	91.5746	17331	60401
3MS	10	35.3453	1101	3359
VGI	10	34.4103	1030	3131

Table 2.1: Each system with its corresponding identifier, total simulation length, cell length, total number of water atoms and total number of atoms in each simulation.

2.2 Materials and Methods

2.2.1 Biomolecular Systems

A crystal structure of NS5B from HCV genotype 1b (PDB ID 1QUV) was used in our simulations (see figure 2.1). 1QUV was selected because it has a relatively high resolution of 2.5 Å and does not contain any ligand. This allows us to clearly observe the impact of ligand binding on enzyme structure and dynamics. 1QUV consists of a single amino acid chain containing 578 residues. However, the last 47 C-terminal residues were

removed because they are not essential for *in vitro* RNA replication.[1, 2] These residues may play a role in docking the virus to the endoplasmic reticulum of the host in vivo.[1, 3] Prior work from our group has shown that these residues reduce the conformational sampling and overall flexibility of the enzyme. They also induce a conformation that may disfavor de novo initiation *in vitro*.[4] The three-dimensional structure of 1QUV was solved without metal ions coordinated within the active site. However, two magnesium ions were added to the catalytic site of the structure in all simulations because they have been shown biochemically to be needed for efficient enzyme function[5, 6] and have been observed to fundamentally alter the structure and dynamics of the enzyme in previous molecular simulations by our group.[2] Two allosteric inhibitors 3MS and VGI were employed in this work. Both inhibitors have been separately co-crystallized with NS5B of genotype 1b and show nanomolar activities against the enzyme individually (3MS: IC_{50} < 10nM [from PDB ID: 3CO9], VGI: IC_{50} = 20nM [from PDB ID: 2WHO]).[7, 8] These ligands were chosen because they bind to non-overlapping allosteric sites that have been implicated in enhanced inhibition of NS5B when used in combination.[9-11] In addition, the ligands individually have been used in previous simulations, providing a way to independently assess the robustness of our current findings.[2, 4] To understand the impact of each inhibitor on the enzyme, we performed simulations where only one of each ligand was bound (these will be referred to as NNI3 and NNI2 for 3MS and VGI, respectively throughout the paper), while our control system was that of the free protein (FREE). Our fourth system consisted of 1QUV bound to both ligands (DUAL). In each simulation, 19 chloride ions were added to neutralize the total system charge. Each system was solvated with explicit TIP3P water molecules within a truncated octahedral unit cell larger than the protein by 10 Å in each dimension, resulting in an edge length of at least 91 Å (**table 2.1**).

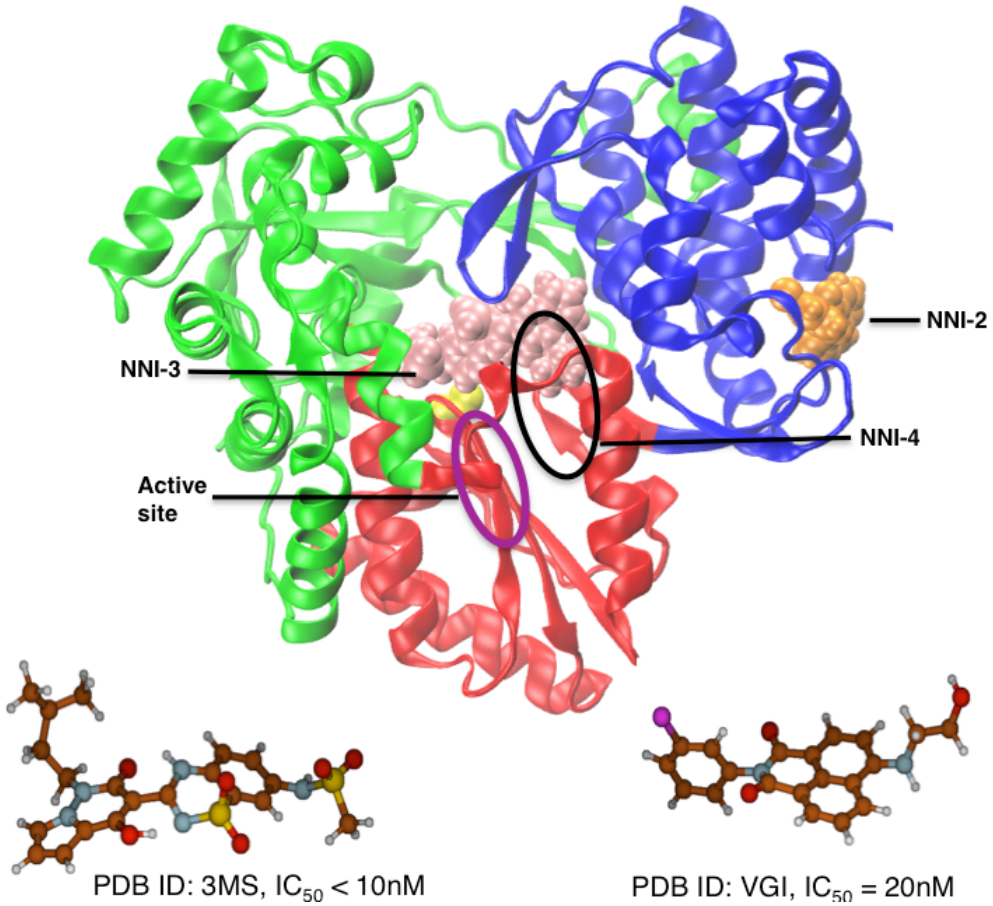


Figure 2.1: Structure of the Hepatitis C Virus Polymerase (NS5B) depicting three allosteric binding sites. The three domains are shown in red (palm), blue (thumb) and green (fingers). The two magnesium ions needed for efficient viral replication are depicted as yellow van der Waals spheres and the active site encircled by the magenta oval. The two allosteric inhibitors employed in the current work are shown as van der Waals spheres in their respective binding sites (VGI in orange and 3MS in pink).

2.2.2 Minimization and Molecular Dynamics Simulations

The force field parameters for VGI and 3MS were taken from the CHARMM general force field versions 2b6 and 2b7, respectively.[12] Parameters not available in the ligand force field were obtained using the procedure described by the MacKerell and co-workers.[13, 14] The CHARMM27/CMAP protein force field was used to describe the protein.[15-17] Before beginning the simulations, the solvated free enzyme and ligand-bound systems were minimized with NAMD version 2.7 using the steepest descent method for a total of 1000 steps while employing the while employing the CHARMM27/CMAP protein force field.[15-18] After minimization, all MD simulations were carried out at a temperature of 300K. Initially the protein backbone for each system was restrained using a force constant of 10 kcal/mol*A^2 applied to α-carbons during 5 ns of initial NVT simulation. This was followed by 5 ns of simulations in the NPT ensemble employing the Berendsen barostat, maintaining a pressure

of 1.01 Bar and without any restraints on the protein backbone. Finally, production runs were carried out in the NVT ensemble. At each stage the temperature was maintained via velocity reassignment for at least every 1000 steps for FREE, NNI2 and NNI3. For DUAL, a Langevin thermostat with friction coefficients of $10ps^{-1}$ and $1ps^{-1}$ was applied to non-hydrogen atoms for the NPT and NVT runs, respectively (a value of 5 ps^{-1} was employed for the first 80ns of NVT). Different thermostats were explored to achieve optimal computational efficiency and their use was not observed to alter the simulation properties. For all systems, we carried out at least a total of 1 μs of production simulations in 10 ns increments with an integration time step of 2 femtoseconds. During the production simulations, trajectory coordinates were written out at least every 100 ps. For all analyses we utilized the last 500ns of each trajectory.

We also performed MD simulations containing only the ligands in order to compute the ligand entropy contribution to relative binding free energies. We utilized the same minimization protocol for the solvated systems of VGI and 3MS (see **table 2.1** for solvated systems) followed by 5 ns of MD in the NPT ensemble using the Berendsen barostat to maintain an average pressure of 1.01 Bar and the Langevin thermostat with a friction coefficient of 1 ps^{-1} applied to non-hydrogen atoms. The Langevin thermostat was also employed to perform 10 ns of NVT production runs for each solvated ligand system.

2.2.3 Covariance

One mechanism by which allosteric inhibitors may work is by disrupting motions that are critical to protein function. Consequently, we performed covariance analysis as shown in equation 1 below to understand the characteristic motions of protein atoms within the trajectories. The covariance C_{ij} between sites i and j is defined by:

$$Ci,j = \frac{\left\langle \left(r_i - \left\langle r_i \right\rangle\right) \cdot \left(r_j - \left\langle r_j \right\rangle\right) \right\rangle}{\sqrt{\left\langle \left(r_i - \left\langle r_i \right\rangle\right)^2 \right\rangle \left\langle \left(r_j - \left\langle r_j \right\rangle\right)^2 \right\rangle}} \qquad (2.1)$$

In equation 2.2, r_i represents the position of the center-of-mass for residue i and the angle brackets denote ensemble averages.

2.2.4 Principal Component Analysis

$$[V][C_{i,j}][V^T] = diag[\lambda_1, \lambda_2, \lambda_3, ...] \quad (2.2)$$

To reduce the dimension of our data and more readily identify the principal motions in each system, we performed principal component analyses (PCA) of the covariance matrix. The covariance matrix (C_{ij}) was diagonalized using the eigenvector matrix (V) to generate eigenvalues (λ). The eigenvectors and eigenvalues comprise the principal components of the NS5B motions. PCA aids in characterizing changes in atomic motions by allowing us to examine the primary vibrational modes within each biomolecular system.

2.2.5 Root-mean squared fluctuations

$$RMSF_i = \sqrt{\frac{\sum_{t_j}^{T} (r_i(t_j) - \tilde{r}_i)^2}{N}} \quad (2.3)$$

The RMSF provides a measure of local flexibility in the protein. We calculated the RMSF for each residue center-of-mass by aligning every snapshot to the initial crystal structure and summing the difference between the instantaneous position of each center-of-mass at time t_j, ($r_i(t_j)$), and its corresponding reference position ($\tilde{r}_i$) for every snapshot. This difference is then divided by the total number of snapshots (N) and the square root taken of this result.

2.2.6 Describing the Enzyme Conformational Space

Previous studies indicate that at least two protein conformations are necessary for viral replication.[19, 20] A closed conformation is thought to be essential for the initiation stage of replication, while an open conformation is important for the elongation phase. To monitor

the different protein conformations adopted throughout the simulations, we used two structural metrics: i) ***Inter-domain Angle***: This is the angle between the fingers, palm and thumb subdomains and allows us to determine how open or closed the enzyme is during a given trajectory. The inter-domain angle was computed via measuring the angle between the centers-of-mass for each protein subdomain. ii) ***Template Channel Width***: the template channel is critical for the binding of RNA template. Consequently, we measured the widths of the channel to understand the impact of inhibitor binding on this crucial structural element. To probe the template channel width, the distance between the centers-of-mass of residues methionine 139 and valine 405 was calculated. More open states of the enzyme are associated with larger values of these two metrics.

2.2.7 Hydrogen Bonding

We employed the hydrogen bond (Hbond) utility in CHARMM to calculate Hbond distances and lifetimes of each system. We used the default hydrogen-acceptor probe distance of 2.4 Å and counted hydrogen bonds as interactions with occupancies of at least 10% for protein-ligand interactions and 20-90% for protein-protein interactions.

2.2.8 Binding Free Energies

To assess relative binding free energies we used the Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) free energy method. We employed the CHARMM Generalized Born Molecular Volume module (analytical method II) for these calculations using a surface area coefficient of 0.00542 kcal/mol*A^2.[21-24] The ligand binding free energy ΔG was computed according to $\Delta G = \Delta H - T\Delta S$ for the singly and doubly bound ligand simulations as follows:

i) for a singly-bound system:

$$\Delta G_x$$

$$NS5B(aq) + L(g) \rightarrow NS5B\text{-}L(aq) \quad (2.4)$$

ii) for the doubly-bound system:

$$NS5B(aq) + L1(g) + L2(g) \xrightarrow{\Delta G_x} NS5B\text{-}L1\text{-}L2(aq) \quad (2.5)$$

In equations 2.4 and 2.5 above, *L* denotes either the NNI2 or NNI3 ligand that is present in the protein-ligand complex on the right hand side, while *x* corresponds to ΔG with respect to NNI2, NNI3 or DUAL. ΔH values were calculated using equation 2.6 below. ΔS was calculated using equation 2.7 and employed the quasiharmonic approximation as implemented in the CHARMM VIBRAN module to evaluate the protein or ligand entropy. In each case only the 1% of modes with lowest frequencies were employed for the entropy calculations. This choice was made because the lowest frequency modes typically account for most of the overall fluctuation in molecular systems. Thus, focusing on these modes generates a more representative description of the entropy changes mediated by large amplitude fluctuations. It is these fluctuations (rather than small amplitude, high frequency vibrations) that are most likely to be functionally relevant in biomolecular systems. The calculation was performed using trajectories of the enzyme residues centers-of-mass for the protein entropy or using ligand all-atom trajectories for the ligand entropy. For equation 2.7, $S_{free\ protein}$ and $S_{free\ ligand}$ refer to the entropies computed from the simulations of only the protein or ligand, respectively. Note that this treatment neglects the changes in rotational and translational ligand entropy and captures the solvation entropy only qualitatively via the impact of the Generalized Born solvation term on ΔH. Nonetheless, we anticipate that these quantities are consistent for either ligand, allowing the trends in binding affinity to be well reproduced when comparing the binding of individual ligands for NNI2 and NNI3 to that of DUAL.

$$\Delta H = H_{\text{complex}} - H_{\text{protein from complex}} - H_{\text{ligand from complex}} \quad (2.6)$$

$$\Delta S = (S_{\text{protein from complex}} - S_{\text{free protein}}) + (S_{\text{ligand from complex}} - S_{\text{free ligand}}) \quad (2.7)$$

In this context, one would anticipate that synergistic binding in DUAL would be reflected in a more favorable binding free energy for the DUAL system than for the sum of the individual binding free energies for NNI2 and NNI3. Thus, synergy in binding would be consistent with $\Delta G_{DUAL} < \Delta G_{NNI2} + \Delta G_{NNI3}$.

2.2.9 Standard Error of ΔH

To compute the standard error of the ΔH for each of the system, we needed to determine the statistical inefficiency, s, which is the time point at which the simulation data becomes uncorrelated. In order to determine s, we used the block averaging method, in which our entire dataset was divided into blocks (b) of varying lengths and the variance of the means for the different blocks $\sigma^2(\langle A \rangle_b)$ computed.[25] Then, we can calculate s via:

$$s = \lim_{t_b \to \infty} \frac{\sigma^2(\langle A \rangle_b)}{\sigma^2(A)} \tag{2.8}$$

where $\sigma^2(A)$ is the variance of the entire dataset and t_b is the number of data point in each block. From a plot of $(t_b * \sigma^2(\langle A \rangle_b)) / \sigma^2(A)$ versus t_b, we can determine s as the value at which this ratio becomes independent of t_b. We determined the value of s for the NNI-2 system because it had the smallest deviation and thus would yield the largest number for s. We then used the s to compute the standard error $\sigma(A)$ via the following equation:

$$\sigma_{(A)} \approx \sigma \sqrt{\frac{s}{M}} \tag{2.9}$$

where σ is the standard deviation of the system under investigation and M = total number of data points (i.e. total simulation steps/frames).

2.3 Results

2.3.1 NNI-2 and NNI-3 Distinctly Impact NS5B Conformations and Dynamics

In order to understand the impact of ligand binding on conformational sampling, we calculated two-dimensional potentials of mean force (PMFs) using both the inter-domain angle and the template channel width as coordinates (**figure 2.2** and **table 2.2**). In the PMF of the FREE enzyme, we observe a minimum centered at an inter-domain angle of about 67° while fluctuations of the template channel width are between 10.3 – 20.2 Å (**figure 2.2.A**). In contrast, the NNI2 system displays a more compact minimum centered at an angle of about 65.5° and displays a narrower range of template channel widths between 14.6 – 19.8 Å (**figure 2.2.B**). Thus, in comparison to FREE, the presence of the NNI2 ligand restricts conformational sampling of the enzyme. Similar observations were noted in other recent work from our group.[4] We believe that the ability of the NNI2 ligand to predominantly stabilize more closed conformations likely prevents the transitions between closed and open conformations that NS5B must undergo to perform its function, hindering replication. In contrast to NNI2, NNI3 has a minimum that is less compact (**figure 2.2.C**). The enzyme samples larger inter-domain angles and template channel widths compared to the FREE and thus explores conformational states that are in general more open. Thus, both allosteric inhibitors are similar in that they restrict conformational sampling, although this restriction is directed towards different regions of the free energy landscape.

System	Inter-Domain Angle (°)	Inter-Domain Angle Standard Deviation	Template Channel Width (Å)	Template Channel Width Standard Deviation
1QUV	70.045		16.77	
1NB7 (RNA template present)	68.57		16.53	
1YUY	70.33		16.59	
1YV2	71.47		18.6	
FREE *	67.04	1.434	15.37	1.645

NNI2*	65.42	0.9460	16.72	0.8050
NNI3*	69.48	2.040	17.48	1.098
DUAL*	64.79	1.573	15.21	1.090

Table 2.2: Inter-domain angles and template channel widths for PDB structures discussed in the text and average structures from each of the 4 simulations.

It is important to note that a majority of the NNI3 population displays quite large inter-domain angle values, some of which are not observed in the FREE enzyme. This suggests that the NNI3 ligand (3MS) is able to induce novel conformations not observed for the FREE. We believe one reason for this observation is the extensive fluctuations of 3MS (**figure 2.3.C**) and its ability to interact with all three protein domains. The primarily horizontal conformation of 3MS within the central cavity of the enzyme in NNI3 makes it difficult for interactions between thumb and fingers domains to stabilize closed conformations as observed in NNI2 or FREE. As the closed conformation is required for the initiation of RNA synthesis, the ability of 3MS to induce the enzyme to sample more open states would likely impede this stage of replication.

In NNI2, we see reduced correlated motions compared to the FREE system (e.g. see the decreased negative correlations in areas i, ii and iv of **figure 2.4.A**). Area i of NNI2 shown in **figure 2.4.A** represents part of the thumb domain that includes key residues with which the ligand interacts directly such as Ser476, Tyr477, Met423 and Ile482. This area of eliminated motions also includes functional region III (residues 401-414) that is thought to play a role in nascent RNA duplex binding.[26] Flexibility in this area may be required to accommodate double stranded RNA. Thus, eliminating these motions may decrease the ability of NS5B to bind the RNA duplex. Comparing the RMSF computed using the protein residues' centers-of-mass for FREE and NNI2 shows that the flexibility of residues in functional region III is reduced due to the presence of the NNI2 ligand (**figure 2.5.A**). In

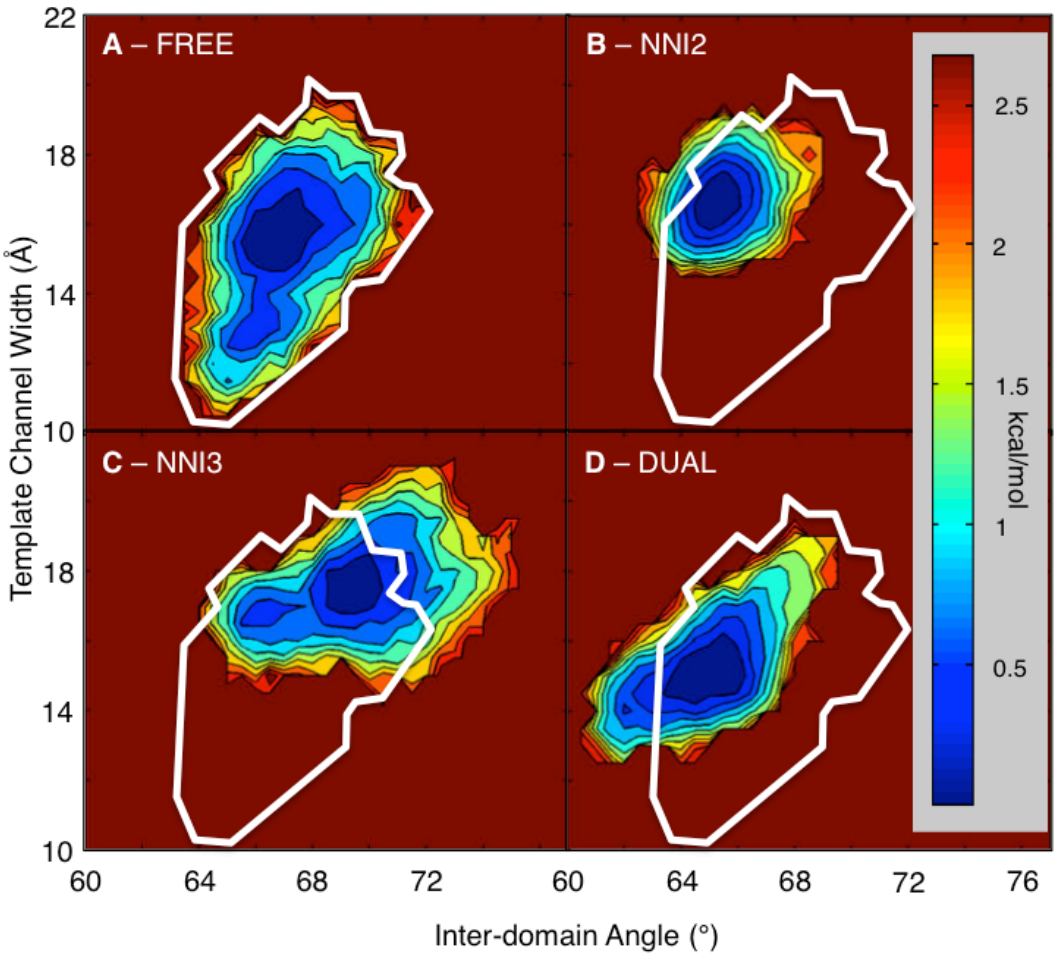


Figure 2.2: Dual inhibitor binding results in novel NS5B conformations. Two-dimensional potential of mean force plots of template channel width versus inter-domain angle for the FREE (A), NNI2 (B), NNI3 (C) and DUAL (D) systems. The color bar represents the relative free energies in units of kcal/mol, where blue indicates more favorable energies while red indicates less favorable energies. The white line demarcates the conformational space sampled in the FREE system to highlight differences between the FREE and ligand bound simulations.

addition to residues within the NNI2 binding pocket, we observed perturbed correlated movements within areas ii and iv of **figure 2.4.A** that span all three enzyme domains. Areas ii and iv constitute motifs B (residues 287 – 306) and E (residues 354 – 372) that function in template and nascent RNA 3'-end binding, respectively.[26] Thus, the NNI2 ligand is able to disrupt correlated motions not only with residues it directly interacts with, but also with residues that are located distal to its binding site. If the correlated motions observed within these key motifs are necessary for interactions between the enzyme and the indicated components of the replication complex to occur, the observed disruption could contribute to inhibiting NS5B function and thus reducing viral replication.

In contrast to NNI2, new correlations are induced in NNI3 when compared to FREE (e.g. the negative correlations in areas iii and v of NNI3 in **figure 2.4.B**). In addition, there are changes in the intensities and patterns of negative correlation in areas i, ii and iv. In the crystal structure containing the NNI3 inhibitor bound to NS5B (PDB ID: 3CO9), the inhibitor forms hydrogen bonds to Ser288, Tyr448 and Gln449 and interacts with other residues such as Phe193, Arg200 and Leu384 located within regions iii and v of **figure 2.4.B.** These interactions are recapitulated in our simulations. However**,** these regions of the enzyme also display new negative correlations compared to the FREE, suggesting that the NNI3 inhibitor induces new protein motions within its binding pocket. Changes in the patterns of correlated movement observed in areas i, ii and iv (the latter two constituting motifs B and E respectively) indicate that the NNI3 inhibitor can also modulate protein correlations distal to its binding pocket in a manner that would tend to reduce enzyme activity. RMSF data for NNI3 displays an increase in the flexibility of residues in motifs E and F that exhibit new correlations (**figure 2.4.B**). As mentioned previously, motif E plays a role in binding the 3' end of nascent RNA, while motif F plays a role in nucleotide and template binding. Unlike NNI2, the presence of the inhibitor in NNI3 seems to increase fluctuations of residues within these motifs. Such fluctuations may work to destabilize these areas of the enzyme to the extent that they weaken interactions between NS5B and RNA template or nucleotides, thus inhibiting the enzyme. Thus, the NNI2 and NNI3 inhibitors may have contrasting effects in modulating the flexibility of these key regions yet both reduce enzyme activity.

2.3.2 Dual Inhibitor Binding Induces Novel Protein Conformations and Dynamics

We find that even though VGI (NNI2) and 3MS (NNI3) have distinct impacts on certain structural features of the enzyme, they are able to bind simultaneously and jointly impact both structure and dynamics. To date, only biochemical data has been available to

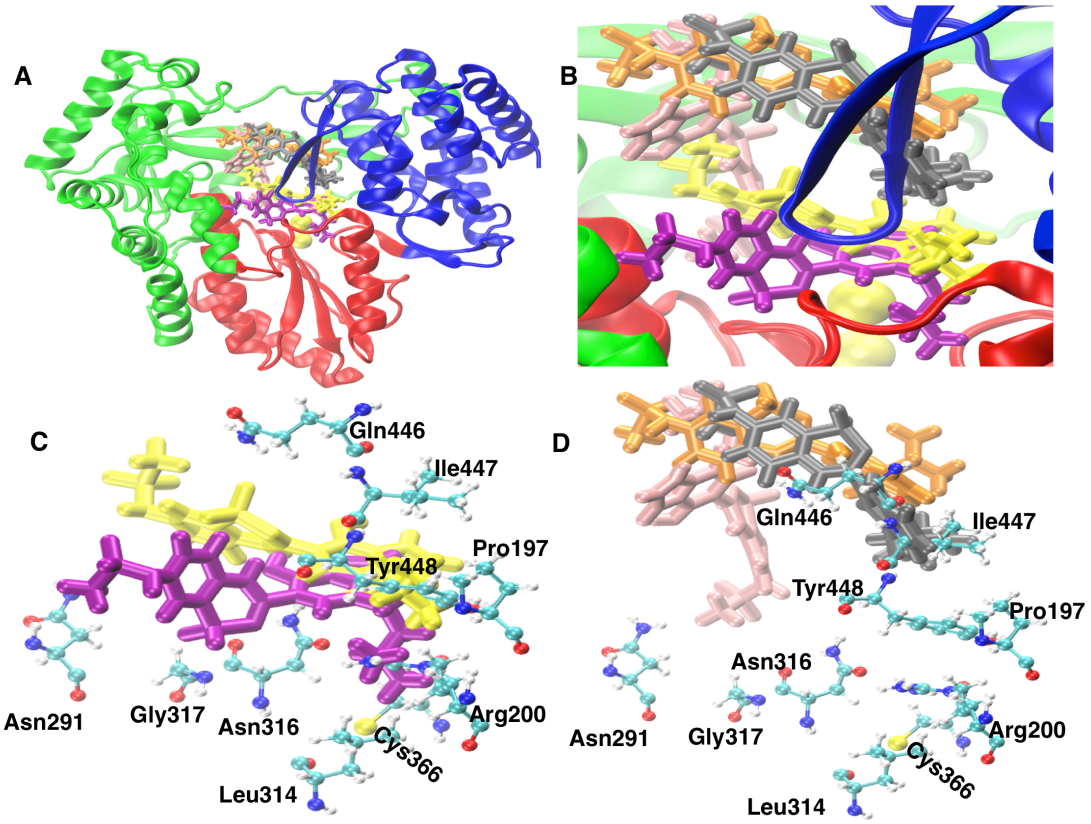


Figure 2.3: Conformations and locations of 3MS in the NNI3 and DUAL systems suggest a single expansive palm-binding cavity. Panel A depicts snapshots of both the NNI3 and DUAL systems relative to the three protein domains, while panel B is a zoomed image of the black encircled region in A. Panel C and D show snapshots of 3MS (licorice representation) from NNI3 (C) and DUAL (D) simulations along with critical protein residues (CPK representation) known to bind different palm allosteric moieties. These residues represent key interactions for specific palm allosteric inhibitors. In panel D, 3MS is shown interacting with palm residues known to be involved in binding to other palm inhibitors. In panel C, the yellow- and purple-colored snapshots of 3MS are taken from the NNI3 trajectory and represent different orientations that facilitate protein-ligand interactions specific to the two known palm allosteric sites. In panels B and D the pink, orange and black snapshots of 3MS highlight different orientations of the ligand within the DUAL trajectory and how these allow for varying protein-ligand interactions.

suggest that the combined presence of thumb inhibitors such as VGI and palm ligands like 3MS results in enhanced reductions of NS5B activity.[9-11] Liu et. al. performed an inhibitor/inhibitor competition assay and observed an additive inhibitory impact on NS5B when simultaneously treated with a known NNI-3 (palm allosteric) inhibitor and a compound later confirmed to be an NNI-1, **figure 2.4A**.[10] Similarly, Le Pogam and coworkers observed a 20-fold synergistic reduction in the colonies of HCV replicons on agarose plates when incubated simultaneously with NNI-3 and NNI-2 allosteric inhibitors (**figure 2.4B**).[9] Although the work by Liu et al. and Le Pogam et al. provided evidence for enhanced

inhibition of NS5B in the presence of palm and thumb allosteric inhibitors, neither provided molecular details such as the changes to protein conformational space and dynamics mediating the observed enhanced inhibitory effects. Thus, we provide the first molecular evidence that inhibitors at the thumb NNI-2 and palm NNI-3 sites are able to bind to the enzyme simultaneously and also delineate molecular mechanisms through which enhanced

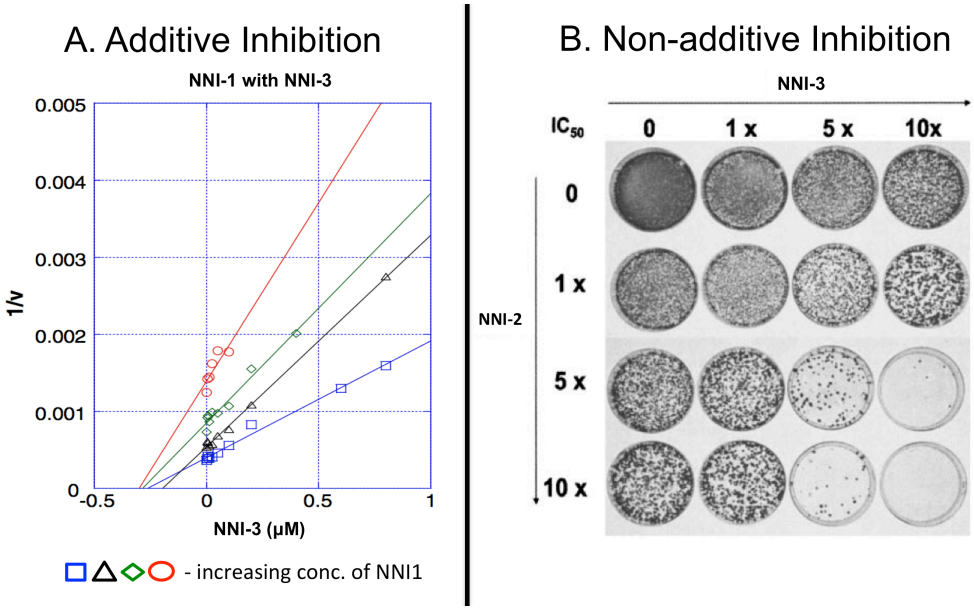


Figure 2.2: Prior evidence that supports enhanced inhibition in the HCV polymerase. Panel A depicts data from an inhibitor/inhibitor competition assay performed by Liu et al highlighting the increased negative impact on the rate of nucleotide incorporation by the polymerase when both NNI-3 and NNI-1 inhibitors are present. [10] The plot shows inverse enzyme rate versus concentration for an NNI-3 (thiadiazine) with each series representing an increased concentration of an NNI-1 (halosalicylamide) going from blue to red. Work by Le Pogam and coworkers in panel B demonstrate that when NNI-3 and NNI-2 inhibitors are present there is a 20-fold reduction in the number of colonies of HCV replicons observed on agarose plates relative to when individual inhibitors are applied. [9] The horizontal arrow indicates increasing concentrations of the NNI-3 inhibitor while the vertical arrow represents increasing concentrations of the NNI-2. The entries along each axis indicate how much larger the concentration at that position is relative to the IC_{50} of the respective inhibitors (i.e. 5 X indicates the inhibitor concentration is five times larger than the IC_{50}).

inhibition of NS5B via the action of dual allosteric inhibitors may occur. When both inhibitors are bound we observe novel fluctuations in addition to patterns of protein motions that exhibit characteristics of the individual singly bound systems. There are patterns of correlation in areas i and ii of the correlation map for the DUAL system that also occur in the other systems, while there are patterns of correlation in area iv of the DUAL system that are more extensive and more intense compared to the other 3 systems (**figure 2.5.C**). The changes within area iv span two motifs: B, which has functional role in primer and metal binding and catalysis as we mentioned previously, and C that is important for triphosphate and nucleotide triphosphate binding and catalysis.[26] Novel fluctuations are also apparent in the significant increase in the flexibility of motif F in DUAL when compared to the NNI3 and FREE systems (**figures 2.6.B** and 2.**6.C**). Motif F has functional roles in nucleotide and template binding.[26] Increased fluctuations in these motifs may drastically reduce or prevent

interactions between the enzyme and template and/or nucleotides. Thus, these new protein motions in the DUAL system may impair viral replication by disrupting catalysis at the active site. This finding may represent one mechanism by which enhanced inhibition of the viral enzyme is mediated in the presence of dual NNIs.

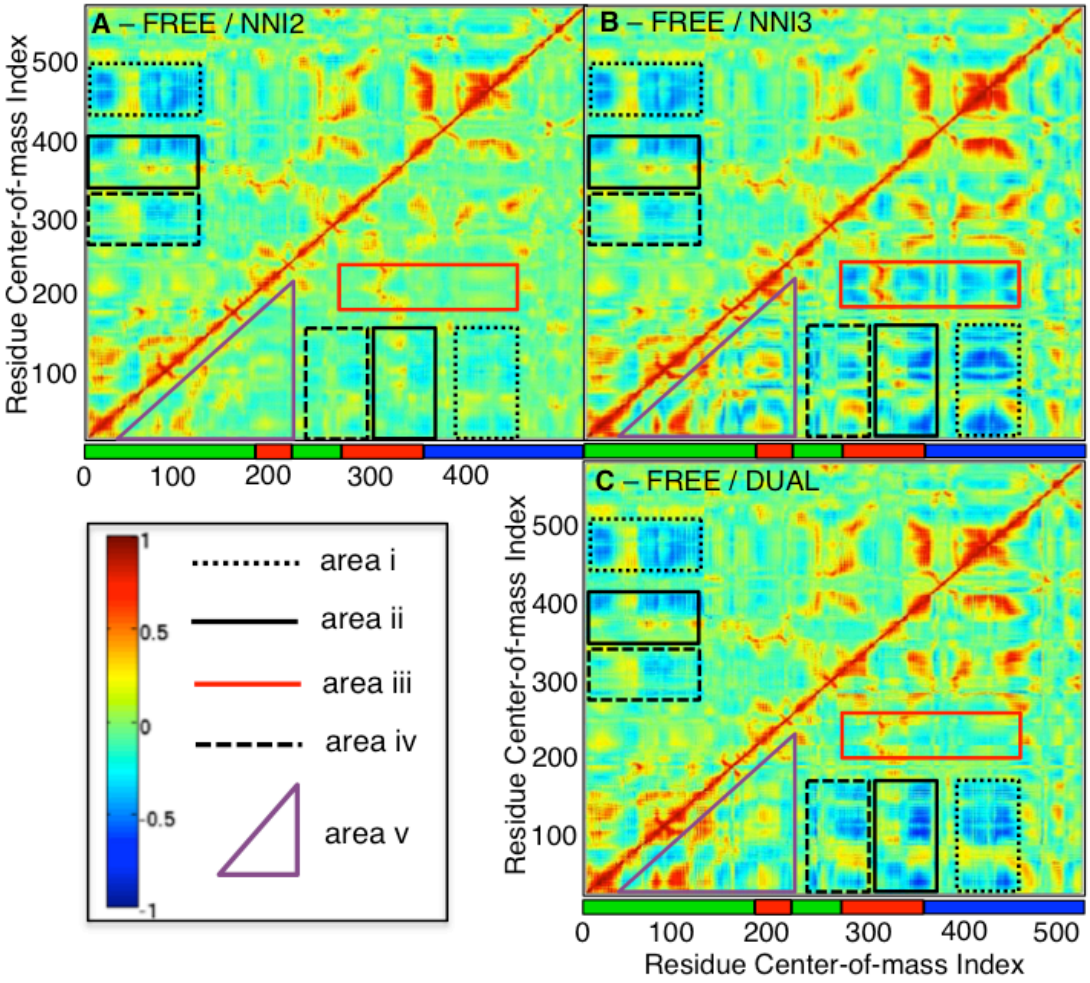


Figure 2.5: Correlation maps showing the impact of inhibitor binding on enzyme dynamics. In each map, the diagonal separates the FREE from the ligand-bound system, with the FREE being above the diagonal. A: FREE versus NNI2, B: FREE versus NNI3 and C: FREE versus DUAL. The outlined areas consist of residues that make up the ligand binding sites or functional motifs that experience changes in protein motions.

In the correlation map of the DUAL system, we do not observe the negative correlations in areas iii and v that were distinctive to NNI3 (**figure 2.5.C**). The absence of these correlations may result from the impact of the VGI ligand in reducing overall enzyme dynamics, a phenomenon noted in our previous studies.[4] We also note reduced intensities of correlations in areas i and ii for the DUAL system compared to NNI3.These areas in the DUAL system are more similar to the corresponding areas in the NNI2 system where there are also reduced correlations relative to the FREE enzyme. Thus, it is likely that the presence

of VGI is the reason some of the characteristic correlation patterns observed in NNI3 are reduced in the DUAL system.

Two-dimensional PMF plots also revealed novel conformations when both allosteric ligands are bound compared to the other three systems (**figure 2.2.C**). Specifically, DUAL exhibits conformations with the highest degree of closure among all four systems. Some of these conformations are more closed (e.g. inter-domain angle < 63°) than the "hyper-closed" conformations previously identified by our group for this enzyme (inter-domain angle ≈ 63°).[4] With the 3MS ligand participating in many more interactions within the fingers and thumb domains, the ligand spends a considerable amount of time in a vertical position at the apex of the central cavity of the enzyme in the DUAL system. In this position, 3MS is less effective in physically hindering the fingers and thumb domains from coming closer together than in NNI3. However, it would be more efficient in directly blocking access of the RNA template to the enzyme active site (**figures 2.7.B** and **2.7.C**). Furthermore, for both NNI2 and DUAL where the conformations are more closed overall than for NNI3, we observe a hydrogen bond between ARG109 and ASP444 that is not seen in NNI3. It is interesting to note that both these residues are located at the top of enzyme: ARG109 within the fingers and ASP444 within the thumb. This hydrogen bond would further strengthen interactions that may be important to allosteric communication between the two domains. Finally, the DUAL system also has a hydrogen bond between HIS95 (fingers domain) and ASN406 (thumb domain) that is not present in the NNI2 system, providing an additional interaction that stabilizes DUAL in a more closed state compared to NNI2. Both of these residues are either part of or are adjacent to residues making up the RNA duplex channel and thus may be important in the elongation phase of RNA replication.

The regions of the binding site explored by 3MS in the NNI3 system coincide with protein residues known to bind various moieties of different palm allosteric inhibitors (**figure 2.3.C**). The yellow snapshot represents a conformation that is very similar to that observed in

the co-crystal structure of NS5B with 3MS (PDB ID: 3CO9) and is representative of most of the conformations adopted throughout the NNI3 trajectory. In contrast, the purple snapshot shows 3MS in the DUAL system occupying regions that are known to be part of the NNI-4 site (**figure 2.1**) and that typically accommodate different palm allosteric inhibitors such as those that contain the benzofuran moiety. These observations, in concert with those discussed above relating to NNI3, suggest that one can consider the two palm sites to be a single large pocket that can accommodate distinct ligands within various regions of the cavity. Thus, the specific location adopted by a given ligand in this extended binding region may depend on its chemical complementarity to different areas of the pocket.

We observe a linear relationship between inter-domain angle and template channel width in DUAL that is less pronounced or absent in the other systems (**figure 2.2**). The correlation between these two metrics is readily observed in the 1% lowest frequency PCA modes of DUAL (see movies in supplemental information). These fluctuations entail a highly concerted twisting movement involving all three domains. During this movement the palm domain is a focal point through which the twisting motion is transferred into the fingers and thumb domains, facilitating their anticorrelated movements relative to one another. In addition, the thumb and fingers domains of the DUAL system move along different axes. This is in contrast to what is observed in the FREE, where both the thumb and fingers domains have fluctuations that are along the same axis. Furthermore, the fluctuations in the DUAL system that allow the fingers and thumb domains to come into close proximity also decrease the template channel width, which may make it much more difficult for RNA template to access the channel and subsequently, the active site. Such structural changes may disfavor the elongation phase of RNA replication.

The appearance of novel protein conformations in the DUAL system is an emergent property that results from the simultaneous presence of both inhibitors and suggests a non-additive impact on the NS5B conformational ensemble. These non-additive effects are

consistent with available biochemical evidence that demonstrates synergistic inhibition of NS5B in the combined presence of palm and thumb allosteric inhibitors.[9-11]

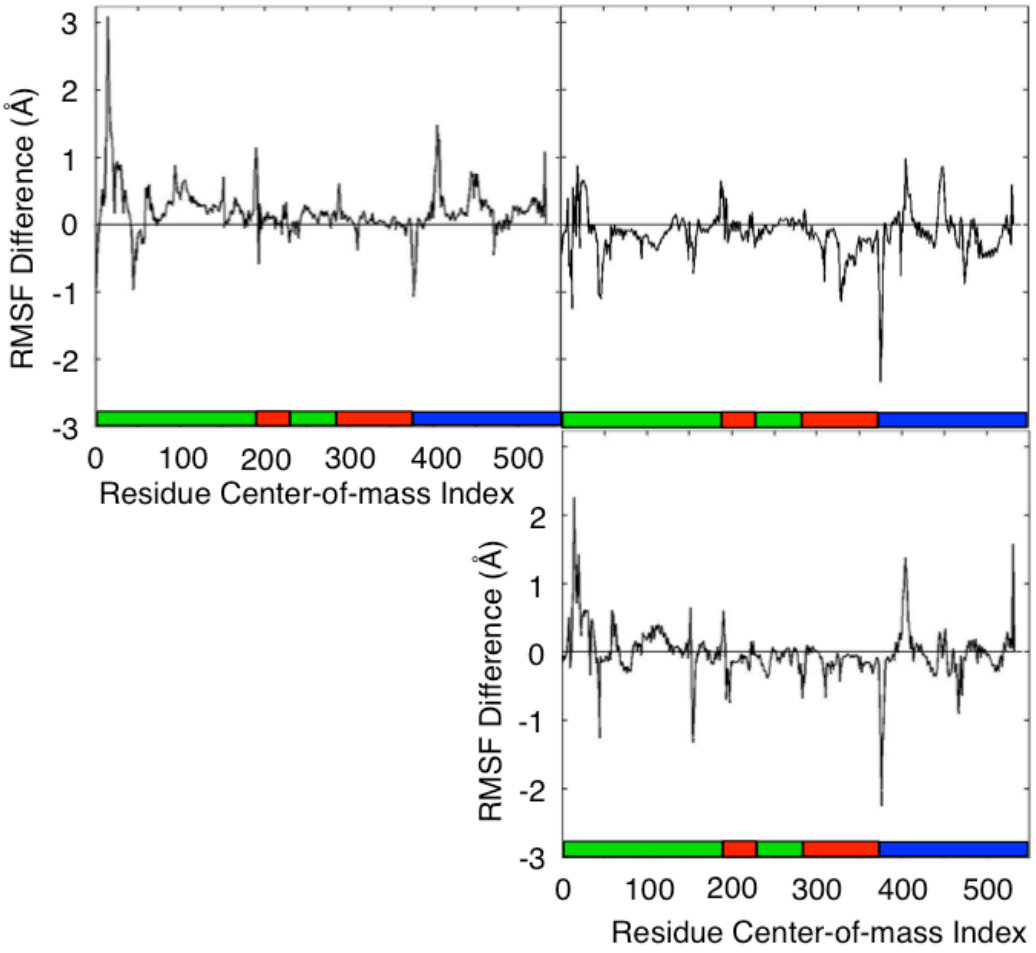


Figure 2.6: RMSF difference plots showing changes to the local enzyme flexibility in the presence of allosteric inhibitors. Each plot represents the difference between free and inhibitor-bound fluctuations. Values greater than zero indicate increased flexibility in the FREE compared to the ligand-bound system of interest while values less than zero denote increased flexibility for the respective ligand bound system NNI2 (A), NNI3 (B) or DUAL (C). The horizontal color bar demarcates the three enzyme domains: palm (red), thumb (blue) and fingers (green). The position of important functional and structural motifs is indicated as shown in the legend in the lower left of the figure.

2.3.3 Novel Conformations in the DUAL System are due to the Altered Interactions of the NNI3 Ligand

As mentioned above, conformations with the highest degree of closure were observed in the DUAL system (**figure 2.2**). The primary determinant of this observation is the flexibility of 3MS, which allows modifications to the hydrogen-bonding network between protein and ligand in the DUAL system. The orientation of 3MS ranges from horizontal to

vertical over the course of the DUAL trajectory. The more vertical orientations correspond to a change in hydrogen bonding that allows 3MS to primarily interact with residues in both fingers and thumb domains, stabilizing the enzyme in an extremely closed conformation. Specifically, the ligand makes direct hydrogen bonds with His95, Ala97 and Gln446 and water-mediated hydrogen bonds with residues such as Gly283 in DUAL (**figure 2.6**). In contrast, it makes direct hydrogen bonds with residues such as Thr292, Tyr191 and Tyr448 and water-mediated hydrogen bonds with residues such as Gly449 and Ser288 in NNI3. The majority of protein-ligand hydrogen bonds observed in NNI3 correspond to those identified crystallographically. These hydrogen bonds are weakened or abolished (e.g. those with residues Tyr448, Tyr191 and Thr292) in the DUAL system. In DUAL we observe new and more stable hydrogen bonds involving residues such as Gln446. In fact, the hydrogen bond between Gln446 and the hydroxyl hydrogen of 3MS has the highest occupancy (84%) of all hydrogen bonds involving protein residues in the DUAL system. This interaction is even more stable than that of Tyr448 (occupancy of 64%), which has the most stable protein-ligand hydrogen bond in NNI3.

3MS adopts numerous conformations and is very mobile in DUAL compared to the NNI3 system (**figures 2.3.C and 2.3.D**). Our studies indicate that these properties are vital to the novel structure and dynamics observed in DUAL. The ability of 3MS to adopt diverse conformations in the DUAL system coincides with the ligand occupying regions of the enzyme that differ from the binding location observed in its original crystal structure (PDB ID: 3CO9). For example, in DUAL, 3MS is able to interact with protein residues such as Gln446 and Ile447 with which it did not in NNI3 (**figure 2.3.D**). These protein residues have not previously been shown to associate with 3MS. However, they have been shown to interact with other palm allosteric inhibitors with chemical moieties such as benzamide and proline sulfonamide (PDB IDs: 3LKH and 2GC8). In addition, the conformations that 3MS adopts in DUAL while interacting with these protein residues (**figure 2.3.D**) are very

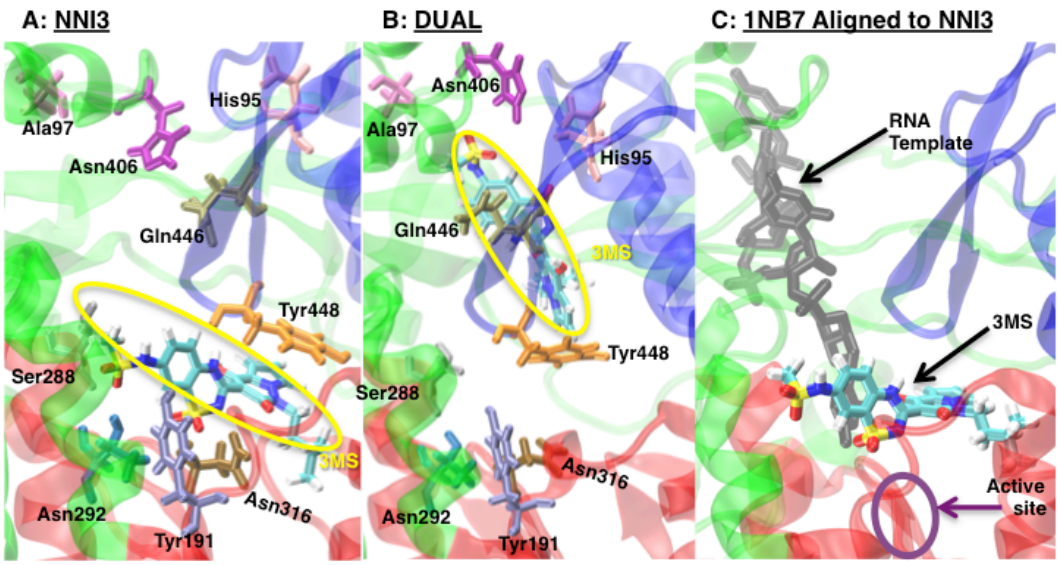


Figure 2.7: Comparison of key hydrogen bonds involving the 3MS ligand in the NNI3 (A) and DUAL (B) systems. We chose two snapshots that are both in a closed conformation (inter-domain angle < 68°), which emphasizes the local differences around the inhibitor even though the global conformations are similar. For each panel, the protein is translucent to allow for better visibility of the important residues in the hydrogen bond network. Panel C: Alignment of NNI3 snapshot 4000 and PDB ID 1NB7, a structure of NS5B containing short RNA template, to show how the presence of 3MS prevents template access to the active site.

different from those observed in NNI3 or for other previously crystallized palm inhibitors. There are instances when 3MS adopts a vertical orientation in the DUAL simulations that allows it to occupy the RNA template channel (**figure 2.3.C** and pink-colored ligand in **figure 2.3.D**). Thus, when dual inhibitors are bound, NNI-3 inhibitors may be particularly effective in preventing RNA template from accessing the active site. In addition, the high flexibility and mobility of 3MS observed in the DUAL system sheds light on the possible difficulty in crystallizing two inhibitors bound to the enzyme, particularly those employed in the current work. This may help explain the absence of a NS5B crystal structures containing these two types of inhibitors bound simultaneously.

Nonetheless, our results clearly indicate that the presence of both inhibitors impacts the interactions each ligand makes with the protein and that there are differences in protein-ligand interactions in the singly bound systems versus DUAL. Above we discussed differences in the hydrogen bonds 3MS makes with the protein in the NNI3 versus DUAL. Specifically, in very closed protein conformations 3MS is able to stabilize interactions

between the thumb and fingers domains in the DUAL system, potentially facilitating allosteric communications that can result in inhibition. We also observe changes to the hydrogen bond network of VGI in NNI2 compared to DUAL. We identified water-mediated hydrogen bonds between VGI and Tyr477 and Leu497 in DUAL that were not observed in the NNI2 system. Both of these residues are part of the VGI binding site. Furthermore, the interaction between VGI and Tyr477 was determined to be crucial for ligand potency.[7] Thus, we can conclude that when both VGI and 3MS are present they facilitate stronger interactions that can disrupt allosteric communication and result in enhanced ligand binding. This observation is further supported by the computed ligand binding energies that are discussed in the next section.

2.3.4 Binding of Dual Ligands is Energetically Favorable

It is useful to employ the simulation results to understand how the action of these inhibitors relates to their binding efficacies. If enhanced binding occurs in the presence of dual inhibitors, one might expect to observe a relative free energy of binding for the DUAL system that is more favorable than the sum of individual binding free energies for the NNI2 and NNI3 systems. Our calculations indicate that DUAL displays the most favorable computed binding free energy, NNI3 the least favorable and NNI2 falls in between these two extremes (**table 2.3**). The relative binding free energy computed for the DUAL system ($\Delta\Delta G_{DUAL}$) is more favorable than the sum of binding free energies for the individual systems $\Delta\Delta G_{NNI3+NNI2}$ (see the last line of **table** 2.**3**), consistent with synergistic binding of both inhibitors. However, given the uncertainty (see standard error values in table 2.3, which were based on a statistical inefficiency of 500 or 50ns in figure 2.8) in the computed values, it is also possible that the inhibitors bind in an additive rather than synergistic manner.

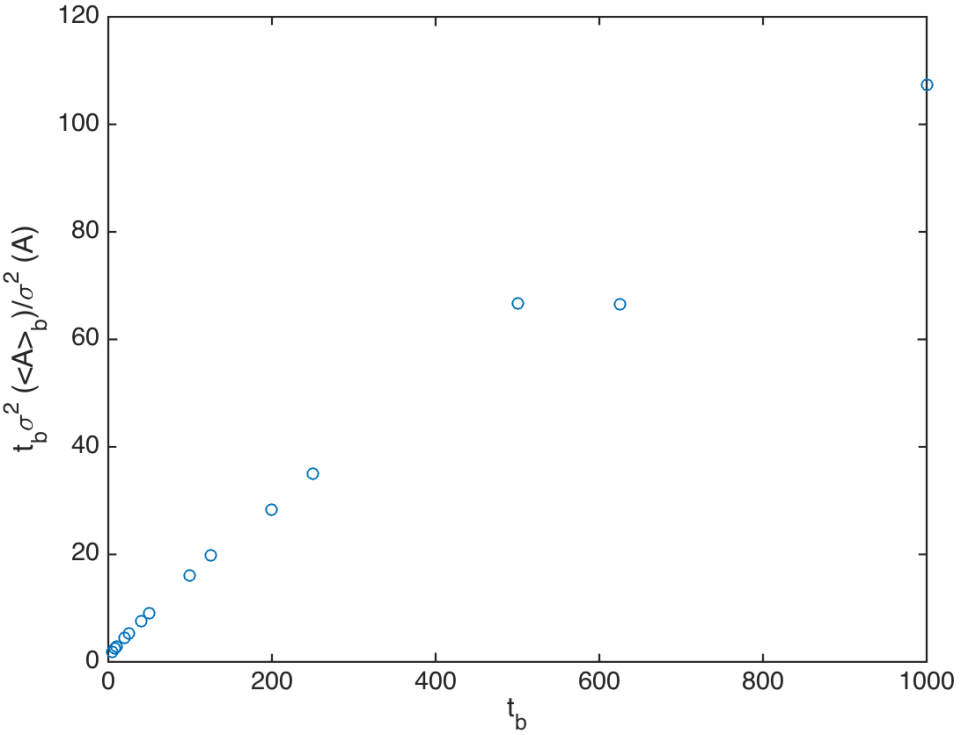


Figure 2.8: Calculating the Statistical Inefficiency of Various Block Size using the Computed ΔEnthalpies of the NNI-2 system. Determination of the statistical inefficiency is required to calculate the standard error. In this plot, the Δenthalpy (i.e. A) data of the last 5000 snapshots of the NNI-2 trajectory were divided into blocks of varying lengths. $\sigma^2(<A>_b)$ is the variance of the mean of the blocks, $\sigma^2(A)$ is the variance of all the Δenthalpy data and t_b equals the number of data per block.

System	ΔH (kcal/mol)	ΔH Standard Error (kcal/mol)	ΔS (kcal/mol*K)	TΔS (kcal/mol)	ΔG (kcal/mol)
NNI3	-19.1	5.1	-0.00677	-2.0	-17.0
NNI2	-23.4	1.2	-0.0150	-4.5	-18.9
DUAL	-47.5	2.6	-0.0211	-6.3	-41.2

$\Delta\Delta G_{NNI3+NNI2}$ = -35.9 ± 6.3 kcal/mol

Table 2.3: The binding free energies of the ligand bound systems along with the enthalpic and entropic contributions corresponding to each.

However, the novel patterns of correlation and unique enzyme conformations that only arise when both ligands are present in the DUAL system (**figure 2.2 and 2.3**) clearly indicate that the two ligands jointly impact the enzyme free energy landscape and dynamics. One might anticipate for this observation to be associated with synergistic binding of these

ligands. Moreover, experimental studies have shown enhanced inhibition when palm and thumb allosteric inhibitors are employed in combination (NNI-4 and NNI-2 or NNI-3 and NNI-1),[9-11] a situation that may be indicative of synergistic binding. Thus, despite the estimated uncertainties for the computed free energy values (**table 2.3**), it is likely that both ligands do bind in a slightly synergistic manner.

Regardless of whether these ligands bind in an additive or synergistic manner, our studies indicate that the presence of dual inhibitors does not lower their binding efficacy. Consequently, it is reasonable to anticipate that the net inhibition will be enhanced in the presence of both ligands as suggested by the experimental evidence. Our studies can thus illuminate the molecular mechanisms underlying these observations. We note that the computed affinity of both inhibitors in the DUAL system is more favorable than the affinity of either of the individual NNIs (which are both known to bind to the enzyme). This suggests that both inhibitors should be able to simultaneously bind to the enzyme quite well. It could even be possible that binding occurs in an additive manner, but that once binding takes place the free energy landscape is synergistically modified so that the dual inhibitors are more effective in modulating the physical properties of the enzyme than either on its own.

2.3.5 Broader Significance

We find that the inhibitors studied in this work induce an allosteric response by modulating the ensemble of protein conformations. Recent work shows a link between local interactions and global conformational changes important to allostery.[27, 28] Similarly, our work reveals that local differences such as altered hydrogen bonding patterns are associated with distinct global protein properties. Numerous studies have discussed the capacity of allosteric proteins to respond to a ligand in either an agonistic or antagonistic manner based on the presence of a secondary allosteric effector.[29-31] Our present work provides molecular evidence describing how an allosteric protein such as NS5B can differentially respond to two

allosteric inhibitors bound at non-overlapping sites. Specifically, our results suggest a model of enhanced allosteric inhibition for the HCV polymerase that may include additive binding, but also incorporates synergistic changes to protein conformations and fluctuations once both ligands are bound.

As mentioned in the introductory paragraph, one way to combat the viral resistance that arises due to mutations in NS5B is to employ multiple polymerase inhibitors in combination to treat HCV infection. Because allosteric inhibitors do not compete with nucleotides for binding to NS5B, they may be able to reduce the need for widely used nucleoside analogs (such as ribavirin and Solvaldi) that can be associated with non-specific cellular toxicities. Importantly, our studies suggest that the molecular mechanisms mediating enhanced inhibition for allosteric inhibitors such as those studied in this work may include non-additive structural and dynamic changes to NS5B.

2.4 Conclusion

We have shown that allosteric inhibitors binding to non-overlapping locations, specifically sites NNI-2 and NNI-3, can jointly modulate the conformations and dynamics of the HCV NS5B polymerase. This is despite the fact that each ligand on its own induces distinct regions of the free energy landscape to be sampled and elicits unique patterns of enzyme motions. We find that the NNI2 inhibitor used in this study elicits its allosteric effect via a conformational selection mechanism in which it stabilizes more closed protein conformations and reduces flexibility compared to the free enzyme. In comparison, the NNI3 ligand not only shifts the population of enzyme conformation sampled in the free enzyme to be more open on average, it also induces novel conformations and correlated motions. The mechanism of action for the NNI3 ligand thus incorporates both conformational selection and induced fit properties. Despite these individual differences, when both NNIs are bound we observe new protein conformations as well as novel dynamics that were not previously

observed in either the free enzyme or the individual enzyme-inhibitor complexes. Although computed binding free energies do not exclude the possibility that the inhibitors bind in an additive manner, they do appear to act synergistically with regard to their ability to modulate enzyme conformations and dynamics. One might anticipate that even if simultaneous binding of both ligands is not more highly favored than the sum of their individual affinities, once both are bound the enzyme may nonetheless undergo dynamics and conformational fluctuations that are more effective in reducing its activity.

These studies provide the first molecularly detailed description of the mechanisms underlying enhanced inhibition of the HCV polymerase in the presence of dual allosteric inhibitors. This knowledge will facilitate efforts to optimize combinations of inhibitors targeting NS5B. In this way it may be possible to circumvent the emergence of viral resistance that is a persistent obstacle to the development of HCV therapeutics.

2.5 Acknowledgements

We employed the University of Maryland, Baltimore County (UMBC) high performance computing facility (HPCF) in this study. This investigation also used the Ranger and Stampede HPC clusters of Extreme Science and Engineering Discovery Environment (XSEDE). We wish to thank Dr. Alexander Sodt for his technical assistance with data analysis. In addition, we acknowledge technical assistance from Meagan Small, Dr. Kenno Vanommeslaeghe and other members of the laboratory of Dr. Alexander MacKerell. Finally, we thank Ester Sesmero and Brittny Davis and other members of the Thorpe laboratory for numerous and valuable discussions. HPCF is funded by the U.S. National Science Foundation through the MRI program (grant numbers: CNS-0821258 and CNS-1228778) and the SCREMS program (grant number: DMS-0821311), with significant support from UMBC. XSEDE is supported by the National Science Foundation (grant number: OCI-1053575). Jodian Brown was funded by NIH F31 predoctoral fellowship (GM106958).

2.6 References

[1] Adachi, T., Ago, H., Habuka, N., Okuda, K., Komatsu, M., Ikeda, S., and Yatsunami, K. (2002) The essential role of C-terminal residues in regulating the activity of hepatitis C virus RNA-dependent RNA polymerase, *Biochimica et Biophysica - Proteins and Proteomics 160*, 38 - 48.

[2] Davis, B. C. T., I. F. (2013) Molecular Simulations Illuminate the Role of Regulatory Components of the RNA Polymerase from the Hepatitis C Virus in Influencing Protein Structure and Dynamics, *Biochemistry 52*, 4541 - 4552.

[3] Harrus, D. A.-E.-S., N.; Simister, P. C.; Miller, S.; Triconnet, M.; Hagedorn, C. H.; Mahlas, K.; Rey, F.; Astier-Gin, T.; Bressanelli, S. (2010) Further insights into the Roles of GTP and the C Terminus of the Hepatitis C Virus Polymerase in the Initiation of the RNA Synthesis, *Journal of Biological Chemistry 285*, 32906-32918.

[4] Davis, B. C. T., I. F. (2013) Thumb inhibitor binding eliminates functionally important dynamics in the hepatitis C virus RNA polymerase, *Porteins: Structure, Function, Bioinformatics 2013*, 40 - 52.

[5] Benzaghou, L. B., I.; Bisaillon, M. (2004) Effect of Metal Ion Binding on the Structural Stability of the Hepatitis C Virus RNA Polymerase, *Journal of Biological Chemistry 279*, 49755-49761.

[6] Bougie, I. C., S.; Bisaillon, M. (2003) Characterization of the Metal Ion Binding Properties of the Hepatitis C Virus RNA Polymerase, *Journal of Biological Chemistry 278*, 3868-3875.

[7] Ontoria, J. M. (2009) Identification and Biological Evaluation of a Series of 1H-Benzo[de]isoquinoline-1,3(2H)-diones as Hepatitis C Virus NS5B Polymerase Inhibitors, *Journal of Medicinal Chemistry 52*, 5217-5227.

[8] Ruebsam, F. (2008) Pyrrolo[1,2-b]pyridazin-2-ones as potent inhibitors of HCV NS5B polymerase, *Bioorganic & Medicinal Chemistry Letters 18*, 3616-3621.

[9] Le Pogam, S. (2006) Selection and Characterization of Replicon Variants Dually Resistant to Thumb- and Palm-Binding Nonnucleoside Polymerase Inhibitors of the Hepatitis C Virus, *Journal of Virology 80*, 6146-6154.

[10] Liu, Y., Donner, P. L., Pratt, J. K., Jiang, W. W., Ng, T., Gracias, V., Baumeister, S., Wiedeman, P. E., Traphagen, L., Warrior, U., Maring, C., Kati, W. M., Djuric, S. W., and Molla, A. (2008) Identification of halosalicylamide derivatives as a novel class of allosteric inhibitors of HCV NS5B polymerase, *Bioorg Med Chem Lett 18*, 3173-3177.

[11] Cheng, C. C. H., Xiaohua ; Shipps Jr., Gerald W.; Wang, Yu-Sen; Wyss, Daniel F.; Soucy, Kyle A.; Jiang, Chuan-kui; Agrawal, Sony; Ferrari, Eric; He, Zhiqing; Huang, H.-C. . (2010) Pyridine Carboxamides: Potent Palm Site Inhibitors of HCV NS5B Polymerase, *ACS Medicinal Chemistry Letters 1*, 466 - 471.

[12] Vanommeslaeghe, K. H., E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields, *Journal of Computational Chemistry 31*, 671-690.

[13] MacKerell, A. D., Brooks, B., Brooks III, C. L., Nilsson, L., Roux, B., Won, Y., and Karplus, M. (1998) CHARMM: The Energy Function and its Parameterization, In *Encyclopedia of Computational Chemistry*, John Wiley and Sons, Ltd.

[14] Vanommeslaeghe, K. H., E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell Jr., A. D. (2010) CHARMM General Force Field: aA Force Field for Drug-Like Molecules Compatible with the CHARMM All-atom Additive Biological Force Fields, *Journal of Computational Chemistry 31*, 671 - 690.

[15] MacKerell, A. D., Bashford, D., Bellot, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher III, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stole, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins, *Journal of Physical Chemistry B 102*, 3586-3616.

[16] Mackerell, A. D., Feig, M., and Brooks, C. L. (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations, *Journal of Computational Chemistry 25*, 1400-1415.

[17] Mackerell, A. D., Feig, M., and Brooks, C. L. (2004) Improved Treatment of the Protein Backbone in Empirical Force Fields, *Journal of the American Chemical Society 126*, 698-699.

[18] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD, *Journal of Computational Chemistry 26*, 1781-1802.

[19] Biswal, B. K. C., M. M.; Wang, M.; Chan, L.; Yannapoulos, C. G.; Bilimoria, D.; Nicolas, O.; Bedard, J.; James, M. N. G. (2005) Crystal Structures of the RNA-dependent RNA Polymerase Genotype 2a of Hepatitis C Virus Reveal Two Conformations and Suggest Mechanisms of Inhibition by Non-nucleosdie Inhibitors, *Journal of Biological Chemistry 280*, 18202-18210.

[20] Chinnaswamy, S. M., A.; Yi, G.; Palaninathan, S.; Kao, C. C. (2010) Conformations of the monomeric hepatitis C virus RNA-dependent RNA polymerase, *Virus Adaptation and Treatment 2*, 21-39.

[21] Lee, M. S., Feig, M., Salsbury Jr., F. R., and Charles III, C. L. (2003) New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations, *Journal of Computational Chemistry 24*, 1348-1356.

[22] Feig, M., Onufriev, A., Lee, M. S., Im, W., Case, D. A., and Brooks III, C. L. (2004) Performance comparison of generalized Born and Poisson methods in the calculation of electrostatic salvation energies for protein structures, *Journal of Computational Chemistry 25*, 265-284.

[23] Chocholousova, J., and Feig, M. (2006) Balancing an accurate representation of the molecular surface in generalized Born formalisms with integrator stability

in molecular dynamics simulations, *Journal of Computational Chemistry 27*, 719-729.
[24] Tzoupis, H. L., G.; Mavromoustakos, T.; Papadopoulos, M. G. (2013) A Comparative Molecular Dynamics, MM-PBSA and Thermodynamic Integration Study of Saquinavir Complexes with Wild-Type HIV-1 PR and L10I, G48V, L63P, A71V, G73S, V82A and I84V Single Mutants, *Journal of Chemical Theory and Computation 9*, 1754-1764.
[25] Leach, A. R. (2001) *Molecular Modeling: Principles and Applications*, 2nd ed., Pearson Education Limited, Essex, England.
[26] Moustafa, I. M. S., H.; Morton, B.; Colina, C. M.; Cameron, C. E. (2011) Molecular Dynamics Simulations of Viral RNA Polymerases Link Conserved and Correlated Motions of Functional Elements to Fidelity, *Journal of Molecular Biology 410*, 159-181.
[27] Motlagh, H. N. H., V. J. (2012) Agonism/antagonism switching in allosteric ensembles, *Proceedings of the National Academy of Sciences 109*, 4134-4139.
[28] Cui, Q. K., M. (2008) Allostery and cooperativity revisited, *Protein Science 17*, 1295-1307.
[29] Hol, T. C., M. B.; Bryant, H. U.; Draper, M. W. (1997) Selective estrogen receptor modulators and postmenopausal women's health, *Journal of Womens Health 6*, 523-531.
[30] Katzenellenbogen, B. S. M., M. M.; Ekena, K.; Herman, M. E.; McInerney, M. E. (1997) Antiestrogens: Mechanisms of action and resistance in breast cancer, *Breast Cancer Research and Treatment 44*, 23-38.
[31] Lu, N. Z. C., J. A. (2005) Translational regulatory mechanisms generate N-terminal glucocorticoid receptor isoforms with unique transcriptional target genes, *Molecular Cell 18*, 331-342.

Chapter 3: Structural Similarity in Viral Polymerases may lead to Comparable Allosteric Binding Sites

This Chapter is reproduced in part from: ***Brown, J.B****., Espiritu, M.V., Abraham, J. and Thorpe, I.F., Virus Research, Submitted.*

3.1 Introduction

The goal of this dissertation is to determine if the structural similarities that exist between the HCV polymerase and the other viral polymerases from DENV, WNV and FMDV correspond to the presence of allosteric sites in these less studied polymerases. To this end, we evaluated four ligand binding site predictors (LBSP) to determine the algorithm with the best predictive power at identifying putative allosteric sites in the DENV, WNV and FMDV polymerases using the HCV enzyme as a model system. Our results point to the LISE tool as being the most apt tool for predicting the target allosteric sites in the HCV polymerase. We then used LISE to identify allosteric sites in the other viral pockets and observed novel pockets in each enzyme that shared structural and/or chemical similarities with two of the allosteric sites of the HCV polymerase. Because of the structural similarities that exist among the predicted DENV, WNV or FMDV sites and the known HCV pockets, we hypothesize that allosteric sites may be a conserved feature of viral RNA polymerases. Finally, some of the observed structural and chemical commonalities suggest the possibility of targeting the identified sites with existing HCV polymerase inhibitors or at the minimum using these molecules as a precursors for further optimization.

3.2 Materials and Methods

3.2.1 Description of LBSPs

FTSite

The FTSite algorithm is an energy-based tool developed on the basis that ligand-binding sites will readily accommodate small organic molecules of various shapes and polarity.[1] FTSite employs sixteen organic probes[2] to detect sites of favorable protein-ligand interactions. Use of such diverse probes is thought to increase the likelihood of predicting a broad range of ligand-binding pockets. The algorithm docks the probes to the query protein then identifies a collection of favorable docked poses. These poses are subsequently minimized with the CHARMM force field[3] using the implicit solvent method Analytic Continuum Electrostatics (ACE) and the Newton-Raphson algorithm.[4] The minimized poses are then clustered based on their energies and the six most energetically favored clusters associated with each probe are retained. The next step is to identify overlap between protein regions predicted by different probes and classify these as consensus clusters. These consensus clusters are then ranked according to the number of non-bonded interactions between the protein and all the probes in the cluster. One potential limitation of FTSite is that the use of organic probes can hinder the identification of ligand binding sites on a protein surface. Unlike large cavities, surface binding sites are often shallow and require rearranging of atoms to reduce steric hindrance when accommodating a ligand. The static protein structures employed in most protein-ligand docking experiments preclude rearrangements within the local environment of such pockets to accommodate the probe(s). This in turn provides a potential bias towards ligand binding sites within cavities of proteins where steric hindrances are reduced.[1]

QsiteFinder

Like FTSite, QsiteFinder is an energy-based tool. However, QsiteFinder differs from the former in that it employs only methyl probe(s) to predict the most favorable ligand-

binding sites.[5] QsiteFinder encompasses a number of separate components. First, the LigandSeek module separates the ligand and protein coordinates. The protein coordinates are rotated around their geometric center to better align the protein to the grid, thereby reducing grid points. The next step consists of docking the methyl probe(s) with the query protein. The LIGGRID module calculates the non-bonded interaction energy of the docked probe-protein poses using GRID force parameters.[6] Poses with the most favorable energies are retained and subsequently clustered if they are within 1 Å of each other. Finally, the coordinates are rotated back into their original orientation and a putative ligand-binding site is predicted if 25% or higher of predicted probe clusters overlap at a given location. One known caveat of applying this method is that predictions are generally more accurate for ligand bound conformations than unbound conformations. In Qsite Finder's initial evaluation by Laurie et al. found that even subtle changes in protein structure caused by ligand binding could alter existing binding sites, leading to a drop in precision [5]. However, we note that ligand induced conformational changes may impact the performance of all of these algorithms since they each employ static protein structures as inputs.

LISE

LISE is a geometric and propensity based algorithm conceptually derived from the docking score function MotifScore[7]. MotifScore works by searching a protein-ligand complex for specific networks of interaction motifs consisting of 3 protein and 2 ligand atoms thought to occur primarily at ligand binding sites. It then orients the ligand into the best position to maximize these interactions[7]. These interaction motifs are termed protein triangles within LISE and are associated with a binding site enrichment factor, which denotes the likelihood of these triangles occurring at a ligand-binding site. The enrichment factor is the ratio of the probabilities of a motif occurring at a specific site relative to the probability of its occurrence throughout the entire protein. Protein triangles with binding site enrichment

factors less than one are eliminated. The enrichment factors of the retained protein triangles are used in conjunction with a weighting factor of 1.7 and conservation scores taken directly from PSI-BLAST's position specific scoring matrix to calculate the triangle scores. A three-dimensional grid at 1 Å intervals is then created around the protein and each grid point denoted as protein occupied or empty based on a specific distance threshold. If a grid point is labeled empty, then a grid point score is computed as the sum of all triangle scores within 4 Å of the grid point. The next step is to calculate the sphere score, which is the sum of all the grid point scores centered within 11 Å of each empty grid point. Finally, the sphere scores are ranked with the top predicted site being associated with the highest value. One limitation of LISE is that it omits protein triangles with enrichment factors that are less than one, which most often are on the protein surface. On the other hand, the statistical propensity approach that LISE employs negates the need for tedious energy calculations such as those used in FTsite and QsiteFinder.

LIGSITE^{csc}

LIGSITE^{csc} is a geometry-based algorithm developed using surface-solvent-surface interactions to predict putative binding sites. The first step of this algorithm is to project the query protein onto a three-dimensional grid and align the protein along the xyz axes. After alignment, each grid point is denoted as one of the following: i) protein if it contains a protein atom within 1.6 Å of the grid point; ii) surface (defined using a combination of the protein vdW surface and solvent accessible surface computed using the Connolly algorithm[8]); and iii) solvent (all remaining volume). The next step involves scanning in all direction to identify surface-solvent-surface events, which are characterized as a sequence of protein grid points that are bounded by surface grid points with solvent grid points in between. Next, if a solvent grid point is part of five surface-solvent-surface interactions it is marked as a potential ligand-binding pocket. These pockets are clustered if they are within 3.0 Å of each other. Finally, the

clusters are ranked according to the number of grid points in each. One drawback of LIGSITE^{csc} is that the use of the Connolly Surface restricts the prediction of interior ligand-binding sites within channels and crevices of the protein. Nonetheless, it is also the use of the Connolly surface that allows this tool to consistently identify surface binding sites.

	FTSite	**Qsite Finder**	**LISE**	**LIGSITE^{csc}**
Type of Algorithm	Energy	Energy	Geometric & Propensity	Geometric
Chemical Probe	16 organic probes	Methyl probe(s)	none	none
Pro(s)	diverse array of probes	prediction of surface sites	negates the need for any energy consideration	prediction of surface sites
Con(s)	organic probes cause bias towards predicting large cavities	predictions more accurate for holo structures than apo structures	surface sites can be under predicted	Connolly Surface restricts the prediction of interior sites

Table 3.1: Features of LBSPs Evaluated

3.2.2 Evaluation of LBSPs

Identification of known Polymerase Binding Sites

As mentioned previously, our primary mode of evaluating LBSPs was to detect NNI-1, NNI-2 and NNI-3 sites of the HCV polymerase. Since these sites have been previously identified via x-ray crystallography, we used the Ligand Explorer utility of PDB.org and the Visual Molecular Dynamics (VMD) program to locate all protein residues within 8 Å from the ligands. This distance threshold allowed us to capture long-range interactions between protein residues and the ligands. These include water mediated protein-ligand hydrogen bonds important to the potency of some NNIs that are not captured using shorter distance cutoffs. Residues within the noted distance threshold were also verified using descriptions in the literature.

We note that during the initial development and testing of several LBSPs including FTSite, LISE and LIGSITE^{csc}, a successful hit is defined when at least one point of a predicted binding site (usually the geometric center) is within 4 Å of any ligand atom. However, such a criterion cannot be applied in cases where the binding site is unknown. In addition, a cutoff this short would omit some of the long-range interactions noted above that are important for binding of NNIs. Consequently, we employ different criteria to assess success in this study. First, we focus on identifying residues that comprise a ligand-binding site rather than on identifying the location of ligand atoms themselves. Since the residues around experimentally confirmed ligand binding pockets are known, we can compare these known residues to the residues surrounding the predicted site. In the case of LISE, which outputs the geometric center of a predicted site, a cutoff of 8 Å was employed to identify the protein residues surrounding the predicted site. This allowed us to compute the overlap between the known and predicted binding pocket residues. The percentage of overlap is defined as the total number of predicted pocket residues in common with the known pocket residues, divided by the total number of known pocket residues and multiplied by 100.

We chose an overlap of 50% or greater as denoting a successful hit. This precision threshold ensures that at least half of the residues comprising a binding pocket are identified within the predicted site. When we compare binding site residues described in the literature to those identified using an 8 Å cutoff from the center of mass of each ligand, we observed an overlap of 56% or more. This observation motivates our decision to use a threshold of 50% or greater to denote a successful hit in the other systems employed in this study. Note that this overlap criterion is significantly more stringent than that employed in the development of LBSP tools such as Qsite Finder and LISE, which can be as low as 25%. Another benefit of defining successful hits based on identifying binding pocket residues is that this criterion can be readily extended to cases where the binding site location is unknown. In such cases we can anticipate that residues within 8 Å of predicted binding site locations have a probability of

being part of an actual ligand binding site that is at least as high as the precision threshold (i.e. 50%).

For additional validation, we used LISE to predict known allosteric sites on the FMDV and Coxsackievirus (CSV) polymerases. FMDV and CSV have known allosteric sites in the fingers domain and are both members of the picornaviridae viral family. As these viruses are more distantly related to HCV than DENV or WNV, they provide a more stringent test of the ability of LISE to perform well for diverse viral polymerases.

Structure Preparation

We used three protein data bank (PDB) structures that are representative of the target sites in the HCV polymerase. These structures are 2BRL(NNI-1), 2WHO(NNI-2) and 3CO9(NNI-3). We prepared each crystal structure by deleting chains B and higher (if the structure was solved with multiple chains), ions, water molecules and bound ligands. The free protein structures were individually submitted to online servers for each of the four LBSPs. Deletion of each ligand exposes its binding pocket and allows the tools to use only the protein structure to predict putative binding sites, as is the case with unknown query structures.

In using LISE to search for allosteric sites within the DENV, WNV, FMDV and CSV polymerases, we employed three different structures for each viral enzyme. We used protein structures that did not possess any small molecule inhibitors or for which bound ligands were deleted before submission to the server. The DENV query structures have PDB ID's 4V0R, 2J7W and 2J7U, while for WNV we used the 2HCN, 2HCS and 2HFZ coordinates. For FMDV we used coordinates 1U09, 2F8E and 2D7S while for CSV we used 3DDK, 4Y2A and 4WFZ.

Analysis

We conducted both direct and cross comparisons to evaluate each tool's performance in predicting the target sites in the HCV and FMDV polymerases. The purpose of direct comparisons is to evaluate how well each tool is able to predict each target site when the query protein was previously solved with a ligand in that particular binding pocket. For the direct comparisons, we compared the residues of each predicted site to the residues of the known binding pocket of the query protein. For example, all the predicted sites for the NNI-1 allosteric pocket from the four LBSP tools were compared to the known 2BRL NNI-1 site. In contrast, we performed cross comparisons to evaluate the robustness of each tool. Specifically, we wanted to determine which LBSP was the best at predicting the largest number of known sites regardless of query structure. For cross comparisons, we compared all putative sites to the known binding sites. Using the example above, all predicted pockets for the 2BRL query structure (containing an NNI-1 site) were compared to known NNI-2 and NNI-3 sites. Note that since there is only one elucidated allosteric site in the FMDV and CSV enzymes, there were no cross comparisons for these systems. Finally, electrostatic potentials were evaluated for protein residues comprising the surface of each predicted binding site. The electrostatic potential maps were generated by using the PDB2PQR web server[9, 10] to convert the respective PDB files to PQR format containing the necessary charge distributions. The PQR files were then employed to calculate the electrostatic potentials with the APBS program.[11]

3.3 Results and Discussion

3.3.1. *LISE was Best at Predicting Allosteric Sites in the HCV Polymerase*

Our results revealed that LISE was the best LBSP for predicting each of the target sites within the HCV polymerase in direct comparisons (**table 3.2**). In contrast, both Qsite Finder and LIGSITE^{csc} were able to identify only the NNI-1 and NNI-2 target sites. FTSite proved to be the least successful, as it only identified the NNI-3 pocket. Although the cross

comparisons suggest that both LISE and Qsite Finder were able to predict all of the target sites when every query structure was considered, overall LISE identified the target sites more consistently (**tables S3.1**, **S3.2** in appendix B).

PDB ID of Query Structures:	**Percent Overlap for Top Predicted Sites‡ for Each LBSP**			
	FTSite	Qsite Finder	LISE	LIGSITE^{CSC}
2BRL (NNI-1)	<50%	88%	81%	81%
2WHO (NNI-2)	<50%	75%	66%	75%
3CO9 (NNI-3)	50%	<50%	63%	<50%

Table 3.2: Percent Overlap for Direct Comparisons of the Top Predicted Site(s) and Query Site for Each LBSP.

One of our initial hypotheses was that, due to the diverse array of organic probes used in FTSite,[1] this program would outperform the other tools. However, as mentioned above it failed to predict locations of the NNI-1 and NNI-2 sites. Furthermore, the percent overlap was barely at the threshold cutoff for the NNI-3 pocket and was significantly lower than those observed for LISE and Qsite Finder (**table 3.2** in main text, **B3.1**, **B3.2** and **B3.3** in appendix B). Both NNI-1 and NNI-2 sites are located outside the large central cavity found in the polymerase palm domain. Thus, this observation is consistent with the known limitation (noted in section 3.2.1 Methods) that this tool tends not to identify surface sites. Results from the cross comparisons for FTSite also confirmed the tendency for FTSite to predict interior cavities (**table B3.3** in appendix B). Specifically, all the sites predicted by FTSite in table 2 are located within the central cavity of the protein regardless of input structure. Thus, we conclude that FTSite possesses a strong bias for predicting binding sites that are contained within large interior cavities and would not be ideal for predicting external surface sites. This may limit the utility of this program in a protein for which there is no a priori knowledge regarding the binding site location.

As stated above, both Qsite Finder and LIGSITE^{csc} were able to identify only the NNI-1 and NNI-2 target sites in the direct comparisons (**table 3.2**). Both of these sites can be

found on the surface of the NS5B polymerase. In addition, both sites are located within the thumb domain of NS5B and are predominantly hydrophobic. As we noted previously, Qsite Finder utilizes methyl probes[5] for binding site predictions, which may help to explain why it primarily identified hydrophobic pockets. Despite a bias towards predicting hydrophobic target sites, Qsite Finder was also able to identify other putative sites within different domains of the NS5B protein. While these sites did not correspond to known allosteric sites within NS5B, this observation suggests that Qsite Finder has less bias for predicting pockets within the central cavity of the enzyme compared to FTSite (which only identified sites within this central cavity). An interesting observation was that Qsite Finder was able to predict the NNI-3 site when 2BRL and 2WHO were input structures but not when the corresponding NNI-3 structure 3CO9 was the query. This finding may suggests that Qsite Finder is particularly sensitive to the conformational differences that exist among the query structures. Thus, the predictions from Qsite Finder may be more susceptible to being negatively impacted by the occurrence of ligand induced protein conformational changes (see section 3.2.1). Unlike the hydrophobic probes employed by Qsite Finder, the geometric criteria employed by LIGSITE^{csc} are anticipated to bias the program towards identifying surface cavities at the expense of detecting deeper cavities and channels within the protein. Our results confirm this expectation, as LIGSITE^{csc} only detected the NNI-1 and NNI-2 sites (**table 3.2**).

LISE, in comparison to the other three tools, made more consistent predictions across all input queries. The only exception to this observation is the NNI-1 site, which was only detected when 2BRL was used as a query. However, we note that the other LBSPs that were able to predict NNI-1 also were only successful using the 2BRL structure. Unlike FTSite, Qsite Finder and LIGSITE^{csc} that exhibit biases towards identifying either internal, hydrophobic or surface target sites, LISE is able to accurately predict all site categories. The statistical propensity approach employed in LISE eliminates limitations that may be

associated with energy-based methods such as the need to extensively sample the protein structure with probes or the longer search times typically required in order to perform energy calculations.[7, 12] One potential caveat to note in employing LISE is that the calculation of the triangle score often results in the omission of over 70% of surface triangle motifs,[13] which may hinder the prediction of surface ligand-binding sites such as the NNI-1 or NNI-2 target sites. However, LISE was still able to identify both NNI-1 and NNI-2 sites, suggesting that surface sites that are true binding pockets are not removed based on the triangle score.

It is interesting that the NNI-1 site was only predicted when the 2BRL structure was used as an input query. One possible reason for this observation is that this allosteric site only exists when a ligand is present. The ligand binding pocket in 2BRL lies in a cleft between the thumb and fingers domains.[14] In the absence of a ligand this cleft is completely filled by a flexible loop (residues 9 – 41) from the fingers domain. Thus, this binding pocket is formed via displacing this loop. In this case, one would anticipate that when the ligand is absent the LBSPs would not be able to predict the presence of this site. Results from the cross comparisons support this hypothesis (see **tables S1 to S4** in appendix B). The finding that the methods of the four tested LBSPs are unable to predict ligand-induced binding sites is not surprising given that these tools generally employ static protein structures for their predictions. A consequence of this occurrence is that natively occurring variations in the protein structures employed for the queries may influence the type and number of functionally relevant sites predicted for a given system. Thus, no LBSP had the same target sites predicted across all the query structures. This result probably is due to the fact that the three HCV polymerase query structures used for validation represent, to some degree, the structural variability that one might expect to occur across a solution ensemble. Thus, the three binding sites likely display different characteristics in each structure. LISE displayed the most consistent performance, as the NNI-2 and NNI-3 sites were predicted for all three of the query structures (**tables 3.2 and B3.1**). The noticeably better performance of LISE for the

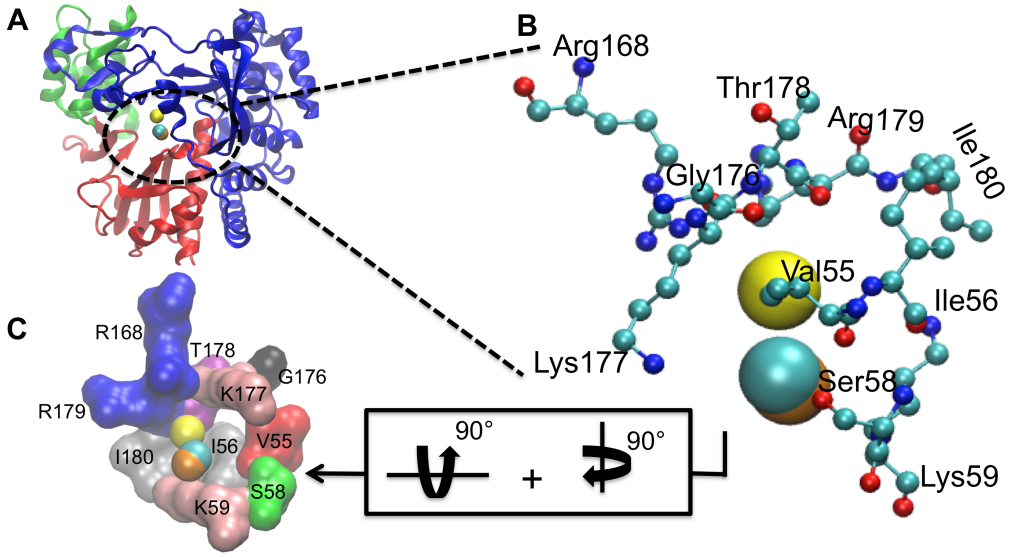
HCV polymerase validates its use to predict polymerase binding sites in the less studied DENV, WNV and FMDV.

Finally, we note that when using these LBSPs it is important to consider their high false positive rates in predicting target sites: every query structure had predictions that were not among any of the known sites. This may be related to the inherent structural variability noted above and sensitivity of LBSPs to this variability. However, the false positive rate can be reduced significantly if one employs multiple query structures of a given protein. Thus, we compare the predictions among the different query structures and choose sites that represent the consensus result. We find that these consensus solutions are more likely to identify functionally relevant pockets. This consideration was the primary reason we employed a minimum of three X-ray crystallographic structures as queries for each of the viral polymerases for which the locations of allosteric sites are not necessarily known.

3.3.2. *LISE Predicts Known Allosteric Sites in FMDV and CSV*

LISE was able to predict the allosteric site in FMDV previously identified by Durk and coworkers. [15] The ligand 5D9 binds to this pocket, which is located in the fingers domain proximal to the NTP binding site (**figure 3.1**). Successful matches were observed with all three input structures. Specifically, comparisons of residues making up the predicted LISE sites and known residues elucidated by mutagenesis resulted in a 100% overlap for the 1U09 query structure and 90% overlap for 2D7S and 2F8E (**figure 3.1**). Furthermore, each of the LISE matches identified Lys59 and Lys177, which have been shown to be critical for inhibitor binding through the mutagenesis studies by Durk et al. noted above. Another test of the robustness of LISE was its ability to predict the presence of the magnesium ions binding site in the FMDV polymerase, identifying each of the residues (Asp238, Asp240, Asp339, Val239, Ile340 and Thr384) that interact with the magnesium ions.[16] Specifically, all of these residues were found in the fourth most highly ranked sites for both the 2D7S and 2F8E query

structures while 5 of the residues were identified as the part of the seventh most highly ranked site for the 1U09 input.

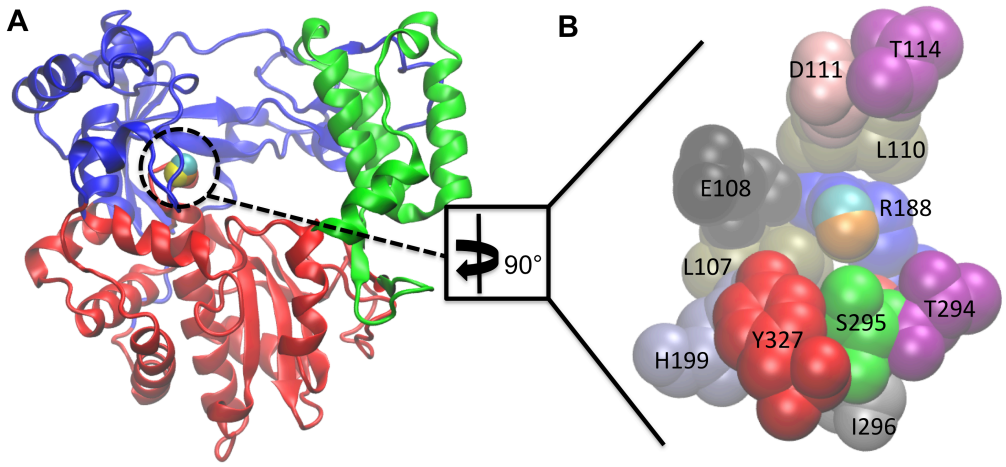


Query Structure	Percent Overlap*
1U09 (476res / yellow)	100%
2D7S (474res / cyan)	90%
2F8E (474res / orange)	90%

*Percent overlap= (total no. of predicted residues in common with known residues) / (total no. of residues identified from mutagenesis [shown in **B** above]) *100

Figure 3.1: LISE accurately predicts the site of a known allosteric inhibitor of the FMDV polymerase. Panel A shows the rear of the enzyme with the LISE predicted spheres from the 3 query structures in van der Waals (VDW) representation (1U09 in yellow, 2F8E in orange and 2D7S in cyan). Panel B is a magnified view of A showing only the residues (in CPK representation) that have been biochemically established to be critical for ligand binding at this site and the LISE predicted spheres (results for query structure 1U09 in yellow, 2F8E in orange and 2D7S in cyan). Panel C displays the residues making up this site depicted in surface representation after 90º rotations about both the y and x axes. Each residue is separately colored and labeled.

LISE was also able to identify a known allosteric site in the polymerase of CSV (**figure 3.2**).[17] CSV can cause a number of ailments including rashes, upper respiratory tract infections and meningitis. While CSV is not our primary focus, the ability of LISE to identify the known CSV allosteric site provides additional support regarding the predictive power of LISE for viral polymerases. This success is further evidence that LISE can capably identify ligand binding sites in diverse viral polymerases. As we will show below, LISE not only performed exceptionally well at predicting previously identified sites, but also predicted new cavities in DENV, WNV and FMDV that have counterparts on the HCV polymerase.



Query Structure	Percent Overlap*
4WFZ (462res / yellow)	83%
4Y2A (466res / cyan)	83%
3DDK (462res / orange)	83%

*Percent overlap= (total no. of predicted residues in common with known residues) / (total no. of residues identified from mutagenesis [shown in **B** above]) *100

Figure 3.2: LISE accurately predicts the site of a known allosteric inhibitor of the CSV polymerase. Panel A shows a frontal view of the enzyme with the LISE predicted spheres from the three query structures in van der Waals (VDW) representation (4WYZ in yellow, 3DDK in orange and 4Y2A in cyan). Panel B is a magnified view of A after a 90º rotation about the y axis showing only the residues (in VDW representation) comprising the ligand binding site and the LISE predicted spheres (results for query structure 4WFZ in yellow, 3DDK in orange and 4Y2A in cyan). Each residue is colored according to its name and is individually labeled. Note: Y195 (in red), which is behind S295 (in green) is the only residue not fully visible in the current view.

3.3.3 *HCV-like Allosteric Pockets Exist in DENV, WNV and FMDV*

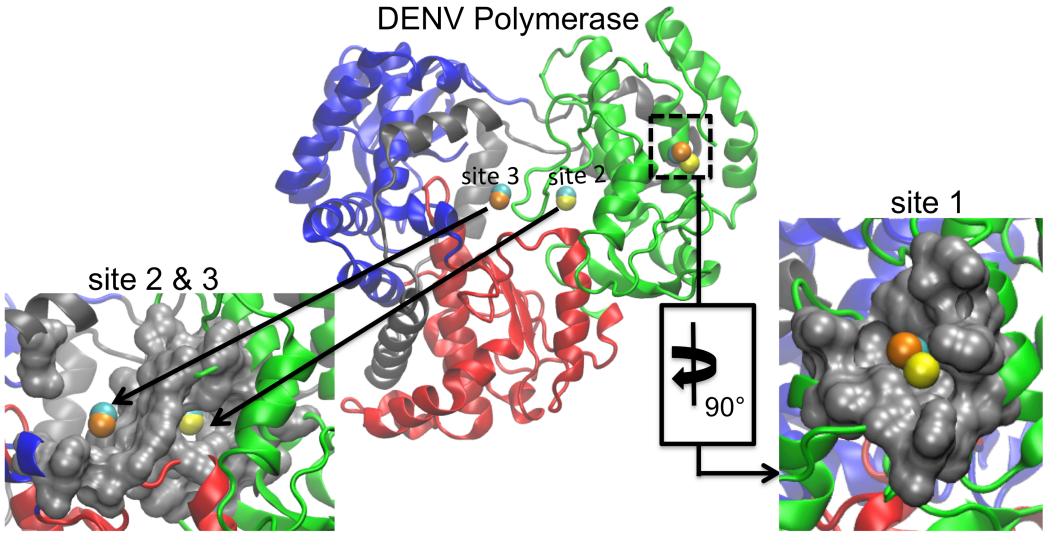
DENV

LISE predicted three putative sites in the DENV polymerase that each had a similar counterpart in the HCV polymerase (**figure 3.3**). Malet and coworkers had previously used the CASTp[18] and PASS[19] LBSP tools to predict two cavities in the DENV polymerase and 5 cavities in the WNV enzyme.[20] These two pockets in the DENV enzyme were subsequently tested for functional utility via mutagenesis studies. One residue in cavity A (Lys756) and four in cavity B (Leu328, Lys330, Trp859, Ile863) were identified as being crucial for viral replication.[21] We compared our LISE predictions with the residues comprising cavities A and B as listed in their work. LISE identified cavity A in the thumb domain for each of the 3

DENV query structures. Specifically, one of the top ten predicted sites (site 1 in **figure 3.3**) for each of 4V0R, 2J7U and 2J7W were matches for cavity A. Overlap values based on the residues provided in Malet et. al. [20] were at or above our 50% threshold (50% for 4V0R, 58% for 2J7U and 67% for 2J7W). We believe site 1 in DENV possesses some similarities to the NNI-2 site within the HCV polymerase. Visually, both sites are found on the lateral surface of the thumb domain in similar locations. This region of the enzyme is displayed in panel B of **figure 1.2** in Chapter 1; further references to the lateral enzyme surface should be understood to denote this location. The DENV site 1 is slightly smaller and is located slightly higher in the thumb domain than the NNI-2 pocket. Thus, only the residues lying at the bottom of the DENV site 1 coincide with the top portion of where the NNI-2 site would be located in the HCV polymerase (**figures 1.1 and 3.3**). Nonetheless, both sites are similarly hydrophobic[22] and posses a deep groove that runs their entire length. However, they differ in that NNI-2 has more positive charge than site 1 in DENV (**figure 3.4**). Although the predicted site 1 did not encompass the critical residue Lys756 identified by Zou et al.,[21] this site warrants further biochemical evaluation as the initial study by Zou et al. only mutated 7 of the residues making up the initially predicted site and our current findings include additional residues that may also be important for ligand binding.

We note that significant differences in structure may contribute to variations in the DENV site predictions. The DENV 4V0R structure is the full-length protein including not only the polymerase domain, but also the methyltransferase (MTase) domain. The MTase has been shown to be involved in the formation of the RNA cap, which plays a role in regulating gene expression.[20, 23, 24] LISE analysis of full-length 4V0R indicates an accumulation of putative sites at the interface of the polymerase and MTase domain. However, these domains are connected by a flexible linker and it is likely that the structure of the domain interface is highly variable, making it difficult to target this region using structure-based drug discovery

efforts. The result obtained using the 4V0R structure also contrast with those obtained for structures containing only the polymerase domain (2J7U and 2J7W).



DENV site 1	S763, R773, S776, N777, C780, D808, M809, L810, W833, E834, D881, T882, M883
DENV site 2	L511, C709, S710, H711, R729, L734, R737, T758, M761, T766, T793, T794, W795, S796, H798, A799, W803
DENV site 3	K401, N492, H495, Q507, G508, R792, W795, S796, I797, H798, A799, H800, H801

Figure 3.3: LISE predicted sites in the DENV polymerase. The center image depicts the DENV enzyme with LISE predicted spheres (4V0R in yellow, 2J7U in orange, 2J7W in cyan). The left image is a magnified surface view of both of the predicted sites 2 and 3 located in the palm domain that share structural and chemical similarities with the HCV NNI-3 pocket. The right image is an enlarged surface view of the predicted site 1, which shares structural and chemical features with the HCV NNI-2 pocket, after 90° rotation about the y axis. The table lists residues in common among the LISE predicted sites.

The latter display predicted binding sites in the various subdomains throughout the enzyme, in agreement with predictions made for WNV and FMDV (see below). We hypothesize that in the 4V0R structure, predicted sites outside of the polymerase/MTase interface exhibit a lower rank (i.e. below the top ten) due to the multitude of predictions within the domain interface. Consequently, we decided to focus on DENV structures containing only the polymerase domain to increase the likelihood that consensus sites in readily targeted regions of the enzyme would be identified. Nonetheless, the occurrence of site 1 in both 2J7U and 2J7W suggests that it represents a realistic binding site location for small molecules. It is possible that this site could be targeted with known NNI-2 chemical scaffolds with altered

functional moieties that are more chemically appropriate for this site. Furthermore, we note that LISE was also able to predict a zinc-binding site that is known to be located at the base

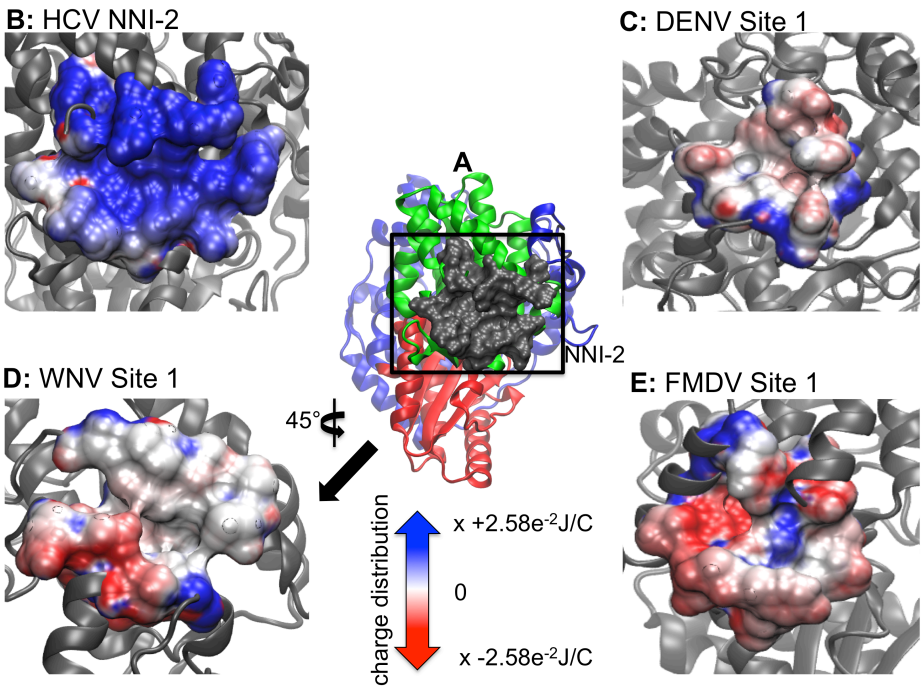


Figure 3.4: Electrostatic potential maps of HCV NNI-2 and sites 1 of DENV, WNV and FMDV highlighting chemical and structural features. Panel A is a lateral view of the HCV polymerase depicting the NNI-2 pocket and the reference used to orient the position of the images from other enzymes. Panels B – E depicts the electrostatic potential of the NNI-2/site 1 in HCV, DENV, WNV and FMDV respectively. The perspective in D has been rotated 45° about the y-axis. Regions of blue and red indicate positively and negatively charged surface regions respectively. The potentials are expressed as multiples of 2.58×10^{-2} Joules per Coulomb (J/C). The color and unit scales are maintained for the remaining electrostatic potential maps.

of the thumb domain at the rear of the enzyme. This gives us high confidence that LISE can make accurate binding site predictions for the DENV polymerase. All four residues (H712, H714, C728 and C847) that coordinate with zinc in this location were identified as highly ranked sites for 2J7U and 2J7W. The inability of LISE to identify a zinc-binding site for the full-length 4V0R structure suggests that presence of the MTase domain may preclude the identification of other viable sites within the polymerase domain. LISE also identified two slightly different sites within the central cavity of the protein that share some structural similarities with the HCV NNI-3 pocket (**figure 3.3**). When DENV sites 2 and 3 are combined they display significant structural similarity to the NNI-3 site, covering a large portion in the middle of the enzyme. Both also contain a mixture of hydrophobic and

hydrophilic residues similar to NNI-3. However, a comparison of the electrostatic potential maps of NNI-3 and DENV sites 2 and 3 demonstrates that the electrostatic potential differs from that present in NNI-3 (**figure 3.5**). In a similar manner to suggestions that have been proposed for NNI-3 inhibitors of HCV polymerase,[25, 26] ligands that bind within the region of the predicted DENV site 3 may block portions of the template channel, preventing the RNA template from accessing the active site and disrupting RNA replication. Site 3 is also in close proximity to the G-loop, while both sites 2 and 3 are located close to the priming loop. The G-loop, located in the fingers subdomain (DENV residues 405-418; WNV residues 407-420), typically protrudes towards the active site of Flavivirus polymerases.[20, 27, 28] Because its location corresponds to the C terminus of the HCV polymerase, the G-loop is hypothesize to play a role in preventing both RNA and incoming NTPs from binding to the enzyme.[20] Also protruding towards the active site, the priming loop in the thumb subdomain (DENV residues 792-804; WNV residues 796-809) is thought to stabilize the *de novo* initiation complex.[20] Small molecules binding within these locations may disrupt changes in the conformation of the priming loop that are needed for the enzyme to shift to the elongation stage of replication.[20]

WNV

In a similar manner to its performance with DENV, LISE was able to predict two sites in the WNV polymerase that share structural features with two allosteric pockets in the HCV polymerase (**figures 3.6**). Cavities A and B that were identified in DENV by Malet et al. (and discussed above) are similar to 2 of the 5 sites that were predicted for the WNV polymerase in the same study. However, we did not observe any matches (i.e. > 50% threshold) between our LISE predictions and residues listed for cavities A and B by Malet and coworkers.[20] We believe that one reason for the lack of a match may be due to differences in the input structures used by Malet et al. and in our work. While we used crystallographic coordinates 2HCN, 2HCS and 2HFZ to make our predictions, it is unclear

whether Malet et. al used an X-ray or modeled structure for their studies. As we have noted above, site predictions can vary appreciably based on the query structure. Even though we did not observe matches to residues surrounding the previously predicted cavities by Malet et al., we were encouraged by the similarities between sites 1 and 2 (**figure 3.6**) and the NNI-2 and NNI-3 pockets in the HCV polymerase. Specifically, all of the query structures predicted sites 1 and 2 shown in **figure 3.6**. Because of the location of site 1 several angstroms away from the expected site of NNI-2, we think this pocket may be novel. In addition, the electrostatic

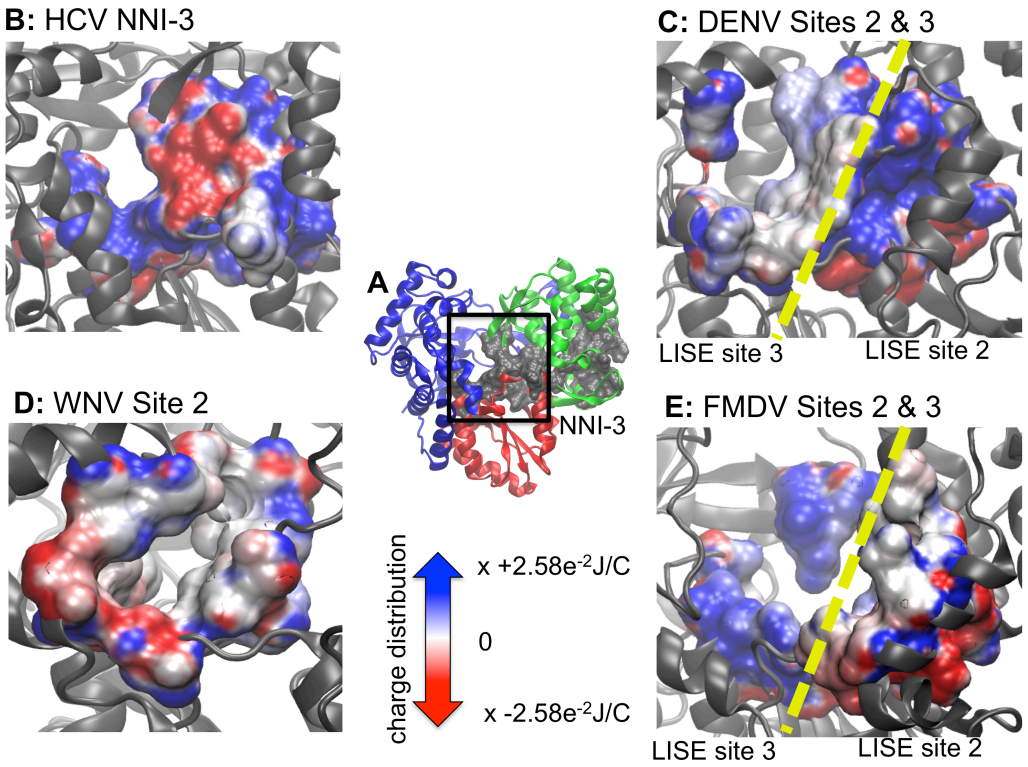
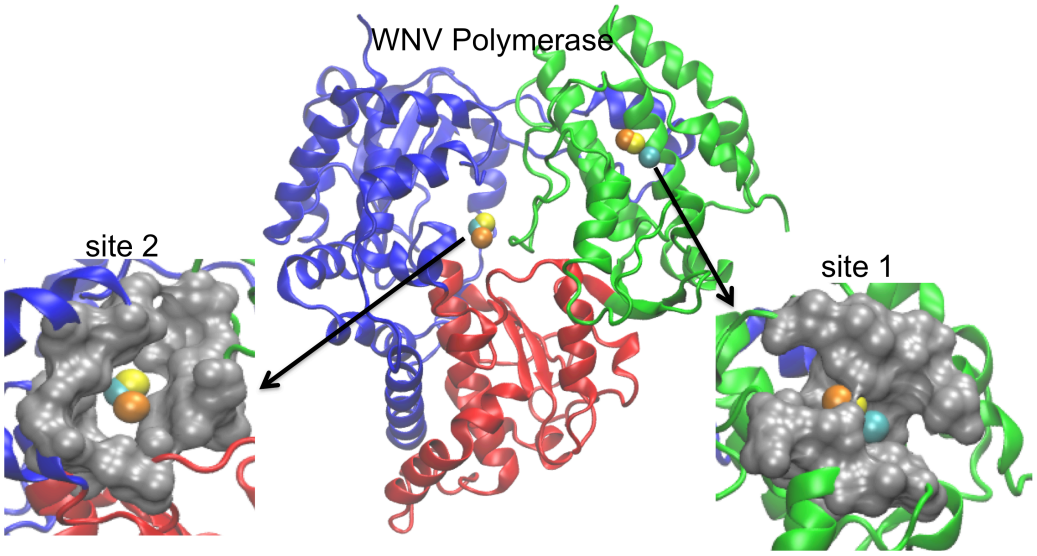


Figure 3.5: Electrostatic potential maps of HCV NNI-3 and predicted palm sites in DENV, WNV and FMDV highlighting chemical and structural features. Panel A is a frontal view of the HCV NNI-3 pocket and the reference used to orient the position of the images from other enzymes. Panels B – E depicts the electrostatic potential of the known or predicted palm allosteric sites in HCV, DENV, WNV and FMDV respectively. Both DENV and FMDV had two separate predicted sites within the palm domain. However, since they are in similar locations both are displayed simultaneously with a yellow dashed line showing the interface between each pair of sites.

potential map of the WNV site 1 (while being almost opposite to that of NNI-2 in the HCV polymerase: see **figures 3.4.B** and **3.4.D**) is very similar to site 1 predicted in the DENV enzyme. Despite the difference in charge distribution, we observe a significant number of residues within the WNV site 1 that exhibit hydrophobicity comparable to that observed for

the NNI-2 pocket. Furthermore, the WNV site 1 possesses a groove that spans the entire cavity, much like that observed in HCV and DENV (**figure 3.4**), which may allow similarly shaped ligands to bind.



WNV site 1	Q765, L768, A781, N782, C785, S786, V790, E812, D813, M814, L815, W838, V885, D886, Y887, S890
WNV site 2	K404, N408, A409, N495, E496, E510, Q606, V607, Y610, R797, T799, I802, H803, A804, G805

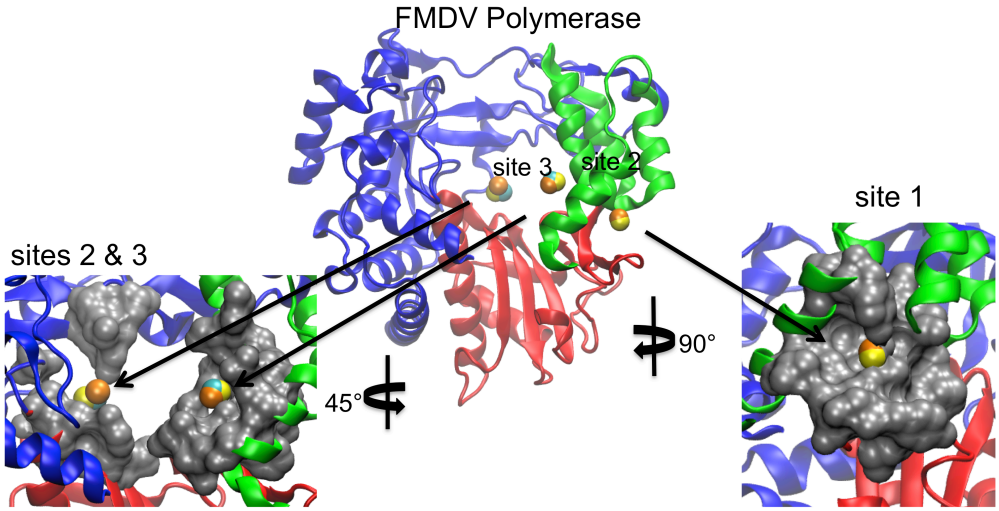
Figure 3.6: LISE predicted sites in the WNV polymerase. The center image depicts the WNV enzyme with LISE predicted VDW spheres in the palm and thumb domains (2HCN in yellow, 2HCS in orange, 2HFZ in cyan). The left image is a magnified surface view of the predicted site 2 located in the palm domain that is structurally similar to the HCV NNI-3 pocket. The right image is an enlarged surface view of the predicted site 1, which is structurally similar to the HCV NNI-2 pocket. The table in the lower panel lists the consensus residues making up each LISE predicted site.

Thus, it may be possible to target site 1 pockets of DENV and WNV with small molecule inhibitors that possess chemical scaffolds that are similar to known HCV NNI-2 ligands (although containing distinct functional groups to match the electrostatic properties of each site). The predicted WNV site 2 is similar to NNI-3 and consists of a mixture of polar and charged residues. It occupies a large fraction of the middle of the enzyme and is reminiscent of the corresponding cavity in the HCV polymerase (**figures 3.5.B** and **3.5.D**). Although the WNV site 2 is more enclosed than the HCV NNI-3, it may be possible to target this site with

established NNI-3 allosteric inhibitors (due to the chemical similarities shown in **figures 3.5.B** and **3.5.D**) or to at least use such ligands as a starting point for developing WNV inhibitors.

FMDV

Like DENV and WNV, LISE predicted sites in the FMDV polymerase that share similar locations and electrostatic properties with the HCV NNI-2 and NNI-3 pockets (**figures 3.4**, **3.5** and **3.7**). In a similar manner to DENV and WNV, the predicted site 1 is not as large as NNI-2 but also possesses a groove that spans the majority of its length. It is located on a region of the enzyme surface that is similar to NNI-2 (**figure 3.7**). However, it



FMDV site 1	H391, M392, D393, Y394, G395, T396, K400, V402, A431, V432, H433, S434, Y455, L458, W462, V463
FMDV site 2	N218, P219, D220, V221, R388, P401, I425, S426, V427, A428, G429, L430, V432, H433, L458, A465, R461, W462
FMDV site 3	K164, Q166, R179, A243, D245, S298, T303, S304, N307, Y336, G337, D338, D339

Figure 3.7: LISE predicted sites in the FMDV polymerase. The center image depicts the FMDV enzyme LISE predicted spheres in the palm and thumb domains (1U09 in yellow, 2F8E in orange and 2D7S in cyan). The left image is a magnified surface view of both of the predicted sites 2 and 3 located in the palm domain that share structural similarities with the HCV NNI-3 pocket. The right image is an enlarged surface view of the predicted site 1 which is structurally similar to the HCV NNI-2 pocket. The table lists the consensus residues making up each LISE predicted site.

differs from NNI-2 in that it is located at the junction of the thumb and palm domains and includes residues from each domain whereas NNI-2 contains only thumb residues. Sites 2 and

3 when combined occupy a location in the central cavity of the enzyme much like NNI-3. Also, sites 2 and 3 display a ratio of positive and negative charges that is similar to that present in the HCV NNI-3 site (**figure 3.5**). These results further support our hypothesis that structural information for one viral polymerase may be used to garner new insights regarding another, even if they possess considerable structural differences. Our findings suggest that despite the genetic and structural differences between the FMDV and HCV polymerases, allosteric sites in the two enzymes have similar geometric and chemical properties. Thus, it may be possible to target these sites with allosteric inhibitors that have similar chemical structures.

3.3.4 Potential Functional Significance of Putative Allosteric Sites

Biochemical experiments combined with computational studies provide insight into how inhibitors bound to the HCV NNI-2 and NNI-3 pockets achieve allosteric inhibition. Both NNI-2 and NNI-3 inhibitors have been shown to disrupt an early stage of replication that precedes elongation.[22, 29] Molecular simulations by our group suggest that inhibitors of the NNI-2 and NNI-3 pockets generally restrict protein fluctuations and disrupt motions observed in the free enzyme.[25, 26, 30] In addition, both type of inhibitors altered correlated motions within enzyme functional motifs in a manner that may disfavor the normal cycle of replication. Specifically, these functional regions are thought to play roles in binding of template, the nascent 3'-end or nascent RNA duplex.[31, 32] Thus, the NNI-2 and NNI-3 allosteric pockets may communicate with residues involved in the binding of template and newly synthesized RNA strands as well as those required to transition to the elongation stage of genome replication. Consequently, the putative allosteric sites identified in the current work may share these functional properties and allow the enzymes to be inhibited in a similar manner to the HCV polymerase.

A recent study using both experimental and computational techniques supports our findings. Manvar and coworkers used in vitro screens and SAR analysis to identify a small molecule scaffold that had inhibitory activity against the DENV polymerase (serotype 2).[33] Fluorescence experiments confirmed that three of the ligands with this scaffold interacted with the polymerase, while docking studies suggest that the ligand with the most favorable affinity binds to a site in the thumb domain.[33] Six of the ten residues identified to interact with this ligand are found within site 1 identified by LISE in our current work. This result further supports the utility of LISE in identifying functionally relevant allosteric sites. Thus, we suggest that the allosteric sites predicted for DENV, WNV and FMDV may also induce inhibition of the viral polymerases and thus should be explored in greater detail for drug discovery.

3.4 Conclusions

Our results suggest that LISE works well to predict allosteric sites for viral polymerases that are structurally similar to the HCV polymerase. LISE was able to identify experimentally elucidated allosteric sites within the HCV, FMDV and CSV polymerases with high confidence. Given that the general architecture of these enzymes is relatively common, LISE may prove useful in identifying allosteric sites for a significant number of other viral polymerases. Moreover, the ability of LISE to identify allosteric sites localized in all three HCV polymerase subdomains suggests that this tool does not possess structural biases that would hinder the identification of allosteric pockets. This characteristic will be of great utility if LISE is used to make predictions for proteins that are not well studied and for which limited structural information is available.

Even though we demonstrate that LISE is the best tool for site prediction among the viral polymerases studied, there are certain caveats that should be considered when LBSPs

such as LISE are used. As discussed in section 3.1, LBSPs are not ideal for predicting ligand-induced sites, as there is no built-in feature within these tools that accounts for the dynamic nature of proteins in sampling various conformations. Thus, there is no way to investigate the presence of sites that only exist within distinct structural states unless one has those states available. To minimize the impact of conformational variation on site prediction, we recommend using as many input structures as feasible and looking for consensus among the predicted sites. These could be obtained from multiple PDB structures or from carrying out molecular dynamics (MD) simulations of the target protein. In either case, one can have greater confidence that the results have converged to describe a realistic binding site when its location is unknown if the results from multiple input structures are consistent with one another. We believe that these consensus sites, such as those identified above for the DENV, WNV and FMDV polymerases, have a higher probability of being functionally relevant. While we consider these sites to be putative in lieu of biochemical studies to definitively establish their existence, they likely represent good starting points for future drug discovery efforts targeting these enzymes.

Our studies suggest that the structural similarities of viral polymerases are associated with similar cavities on the protein surface that could be targeted with small molecules. We identified putative allosteric sites in DENV, WNV and FMDV that possess varying degrees of similarity to the NNI-2 and NNI-3 pockets in the HCV polymerase. Even though the chemical environment of the NNI-2 site in HCV polymerase differs from that of site 1 in DENV, WNV and FMDV, their similar shapes suggest that this pocket is a conserved evolutionary feature. Consequently, this location may have functional significance such as interacting with RNA strands during viral replication or interacting with other viral or host proteins. Thus, it is possible that the activities of the different enzymes can be modulated in similar ways when allosteric ligands bind at these locations and that allosteric inhibition is a shared feature of these polymerases. Moreover, in light of their structural similarities it may

be possible to identify new allosteric inhibitors for one viral enzyme by using the known allosteric inhibitors of another enzyme as a starting point for further optimization. For example, it may be enough to simply change the appropriate functional groups on known chemical scaffolds. We also observe that every enzyme possesses a predicted site within the palm domain that is much like the HCV polymerase NNI-3 pocket with regard to its structural and chemical environment. Thus, it may be possible to target these locations in the different enzymes with very similar small molecules. Finally, we emphasize that the strategy employed here of evaluating LBSPs on a well-studied enzyme to predict sites in less studied enzymes may facilitate the identification of novel binding sites. Ultimately this knowledge can assist in the process of discovering new drugs that target these enzymes, especially in cases for which efficacious and cost-effective treatments are not widely available.

3.5 Acknowledgements

Jodian A. Brown was funded by NIH F31 pre-doctoral grant (GM-106958-01). We wish to thank Paul Zhong-ru Xie for his technical assistance regarding the LISE program.

3.6 References

[1] Ngan, C. H. H., D. R.; Zerbe, B.; Grove, L. E.; Kozakov, D.; Vajda, S. (2012) FTSite: high accuracy detection of ligand binding sites on unbound protein structures, *Bioinformatics 28*, 286 - 287

[2] Kozakov, D. H., D. R.; Chuang, G.-Y.; Cencic, R.; Brenke, R.; Grove, L. E.; Beglov, D.; Pelletier, J.; Whitty, A.; Vajda, S. (2011) Structural conservation of druggable hot spots in protein-protein interfaces *Proceedings of the National Academy of Sciences 108*, 13528 - 13533.

[3] Brooks, B. R. B., R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. . (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, *Journal of Computational Chemistry 4*, 187 - 217.

[4] Brenke, R. K., D.; Chuang, G. Y.; Beglov, D.; Hall, D.; Landon, M. R.; Mattos, C.; Vajda, S. (2009) Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques, *Bioinformatics 25*, 621 - 627.

[5] Laurie, A. T. J., R. M. . (2005) Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites. , *Bioinformatics 21*, 1908 - 1916
[6] Jackson, R. M. (2002) Q-fit: a probabilistic method for docking molecular fragments by sampling low energy conformational space, *Journal of Computer-Aided Molecular Design 16*, 43 - 57.
[7] Xie, Z. R. H., M. J. (2010) An interaction-motif-based scoring function for protein-ligand docking, *BMC Bioinformatics 11*, 298.
[8] Connolly, M. L. (1983) Analytical molecular surface calculation, *Journal of Applied Crystallography 16*, 548 - 558.
[9] Dolinsky, T. J. C., P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. (2007) PDB2PQR: Expanding and updgrading automated preparation of biomolecular structures for molecular simulations *Nucleic Acids Research 35*, W522 - W525.
[10] Dolinsky, T. J. N., J. E.; McCammon, J. A.; Baker, N. A. (2004) PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations, *Nucleic Acids Research 32*, W665 - W667.
[11] Baker, N. A. S., D.; Joseph, S.; Holst, M. J.; McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome, *Proceedings of the National Academy of Sciences 98*, 10037 - 10041.
[12] Gu, J. Q., Z.; Wang, X. (2012) Identification of Ligand Binding Pockets for Unbound State Proteins Using Local Residue References In *International Conference on Artificial Intelligence and Embedded Systems*, pp 14 - 18 Singapore
[13] Xie, Z. R. (2012) Ligand-binding site prediction using ligand-interacting and binding site-enriched protein triangles, *Bioinformatics 28* 1579 - 1585
[14] Di Marco, S. V., C.; Tomei, L.; Altamura, S.; Harper, S.; Narjes, F.; Koch, U.; Rowley, M.; De Francesco, R.; Migliaccio, G.; Carfi, A. (2005) Interdomain communication in hepatitis C virus polymerase abolished by small molecule inhibitors bound to a novel allosteric site, *Journal of Biological Chemistry 280*.
[15] Durk, R. C. S., K.; Cornelison, C. A.; Rai, D. K.; Matzek, K. B.; Leslie, M. D.; Schafer, E.; Marchand, B.; Adedeji, A.; Michailidis, E.; Dost, C. A.; Moran, J.; Pautler, C.; Rodriguez, L. L.; McIntosh, M.A.; Rieder, E.; Sarafianos, S. G. (2010) Inhibitors of Foot and Mouth Disase Virus Targeting a Novel Pocket of the RNA-dependent RNA Polymerase, *PLoS ONE 5*, e15049.
[16] Ferrer-Orta, C. A., A.; Perez-Lugue, R., Escarmís, C.; Domingo, E.; Verdaguer, N. (2004) Structure of Foot-and-Mouth Disease Virus RNA-dependent RNA Polymerase and Its Complex with a Template-Primer RNA, *Journal of Biological Chemistry 279*, 47212 - 47221.
[17] Van Der Linden, L. V.-A., L.; Selisko, B.; Ferrer-Orta, C.; Liu, X.; Lanke, K; Ulferts, R.; De Palma, A.M.; Tanchis, F.; Goris, N.; Lefebvre, D. (2015) The RNA template channel of the RNA-dependent RNA polymerase as a target for development of antiviral therapy of multiple genera within a virus family, *PLoS Pathogens 11*, e1004733.

[18] Dundas, J. O., Z.; Tseng, J.; Binkowski, A.; Turpaz, Y.; Liang, J. (2006) CASTp: computed atas of surface topography of proteins with structural and topographical mapping of functionally annotated residues, *Nucleic Acids Research 34*, W116 - W118.
[19] Brady, G. P. S., P. F. W. (2000) Fast Prediction and Visualization of Protein Binding Pockets With PASS, *Journal of Computer-Aided Molecular Design 14*, 383 - 401
[20] Malet, H. M., N.; Selisko, B.; Romette, J.-L.; Alvarez, K.; Guillemot, J. C.; Tolou, H.; Yap, T. L.; Vasudevan, S.; Lescar, J.; Canard, B. (2008) The flavivirus polymerase as a target for drug discovery, *Antiviral Research 80*, 23 - 35.
[21] Zou, G. C., Y.-L.; Dong, H.; Lim, C. C.; Yap, L. J.; Yau, Y. H.; Shochat, S. G.; Lescar, J.; Shi, P.-Y. (2011) Functional Analysis of Two Cavities in Flavivirus NS5 Polymerase *Journal of Biological Chemistry 286*, 14362 - 14372.
[22] Ontoria, J. M. (2009) Identification and Biological Evaluation of a Series of 1H-Benzo[de]isoquinoline-1,3(2H)-diones as Hepatitis C Virus NS5B Polymerase Inhibitors, *Journal of Medicinal Chemistry 52*, 5217-5227.
[23] Dong, H. F., K.; Züst, R.; Lim, S. P.; Qiin, C.-P.; Shi P.-Y. . (2014) Flavivirus RNA methylation *Journal of General Virology 95* 763 - 768
[24] Zhou, Y. R., D.; Zhao, Y.; Dong, H.; Ren, S.; Li, Z.; Guo, Y.; Bernard, K. A.; Shi, P.-Y.; Li, H. (2007) Structure and Function of Flavivirus NS5 Methyltransferase, *Journal of Virology 81*, 3891-3903.
[25] Brown, J. A. T., I. F. (2015) Dual Allosteric Inhibitors Jointly Modulate Protein Structure and Dynamics in the Hepatitis C Virus Polymerase *Biochemistry 54*, 4131 - 4141.
[26] Davis, B. C. B., J. A.; Thorpe, I. F. (2015) Allosteric Inhibitors Have Distinct Effects, but Also Common Modes of Action, in the HCV Polymerase *Biophysical Journal 108*, 1785-1795
[27] Malet, H. E., M.-P.; Selisko, B.; Butcher, R. E.; Wright, P. J.; Roberts, M.; Gruez, A.; Sulzenbacher, G.; Vonrhein, C.; Bricogne, G.; Mackenzie, J. M.; Khromykh, A. A.; Davidson, A. D.; Canard, B. . (2007) Crystal Structure of the RNA Polymerase Domain of the West Nile Virus Non-structural Protein 5, *Journal of Biological Chemistry 282*, 10678 - 10689
[28] Yap, T. L. X., T.; Chen, Y.-L.; Malet, H.; Egloff, M.-P.; Canard, B.; Vasudevan, S. G.; Lescar, J. . (2007) Crystal Structure of the Dengue Virus RNA-Dependent RNA Polymerase Catalytic Domain at 1.85-Angstrom Resolution, *Journal of Virology 81*, 4753 - 4765
[29] Ruebsam, F. (2008) Pyrrolo[1,2-b]pyridazin-2-ones as potent inhibitors of HCV NS5B polymerase, *Bioorganic & Medicinal Chemistry Letters 18*, 3616-3621.
[30] Davis, B. C. T., I. F. (2013) Thumb inhibitor binding eliminates functionally important dynamics in the hepatitis C virus RNA polymerase, *Proteins: Structure, Function and Bioinformatics 81*, 40-52.
[31] Moustafa, I. M. S., H.; Morton, B.; Colina, C. M.; Cameron, C. E. (2011) Molecular Dynamics Simulations of Viral RNA Polymerases Link Conserved and Correlated Motions of Functional Elements to Fidelity, *Journal of Molecular Biology 410*, 159-181.

[32] Sesmero, E. T., I. F. . (2015) Using the Hepatitis C Virus RNA-Dependent RNA Polymerase as a Model to Understand Viral Polymerase Structure, Function and Dynamics, *Viruses 7*, 3974 - 3994.
[33] Manvar, D. K. c. u. k. z., İ.; Erensoy, G.; Tatar, E.; Deryabaşoğulları, G.; Reddy, H.; Talele, T. T.; Cevik, O.; Kaushik-Basu, N. (2015) Discovery of conjugated thiazolidinone-thiadiazole scaffold as anti-dengue virus polymerase inhibitors, *Biochemical and Biophysical Research Communications in press*.

Chapter 4: Summary, Significance and Future Directions

4.1 Dissertation Summary: Major Findings of this Dissertation

Our overall findings have contributed to a better comprehension of the mechanism(s) of enhanced inhibition in the context of allosteric inhibitors of NS5B, while simultaneously furthering our understanding of allostery. Specifically, we have identified a model of enhanced inhibition that results in synergistic modulation of the free energy landscape and protein dynamics when a palm and thumb inhibitor are simultaneously bound. We have also identified a ligand binding site predictor (LBSP) tool that is apt for allosteric site prediction in viral polymerases. Using this LBSP we predicted putative allosteric sites within the polymerases of the Dengue (DENV), West Nile (WNV) and foot-and-mouth disease (FMDV) viruses that share structural and/or chemical similarities with the HCV polymerase.

4.1.1 Enhanced Inhibition in NS5B when NNI-2 and NNI-3 Inhibitors are Present (Chapters 2)

An understanding of the molecular details mediating enhanced inhibition is critical to furthering our knowledge of the functional behavior of NS5B and will yield insights that may allow for the discovery and/or optimization of more effective HCV therapies. Specifically, our findings suggest that the simultaneous binding of inhibitors from non-overlapping (NNI-2 and NNI-3) allosteric sites is energetically favorable. Moreover, once they are bound these molecules can jointly modulate the protein's conformational space and dynamics. Of note, we observed conformations in

the dually bound system that were novel in comparison to those present in the free enzyme or when it is bound to individual inhibitors. It is likely that these conformations are not conducive to enzyme function. These conformations sampled by the dually bound system exhibited a mixture of the induced-fit and conformational selection models of allostery (discussed in Chapters 1 and 2). Finally, we identified specific changes to the hydrogen bonding network critical to the observed impacts on the structure and dynamics of the dually bound enzyme. Taken together, the results provide molecular evidence to support the use of NNI-2 and NNI-3 inhibitors to simultaneously challenge the HCV polymerase. Furthermore, the insights described here suggest that one possible approach for inhibition of the DENV, WNV and FMDV polymerases that are not as widely studied as HCV is to employ multiple allosteric inhibitors that bind to nonoverlapping sites.

4.1.2 Identification of Putative Allosteric Sites in the Polymerases of DENV, WNV and FMDV (Chapter 3)

The first step to understanding allosteric inhibition in the DENV, WNV and FMDV polymerases is to identify allosteric pockets. To this end, the work carried out in chapter 4 identified a site prediction tool that was effective at allosteric pocket prediction in viral polymerases. Secondly, our study revealed that there are 2 allosteric pockets present in each of the DENV, WNV and FMDV polymerases that share structural and/or chemical similarities with sites present in the HCV enzyme. With the exception of one of the putative allosteric sites identified in the DENV polymerases, all pockets discussed in Chapter 4 were novel. Due to the structural

and/or chemical similarities that exist among these enzymes and the HCV polymerase, we propose that it may be possible to target these novel pockets with existing allosteric inhibitors of the HCV polymerase or, at the very minimum, to use these current inhibitors as initial scaffolds that could be further optimized to match the chemical environments of the novel pockets.

4.2 Significance

There were two major findings from this work: i) identification of an allosteric mechanism of inhibition when two inhibitors are bound to NS5B and ii) identification of novel putative allosteric pockets in the DENV, WNV and FMDV polymerases.

In Chapter 2, we delineated a model of allosteric inhibition for the HCV polymerase in the context of dual inhibitor binding. The model reveals that there are local changes to hydrogen bond interactions and correlated motions that occur and lead to more global protein impacts on conformational sampling. Specifically, we have identified key protein-ligand interactions and protein residues involved in generating correlated motions or novel conformations that may be nonfunctional. This information can be used in the optimization of new or existing combination therapies, potentially adding to the available HCV treatment regiments. As mentioned in Chapter 1, combination therapies are particularly important in combating drug resistance, a major challenge that continues to plague advancement in the treatment of the HCV infection. Furthermore, the high cost for recent available HCV treatments excludes access to millions of patients, which warrants the need for additional therapies that effective yet economical.

Because the HCV, DENV, WNV and FMDV polymerases share some genetic and structural properties, we hypothesized that the HCV enzyme could be used to evaluate LBSPs apt for allosteric pocket identification in the other viral polymerases. We tested this hypothesis and work in Chapter 4 revealed that the HCV polymerase allowed us to effectively evaluate the predictive power of several LBSPs with respect to allosteric sites. Our application of the most successful LBSP indicates that there are putative allosteric sites in the DENV, WNV and FMDV polymerases that share structural and chemical commonalities with those found in the HCV enzyme. DENV, WNV and FMDV are not as well studied as HCV, primarily because the pharmaceutical market is not considered to be as lucrative as for other viral infections such as HCV or HIV. However, the approach employed in Chapter 3 provides a cost effective and faster means to the first step of drug discovery, which is binding site identification. The putative allosteric sites identified in this work can be further probed e.g. through virtual screening of existing HCV inhibitors in attempts to discovery new small molecule inhibitors of the DENV, WNV and FMDV polymerases. Finally, our approach can be extended to other viral systems that are not well studied.

4.3 Future Implications

The work discussed in the dissertation has broadened our knowledge of the mechanisms of allosteric inhibition in the HCV polymerase, while simultaneously serving as a springboard for future research efforts to identify allosteric inhibitors for

the DENV, WNV and FMDV polymerases using both experimental and computational approaches. There are five main research projects that should follow:

1) MD simulations of inhibitors from different combinations of non-overlapping allosteric sites within the HCV polymerase to determine the similarities and differences in comparison with the structural, dynamic and thermodynamic changes observed in Chapter 2. Insights from such a study will allow us to understand if the changes observed in Chapter 2 are a general feature of different inhibitor and site combinations or whether one combination is more favorable for enhanced inhibition. One calculation that could be immediately conducted is to perform MD simulations of the HCV polymerase individually and doubly bound to known NNI-1 and NNI-3 ligands to evaluate and compare the structural, dynamic and thermodynamic properties of the enzyme. To this end, I have led the efforts to parameterize NNI-1 (PDB ID: 23E) and NNI-3 (PDB ID: 26F) inhibitors that are simultaneously present in the PDB IB: 3QGE crystal of the HCV polymerase. Preliminary MD simulations suggested that ligand 23E is very bulky, resulting in detrimental steric clashes with protein residues found within the delta 1 loop shown in figure 4.1. To probe a site combination involving the NNI-1 site, it may be better to use the PDB ID: POO ligand. This small molecule has also been shown experimentally to bind to this thumb site and is smaller than 23E, resulting in less protrusion from the site once bound (figure 4.1).[1] Furthermore, we have previously published MD simulations of the HCV polymerase bound to the POO ligand and several other allosteric inhibitors.[2] I

was responsible for obtaining and optimizing parameters using the methodology referenced in Chapter 2 for the palm allosteric inhibitors needed to describe the ligands in the subsequent MD simulations. The major findings from this work were that the allosteric inhibitors worked in one of three ways: restricting conformational space (e.g. NNI-2), destabilizing conformational sampling (e.g. NNI-1), or blocking template from accessing the active site (e.g. NNI-3).[2] The success of this prior study further substantiates the viability of the POO ligand (NNI-1) in the proposed study. The parameters necessary to describe these small molecule inhibitors (23E, 26F, POO) to be used in future MD simulations are provided in Appendix A.

2) Hydrogen-deuterium exchange experiments coupled to mass spectrometry may be employed to evaluate the impact of inhibitor binding on the dynamics of the HCV polymerase. Results from such a study may provide additional evidence for the changes to dynamics predicted in this dissertation and further validate the use of computational approaches in understanding protein allostery in viral polymerases. Secondly, isothermal calorimetry experiments of free and ligand-bound HCV polymerase can serve to validate the trends in binding free energy that were predicted in this dissertation.
3) MD simulations of free and ligand-bound HCV polymerase in the presence of RNA template to evaluate any impact on the conformational sampling and dynamics mediated by inhibitor binding (e.g. such as that predicted in Chapter 2). Our work revealed that the free enzyme sampled closed and open conformations while ligand binding shifted the population equilibrium to a

dominant state where the conformations were more closed. We also observed altered patterns of motions in key functional motifs such as for template binding. The presence of template may shift the free energy landscape of free and inhibitor bound HCV polymerase in ways that may modify the predicted trends discussed in Chapter 2. Any additional changes to the free energy landscape and conformational space in the presence of template may provide us with a better framework for understanding the allosteric behavior of the HCV polymerase in the context of dual inhibitor binding.

4) Use of structure-activity relationships of the DENV, WNV and FMDV sites identified in this dissertation to screen existing HCV NNI-2 and NNI-3 inhibitors and additional drug databases. Currently, there are limited therapies available for DENV, WNV and FMDV. Virtual Screening provides us with an economical and fast approach to identifying potential drug candidates. In support of the validity of such an approach, I was involved in studies to construct a pharmacophore model based on protein-ligand interactions of the HCV polymerase and use it to screen numerous databases including a library of small molecule inhibitors that I curated from literature.[3] We identified an allosteric inhibitor of HCV polymerase that had better potency than previously reported inhibitors, which strongly validated our computational approach and its ability to identify meaningful drug candidate(s) of viral polymerases.

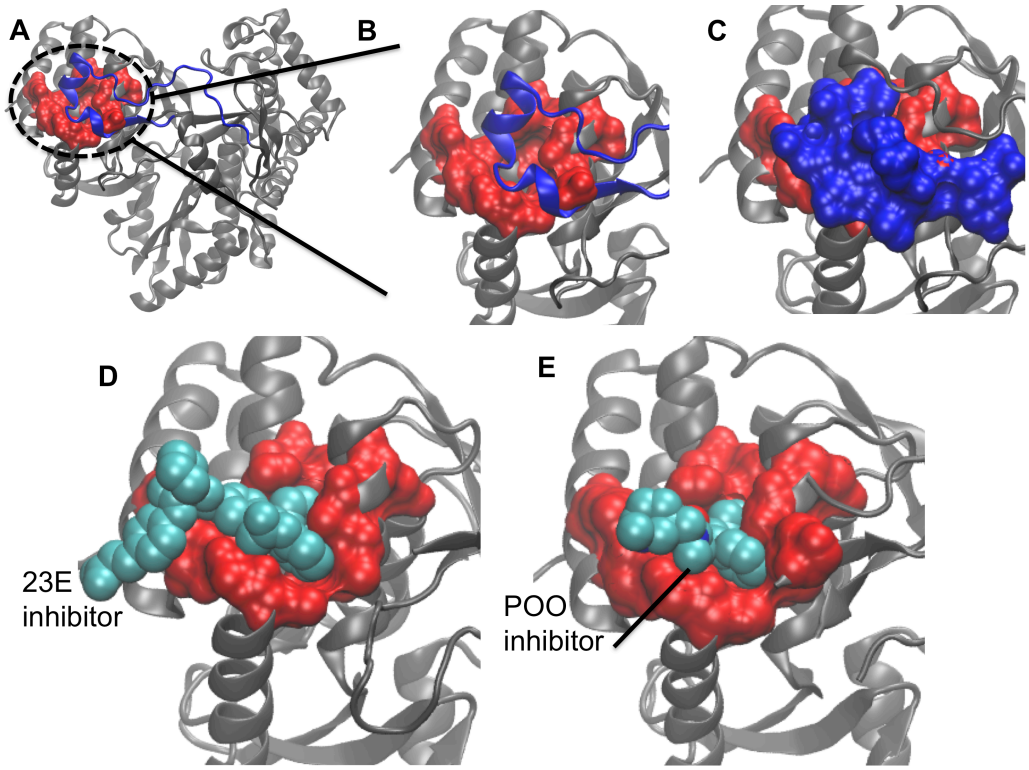


Figure 4.1: The NNI-1 site of the HCV polymerase is occluded by the delta one loop in the absence of ligand. Panel A: Rear view of the polymerase using the crystal PDB ID: 1QUV that was solved without inhibitor where the NNI-1 pocket is shown in red surface representation and the delta one loop in blue new cartoon view. Panels B and C are both magnified views of A; B highlights the NNI-1 pocket depicted in red surface representation while C demonstrates that the delta one loop (residues 9 to 41) in blue surface representation blocks entry to the NNI-1 pocket. Panel D highlights the 23E inhibitor (in cyan) in the NNI-1 pocket of the 3QGE crystal. Panel E: shows that the POO inhibitor that also binds to the NNI-1 site is a smaller ligand than 23E and thus reduces the potential for steric clashes with the delta one loop.

4.4 Acknowledgements

We employed the University of Maryland, Baltimore County (UMBC) high performance computing facility (HPCF) and the Stampede HPC cluster of Extreme Science and Engineering Discovery Environment (XSEDE) in preliminary MD studies mentioned in section 4.3. We wish to thank Kendra Johnson who was a part of the team instrumental for parameterizing the 26F and 23E small molecule inhibitors. HPCF is funded by the U.S. National Science Foundation through the MRI program (grant numbers: CNS-0821258 and CNS-1228778) and the SCREMS program (grant number: DMS-0821311), with significant support from UMBC. XSEDE is supported by the National Science Foundation (grant number: OCI-1053575). Jodian Brown was funded by NIH F31 predoctoral fellowship (GM106958).

4.4 References

[1] Di Marco, S. V., C.; Tomei, L.; Altamura, S.; Harper, S.; Narjes, F.; Koch, U.; Rowley, M.; De Francesco, R.; Migliaccio, G.; Carfi, A. (2005) Interdomain communication in hepatitis C virus polymerase abolished by small molecule inhibitors bound to a novel allosteric site, *Journal of Biological Chemistry 280*.

[2] Davis, B. C. B., J. A.; Thorpe, I. F. (2015) Allosteric Inhibitors Have Distinct Effects, but Also Common Modes of Action, in the HCV Polymerase *Biophysical Journal 108*, 1785-1795

[3] Weidlich, I. E. F., I. V.; Brown, J.; Kaushik-Basu, N.; Krishnan, R.; Nicklaus, M. C.; Thorpe, I. F. (2013) Inhibitors for the hepatitis C virus RNA polymerase explored by SAR with advanced machine learning methods, *Bioorganic & medicinal chemistry 21*, 3127-3137.

Appendix A

Table A2.1: Ligand Parameters for the 3MS Molecule

GROUP		CHARGE
ATOM C1	CG2R51	-0.23
ATOM C2	CG2R51	-0.29
ATOM C3	CG2RC0	-0.19
ATOM N1	NG2RC0	-0.07
ATOM C4	CG2R51	0.08
ATOM N2	NG2R61	0.53
ATOM C5	CG2R63	0.07
ATOM C6	CG2R67	0.09
ATOM C7	CG2R62	-0.20
ATOM C8	CG2R67	0.50
ATOM N3	NG2R62	-0.58
ATOM S1	SG3O2	0.58
ATOM C9	CG2R61	0.07
ATOM C10	CG2R61	0.25
ATOM N4	NG2R61	-0.37
ATOM C11	CG2R61	-0.115
ATOM C12	CG2R61	0.185
ATOM C13	CG2R61	-0.115
ATOM C14	CG2R61	-0.115
ATOM N5	NG311	-0.48
ATOM S2	SG3O2	0.42
ATOM O1	OG2P1	-0.36
ATOM O2	OG2P1	-0.41
ATOM O3	OG2P1	-0.41
ATOM O4	OG311	-0.40
ATOM O5	OG2D4	-0.40
ATOM C15	CG321	0.085
ATOM C16	CG321	-0.18
ATOM C17	CG311	-0.09
ATOM C18	CG331	-0.27
ATOM C19	CG331	-0.27
ATOM C20	CG331	-0.27
ATOM O6	OG2P1	-0.36
ATOM H1	HGR51	0.23
ATOM H2	HGR61	0.115
ATOM H3	HGR52	0.14
ATOM H4	HGA3	0.090
ATOM H5	HGA3	0.090
ATOM H6	HGA3	0.090
ATOM H7	HGA3	0.090
ATOM H8	HGA3	0.090
ATOM H9	HGA3	0.090
ATOM H10	HGA1	0.090

ATOM H11 HGP1 0.33
ATOM H12 HGR61 0.115
ATOM H13 HGR61 0.115
ATOM H14 HGR51 0.19
ATOM H15 HGA2 0.090
ATOM H16 HGA2 0.090
ATOM H17 HGA2 0.090
ATOM H18 HGA2 0.090
ATOM H19 HGA3 0.090
ATOM H20 HGA3 0.090
ATOM H21 HGA3 0.090
ATOM H22 HGP1 0.31
ATOM H23 HGP1 0.51
BOND C1 C2
BOND C1 C4
BOND C1 H14
BOND C2 C3
BOND C2 H1
BOND C3 N1
BOND C3 C7
BOND N1 C4
BOND N1 N2
BOND C4 H3
BOND N2 C5
BOND N2 C15
BOND C5 C6
BOND C5 O5
BOND C6 C7
BOND C6 C8
BOND C7 O4
BOND C8 N3
BOND C8 N4
BOND N3 S1
BOND S1 C9
BOND S1 O2
BOND S1 O3
BOND C9 C10
BOND C9 C11
BOND C10 N4
BOND C10 C14
BOND N4 H22
BOND C11 C12
BOND C11 H2
BOND C12 C13
BOND C12 N5
BOND C13 C14
BOND C13 H12
BOND C14 H13
BOND N5 S2

BOND N5 H11
BOND S2 O1
BOND S2 C20
BOND S2 O6
BOND O4 H23
BOND C15 C16
BOND C15 H15
BOND C15 H16
BOND C16 C17
BOND C16 H17
BOND C16 H18
BOND C17 C18
BOND C17 C19
BOND C17 H10
BOND C18 H19
BOND C18 H20
BOND C18 H21
BOND C19 H4
BOND C19 H5
BOND C19 H6
BOND C20 H7
BOND C20 H8
BOND C20 H9
IMPR C5 C6 N2 O5
END
read param card flex append
* Parameters generated by analogy by
* CHARMM General Force Field (CGenFF) program version 0.9.6 beta
*
! Penalties lower than 10 indicate the analogy is fair; penalties between 10
! and 50 mean some basic validation is recommended; penalties higher than
! 50 indicate poor analogy and mandate extensive validation/optimization.
BONDS
CG2R61 NG2R61 370.00 1.3790 ! 3MSA , from CG2R61 NG2RC0, penalty= 10
CG2R62 CG2R67 305.00 1.3750 ! 3MSA , from CG2R61 CG2R67, penalty= 5
CG2R62 CG2RC0 300.00 1.3600 ! 3MSA , from CG2R61 CG2RC0, penalty= 5
CG2R62 OG311 334.30 1.4110 ! 3MSA , from CG2R61 OG311, penalty= 5
CG2R63 CG2R67 305.00 1.3750 ! 3MSA , from CG2R61 CG2R67, penalty= 25
CG2R67 NG2R61 370.00 1.3790 ! 3MSA , from CG2R61 NG2RC0, penalty= 40
CG2R67 NG2R62 450.00 1.3050 ! 3MSA , from CG2R61 NG2R62, penalty= 30
CG321 NG2R61 400.00 1.4560 ! 3MSA , from CG331 NG2R61, penalty= 6
NG2R61 NG2RC0 550.00 1.3600 ! 3MSA , from NG2D1 NG2S1, penalty= 430
NG2R62 SG3O2 180.00 1.7920 ! 3MSA , from NG2S3 PG1, penalty= 400
ANGLES

CG2R61 CG2R61 NG2R61 100.00 121.40 ! 3MSA , from CG2R61 CG2R61 NG2RC0, penalty= 11
CG2R67 CG2R62 CG2RC0 40.00 118.00 ! 3MSA , from CG2R62 CG2R62 CG2R62, penalty= 4.5
CG2R67 CG2R62 OG311 45.20 120.00 ! 3MSA , from CG2R61 CG2R61 OG311, penalty= 5.5
CG2RC0 CG2R62 OG311 45.20 120.00 ! 3MSA , from CG2R61 CG2R61 OG311, penalty= 6.5
CG2R67 CG2R63 NG2R61 70.00 113.50 ! 3MSA , from CG2R62 CG2R63 NG2R61, penalty= 1.5
CG2R67 CG2R63 OG2D4 100.00 124.60 ! 3MSA , from CG2R62 CG2R63 OG2D4, penalty= 1.5
CG2R62 CG2R67 CG2R63 40.00 120.00 ! 3MSA , from CG2R61 CG2R67 CG2R61, penalty= 3.5
CG2R62 CG2R67 CG2R67 40.00 120.00 ! 3MSA , from CG2R61 CG2R67 CG2R67, penalty= 0.5
CG2R63 CG2R67 CG2R67 40.00 120.00 ! 3MSA , from CG2R61 CG2R67 CG2R67, penalty= 3
CG2R67 CG2R67 NG2R61 40.00 120.00 ! 3MSA , from CG2R61 CG2R67 CG2R67, penalty= 47
CG2R67 CG2R67 NG2R62 40.00 120.00 ! 3MSA , from CG2R61 CG2R67 CG2R67, penalty= 45
NG2R61 CG2R67 NG2R62 70.00 122.20 ! 3MSA , from NG2R61 CG2R64 NG2R62, penalty= 45
CG2R51 CG2RC0 CG2R62 130.00 132.00 ! 3MSA , from CG2R51 CG2RC0 CG2R61, penalty= 0.5
CG2R62 CG2RC0 NG2RC0 80.00 118.80 ! 3MSA , from CG2R61 CG2RC0 NG2RC0, penalty= 0.5
CG321 CG321 NG2R61 70.00 113.50 ! 3MSA , from CG321 CG321 NG2S1, penalty= 2
NG2R61 CG321 HGA2 51.50 109.50 ! 3MSA , from NG2S1 CG321 HGA2, penalty= 2
CG2R61 NG2R61 CG2R67 30.00 120.00 ! 3MSA , from CG2R62 NG2R61 CG2R62, penalty= 3
CG2R61 NG2R61 HGP1 32.00 117.40 ! 3MSA , from CG2R62 NG2R61 HGP1, penalty= 1.5
CG2R63 NG2R61 CG321 70.00 115.40 ! 3MSA , from CG2R63 NG2R61 CG331, penalty= 0.9
CG2R63 NG2R61 NG2RC0 70.00 122.00 ! 3MSA , from CG2R62 NG2R61 CG2R63, penalty= 48.5
CG2R67 NG2R61 HGP1 32.00 117.40 ! 3MSA , from CG2R62 NG2R61 HGP1, penalty= 1.5
CG321 NG2R61 NG2RC0 70.00 120.50 ! 3MSA , from CG2R62 NG2R61 CG331, penalty= 49.4
CG2R67 NG2R62 SG3O2 40.00 110.50 ! 3MSA , from CG2R61 NG2R62 CG2R64, penalty= 333
CG2R51 NG2RC0 NG2R61 15.00 130.50 ! 3MSA , from CG2R51 NG2RC0 CG2R61, penalty= 47
CG2RC0 NG2RC0 NG2R61 15.00 119.80 ! 3MSA , from CG2R61 NG2RC0 CG2RC0, penalty= 47
CG2R62 OG311 HGP1 65.00 108.00 ! 3MSA , from CG2R61 OG311 HGP1, penalty= 0.5

CG2R61 SG3O2 NG2R62 60.00 101.00 ! 3MSA , from CG2R61 SG3O2 OG2P1, penalty= 44
NG2R62 SG3O2 OG2P1 85.00 121.00 ! 3MSA , from OG2P1 SG3O2 OG2P1, penalty= 44
DIHEDRALS
CG2R51 CG2R51 CG2RC0 CG2R62 3.0000 2 180.00 ! 3MSA , from CG2R51 CG2R51 CG2RC0 CG2R61, penalty= 0.5
HGR51 CG2R51 CG2RC0 CG2R62 2.8000 2 180.00 ! 3MSA , from HGR51 CG2R51 CG2RC0 CG2R61, penalty= 0.5
CG2R51 CG2R51 NG2RC0 NG2R61 3.0000 2 180.00 ! 3MSA , from CG2R51 CG2R51 NG2RC0 CG2R61, penalty= 47
HGR52 CG2R51 NG2RC0 NG2R61 1.4000 2 180.00 ! 3MSA , from HGR52 CG2R51 NG2RC0 CG2R61, penalty= 47
CG2R61 CG2R61 CG2R61 NG2R61 1.5000 2 180.00 ! 3MSA , from CG2R61 CG2R61 CG2R61 NG2RC0, penalty= 11
NG2R61 CG2R61 CG2R61 SG3O2 3.0000 2 180.00 ! 3MSA , from CG2R61 CG2R61 CG2R61 SG3O2, penalty= 47
NG2R61 CG2R61 CG2R61 HGR61 0.8000 2 180.00 ! 3MSA , from NG2RC0 CG2R61 CG2R61 HGR61, penalty= 11
CG2R61 CG2R61 NG2R61 CG2R67 4.0000 2 180.00 ! 3MSA , from CG2R62 CG2R62 NG2R61 CG2R62, penalty= 13
CG2R61 CG2R61 NG2R61 HGP1 1.0000 2 180.00 ! 3MSA , from CG2R62 CG2R62 NG2R61 HGP1, penalty= 11.5
CG2R61 CG2R61 SG3O2 NG2R62 0.0000 6 0.00 ! 3MSA , from CG2R61 CG2R61 SG3O2 OG2P1, penalty= 44
CG2RC0 CG2R62 CG2R67 CG2R63 3.1000 2 180.00 ! 3MSA , from CG2R61 CG2R61 CG2R67 CG2R61, penalty= 9.5
CG2RC0 CG2R62 CG2R67 CG2R67 3.1000 2 180.00 ! 3MSA , from CG2R61 CG2R61 CG2R67 CG2R67, penalty= 6.5
OG311 CG2R62 CG2R67 CG2R63 3.1000 2 180.00 ! 3MSA , from CG2R61 CG2R61 CG2R61 OG311, penalty= 38
OG311 CG2R62 CG2R67 CG2R67 3.1000 2 180.00 ! 3MSA , from CG2R61 CG2R61 CG2R61 OG311, penalty= 82.5
CG2R67 CG2R62 CG2RC0 CG2R51 4.0000 2 180.00 ! 3MSA , from CG2R61 CG2R61 CG2RC0 CG2R51, penalty= 5.5
CG2R67 CG2R62 CG2RC0 NG2RC0 3.5000 2 180.00 ! 3MSA , from CG2R61 CG2R61 CG2RC0 NG2RC0, penalty= 5.5
OG311 CG2R62 CG2RC0 CG2R51 3.1000 2 180.00 ! 3MSA , from CG2R61 CG2R61 CG2R61 OG311, penalty= 53.5
OG311 CG2R62 CG2RC0 NG2RC0 1.0000 2 180.00 ! 3MSA , from NG2R60 CG2R61 CG2R61 OG311, penalty= 67
CG2R67 CG2R62 OG311 HGP1 0.9900 2 180.00 ! 3MSA , from CG2R61 CG2R61 OG311 HGP1, penalty= 5.5
CG2RC0 CG2R62 OG311 HGP1 0.9900 2 180.00 ! 3MSA , from CG2R61 CG2R61 OG311 HGP1, penalty= 6.5
NG2R61 CG2R63 CG2R67 CG2R62 1.8000 2 180.00 ! 3MSA , from CG2R62 CG2R62 CG2R63 NG2R61, penalty= 40
NG2R61 CG2R63 CG2R67 CG2R67 1.0000 2 180.00 ! 3MSA , from NG2R60 CG2R61 CG2R67 CG2R67, penalty= 47

OG2D4 CG2R63 CG2R67 CG2R62 1.0000 2 180.00 ! 3MSA , from CG2R62 CG2R62 CG2R63 OG2D4, penalty= 40
OG2D4 CG2R63 CG2R67 CG2R67 1.0000 2 180.00 ! 3MSA , from CG2R62 CG2R62 CG2R63 OG2D4, penalty= 88.5
CG2R67 CG2R63 NG2R61 CG321 11.0000 2 180.00 ! 3MSA , from CG2R62 CG2R62 NG2R61 CG331, penalty= 17.4
CG2R67 CG2R63 NG2R61 NG2RC0 1.5000 2 180.00 ! 3MSA , from CG2R62 CG2R63 NG2R61 CG2R62, penalty= 50
OG2D4 CG2R63 NG2R61 CG321 11.0000 2 180.00 ! 3MSA , from OG2D4 CG2R63 NG2R61 CG331, penalty= 0.9
OG2D4 CG2R63 NG2R61 NG2RC0 1.6000 2 180.00 ! 3MSA , from OG2D4 CG2R63 NG2R61 CG2R62, penalty= 48.5
CG2R62 CG2R67 CG2R67 NG2R61 0.8900 2 180.00 ! 3MSA , from CG2R61 CG2R67 CG2R67 CG2R61, penalty= 47.5
CG2R62 CG2R67 CG2R67 NG2R62 0.8900 2 180.00 ! 3MSA , from CG2R61 CG2R67 CG2R67 CG2R61, penalty= 45.5
CG2R63 CG2R67 CG2R67 NG2R61 0.8900 2 180.00 ! 3MSA , from CG2R61 CG2R67 CG2R67 CG2R61, penalty= 50
CG2R63 CG2R67 CG2R67 NG2R62 0.8900 2 180.00 ! 3MSA , from CG2R61 CG2R67 CG2R67 CG2R61, penalty= 48
CG2R67 CG2R67 NG2R61 CG2R61 4.0000 2 180.00 ! 3MSA , from CG2R62 CG2R62 NG2R61 CG2R62, penalty= 90
CG2R67 CG2R67 NG2R61 HGP1 1.0000 2 180.00 ! 3MSA , from CG2R62 CG2R62 NG2R61 HGP1, penalty= 88.5
NG2R62 CG2R67 NG2R61 CG2R61 0.2000 2 180.00 ! 3MSA , from NG2R62 CG2R64 NG2R61 CG2R63, penalty= 48.5
NG2R62 CG2R67 NG2R61 HGP1 3.6000 2 180.00 ! 3MSA , from NG2R62 CG2R64 NG2R61 HGP1, penalty= 45
CG2R67 CG2R67 NG2R62 SG3O2 2.0000 2 180.00 ! 3MSA , from CG2R61 CG2R61 NG2R62 CG2R64, penalty= 230.5
NG2R61 CG2R67 NG2R62 SG3O2 2.0000 2 180.00 ! 3MSA , from NG2R61 CG2R64 NG2R62 CG2RC0, penalty= 197
CG2R51 CG2RC0 NG2RC0 NG2R61 1.0000 2 180.00 ! 3MSA , from CG2R51 CG2RC0 NG2RC0 CG2R61, penalty= 47
CG2R62 CG2RC0 NG2RC0 CG2R51 1.0000 2 180.00 ! 3MSA , from CG2R61 CG2RC0 NG2RC0 CG2R51, penalty= 0.5
CG2R62 CG2RC0 NG2RC0 NG2R61 0.8000 2 180.00 ! 3MSA , from CG2R61 CG2RC0 NG2RC0 CG2R61, penalty= 47.5
CG311 CG321 CG321 NG2R61 0.2000 3 0.00 ! 3MSA , from CG321 CG321 CG321 NG2S1, penalty= 2.6
NG2R61 CG321 CG321 HGA2 0.1950 3 0.00 ! 3MSA , from NG2S1 CG321 CG321 HGA2, penalty= 2
CG321 CG321 NG2R61 CG2R63 1.8000 1 0.00 ! 3MSA , from CG321 CG321 NG2S1 CG2O1, penalty= 100
CG321 CG321 NG2R61 NG2RC0 1.8000 1 0.00 ! 3MSA , from CG321 CG321 NG2S1 CG2O1, penalty= 125
HGA2 CG321 NG2R61 CG2R63 0.1900 3 0.00 ! 3MSA , from HGA3 CG331 NG2R61 CG2R63, penalty= 6
HGA2 CG321 NG2R61 NG2RC0 0.0000 3 0.00 ! 3MSA , from HGA3 CG331 NG2R61 CG2R62, penalty= 54.5

CG2R63 NG2R61 NG2RC0 CG2R51 0.0000 2 180.00 ! 3MSA , from CG2R61 CG2R61 NG2RC0 CG2R51, penalty= 503
CG2R63 NG2R61 NG2RC0 CG2RC0 0.5000 2 180.00 ! 3MSA , from CG2R61 NG2R62 NG2R62 CG2R64, penalty= 502
CG321 NG2R61 NG2RC0 CG2R51 11.0000 2 180.00 ! 3MSA , from CG2R62 CG2R62 NG2R61 CG331, penalty= 540.9
CG321 NG2R61 NG2RC0 CG2RC0 11.0000 2 180.00 ! 3MSA , from CG2R62 CG2R62 NG2R61 CG331, penalty= 533.9
CG2R67 NG2R62 SG3O2 CG2R61 0.0000 6 0.00 ! 3MSA , from CG2R61 CG2R61 SG3O2 OG2P1, penalty= 537.5
CG2R67 NG2R62 SG3O2 OG2P1 0.0000 6 0.00 ! 3MSA , from CG2R61 CG2R61 SG3O2 OG2P1, penalty= 480.5
IMPROPERS
CG2R63 CG2R67 NG2R61 OG2D4 90.0000 0 0.00 ! 3MSA , from CG2R63 CG2R62 NG2R61 OG2D4, penalty= 4
END

Table A2.2: Ligand Parameters for the VGI Molecule.

GROUP		CHARGE
ATOM C14	CG2R61	-0.03
ATOM C15	CG2R61	0.23
ATOM BR1	BRGR1	-0.415
ATOM C16	CG2R61	-0.03
ATOM C17	CG2R61	-0.155
ATOM C18	CG2R61	-0.10
ATOM C13	CG2R61	0.19
ATOM N2	NG2S1	-0.39
ATOM C9	CG2O1	0.64
ATOM O1	OG2D1	-0.48
ATOM C12	CG2O1	0.64
ATOM O2	OG2D1	-0.49
ATOM C8	CG2R61	-0.26
ATOM C10	CG2R61	0.06
ATOM C11	CG2R61	-0.215
ATOM C7	CG2R61	-0.065
ATOM C1	CG2R61	-0.06
ATOM C2	CG2R61	0.19
ATOM C3	CG2R61	-0.33
ATOM C4	CG2R61	-0.085
ATOM C5	CG2R61	-0.285
ATOM C6	CG2R61	0.43
ATOM N1	NG2S1	-0.8
ATOM C19	CG321	0.14
ATOM C20	CG321	0.17
ATOM O3	OG311	-0.66

ATOM	H1	HGP1	0.375					
ATOM	H2	HGR62	0.12					
ATOM	H3	HGR62	0.12					
ATOM	H4	HGR61	0.115					
ATOM	H5	HGR61	0.125					
ATOM	H6	HGR61	0.125					
ATOM	H7	HGR61	0.115					
ATOM	H8	HGR61	0.115					
ATOM	H9	HGR61	0.115					
ATOM	H10	HGR61	0.115					
ATOM	H11	HGP1	0.36					
ATOM	H12	HGA2	0.09					
ATOM	H13	HGA2	0.09					
ATOM	H14	HGA2	0.09					
ATOM	H15	HGA2	0.09					
BOND	BR1	C15	C15	C16	C16	C17	C17	C18	C18	C13
BOND	C13	C14	C14	C15	C13	N2	N2	C9	C9	C3
BOND	C3	C2	C2	C8	C8	C12	C12	N2	C3	C4
BOND	C4	C5	C5	C6	C6	C1	C1	C2	C1	C7
BOND	C7	C11	C11	C10	C10	C8	C6	N1	N1	C19
BOND	C19	C20	C20	O3	C9	O1	C12	O2	C16	H3
BOND	C17	H4	C18	H5	C14	H2	C4	H9	C5	H10
BOND	C7	H8	C11	H7	C10	H6	N1	H11	C19	H12
BOND	C19	H13	C20	H14	C20	H15	O3	H1		
DONO	H1	O3								
DONO	H11	N1								
ACCE	O3	C20								
ACCE	O2	C12								
ACCE	O1	C9								
IC	C15	C16	C17	C18	1.3958	118.84	0.00	120.75	1.3944	
IC	C16	C17	C18	C13	1.3966	120.75	0.00	119.03	1.3944	
IC	C17	C18	C13	C14	1.3944	119.03	0.00	121.53	1.3951	
IC	C16	C15	C14	C13	1.3958	121.65	0.00	118.20	1.3951	
IC	C17	C18	C13	N2	1.3944	119.03	180.0	119.51	1.4355	
IC	C18	C13	N2	C9	1.3944	119.51	70.62	117.00	1.4120	
IC	C13	N2	C9	C3	1.4355	117.00	180.0	115.85	1.4779	
IC	N2	C9	C3	C2	1.3944	115.85	0.0	120.93	1.4164	
IC	C9	C3	C2	C8	1.4779	120.93	0.0	120.17	1.4164	
IC	C3	C2	C8	C12	1.4164	120.17	0.0	120.54	1.4827	
IC	C2	C8	C12	N2	1.4164	120.54	0.0	115.98	1.4097	
IC	N2	C9	C3	C4	1.4120	115.85	180.0	118.87	1.3833	
IC	C9	C3	C4	C5	1.4779	118.87	180.0	120.19	1.4049	
IC	C3	C4	C5	C6	1.3833	120.19	0.0	121.45	1.3883	
IC	C4	C5	C6	C1	1.4049	121.45	0.0	119.14	1.4309	
IC	C5	C6	C1	C2	1.3883	119.14	0.0	118.62	1.4270	
IC	C6	C1	C2	C3	1.4309	118.62	0.0	119.85	1.4164	
IC	C3	C2	C1	C7	1.4164	119.85	180.0	118.82	1.4172	

IC	C2	C1	C7	C11	1.4270	118.82	0.0	120.40	1.3833
IC	C1	C7	C11	C10	1.4172	120.40	0.0	120.88	1.4077
IC	C7	C11	C10	C8	1.3833	120.88	0.0	119.69	1.3841
IC	C2	C1	C6	N1	1.4270	118.62	175.6	118.73	1.4143
IC	C1	C6	N1	C19	1.4309	118.73	-74.9	115.30	1.4728
IC	C6	N1	C19	C20	1.4143	115.30	163.9	109.88	1.5219
IC	N1	C19	C20	O3	1.4728	109.88	180.0	111.52	1.4247
IC	C13	C14	C15	BR1	0.0	0.0	180.0	119.0	1.9063
IC	C14	C15	C16	H3	0.0	0.0	0.0	0.0	0.0
IC	C15	C16	C17	H4	0.0	0.0	180.0	0.0	0.0
IC	C16	C17	C18	H5	0.0	0.0	180.0	0.0	0.0
IC	C18	C13	C14	H2	0.0	0.0	180.0	0.0	0.0
IC	C12	N2	C9	O1	0.0	125.92	180.0	120.0	1.2280
IC	C9	N2	C12	O2	0.0	0.0	180.0	120.0	1.2280
IC	C2	C3	C4	H9	0.0	0.0	180.0	0.0	0.0
IC	C2	C8	C10	H6	0.0	0.0	180.0	0.0	0.0
IC	C1	C6	C5	H10	0.0	0.0	180.0	0.0	0.0
IC	C1	C7	C11	H7	0.0	0.0	180.0	0.0	0.0
IC	C6	N1	C19	H12	0.0	0.0	-73.6	0.0	0.0
IC	C6	N1	C19	H13	0.0	0.0	44.6	0.0	0.0
IC	N1	C19	C20	H14	0.0	0.0	56.9	0.0	0.0
IC	N1	C19	C20	H15	0.0	0.0	-61.3	0.0	0.0
IC	C19	C20	O3	H1	0.0	0.0	-70.0	107.57	0.0
END									
read param card flex append									
* Parameters needed for new molecules									
*									
ANGLES									
CG2O1 NG2S1 CG2O1 50.00 125.92 ! VGI; INITIAL GUESS by _IT_									
CG2R61 NG2S1 CG321 70.00 115.00 ! VGI; INITIAL GUESS by _IT_									
DIHEDRALS									
CG2R61 NG2S1 CG2O1 CG2R61 2.50 1 180.00 ! Retinol CROT; _IT_ inital guess									
CG2O1 NG2S1 CG2O1 OG2D1 2.50 1 0.00 ! Retinal PACP; _IT_ inital guess									
CG2O1 NG2S1 CG2O1 CG2R61 1.80 1 180.0 ! _IT_ optimized									
CG2R61 CG2R61 NG2S1 CG321 2.00 1 106.4 ! _IT_ optimized									
CG2R61 NG2S1 CG321 CG321 0.30 2 15.25 ! _IT_ optimized									
CG2R61 NG2S1 CG321 HGA2 0.20 3 3.75 ! _IT_ optimized									
NG2S1 CG321 CG321 OG311 0.20 3 0.00 ! _IT_ optimized									

Table A2.3: Ligand Parameters for the 26F Molecule

GROUP		! CHARGE
ATOM C1	CG2R61	-0.116
ATOM C2	CG2R61	-0.116

ATOM	C3	CG2R61 -0.110
ATOM	C4	CG2R61 -0.110
ATOM	C5	CG2R61 -0.115
ATOM	C6	CG2R61 -0.115
ATOM	C7	CG2R61 -0.116
ATOM	C8	CG2R61 -0.116
ATOM	C9	CG2R61 -0.091
ATOM	C10	CG2R61 -0.027
ATOM	C11	CG2R61 0.219
ATOM	C12	CG2R61 0.219
ATOM	C13	CG2R61 0.213
ATOM	C14	CG2R64 0.550
ATOM	C15	CG2R61 0.609
ATOM	C16	CG2R64 0.712
ATOM	C17	CG2O1 0.530
ATOM	C18	CG321 -0.096
ATOM	C19	CG321 0.044
ATOM	C20	CG321 -0.102
ATOM	C21	CG311 0.179
ATOM	C22	CG331 -0.057
ATOM	C23	CG331 -0.097
ATOM	C24	CG331 -0.035
ATOM	C25	CG321 -0.044
ATOM	C26	CG302 0.588
ATOM	N27	NG2R62 -0.603
ATOM	N28	NG2R62 -0.745
ATOM	N29	NG301 -0.450
ATOM	N30	NG301 -0.190
ATOM	N31	NG2S1 -0.470
ATOM	O32	OG2D1 -0.540
ATOM	O33	OG2P1 -0.386
ATOM	O34	OG2P1 -0.372
ATOM	O35	OG301 -0.398
ATOM	O36	OG301 -0.453
ATOM	O37	OG301 -0.440
ATOM	O38	OG301 -0.367
ATOM	F39	FGA3 -0.140
ATOM	F40	FGA3 -0.140
ATOM	F41	FGA3 -0.140
ATOM	S42	SG3O2 0.460
ATOM	H43	HGR61 0.115
ATOM	H44	HGR61 0.115
ATOM	H45	HGR61 0.115
ATOM	H46	HGR61 0.115
ATOM	H47	HGR61 0.115
ATOM	H48	HGR61 0.115
ATOM	H49	HGR61 0.115
ATOM	H50	HGR61 0.115
ATOM	H51	HGR61 0.115

ATOM H52 HGA2 0.090
ATOM H53 HGA2 0.090
ATOM H54 HGA2 0.090
ATOM H55 HGA2 0.090
ATOM H56 HGA2 0.090
ATOM H57 HGA2 0.090
ATOM H58 HGA1 0.090
ATOM H59 HGA3 0.090
ATOM H60 HGA3 0.090
ATOM H61 HGA3 0.090
ATOM H62 HGA3 0.090
ATOM H63 HGA3 0.090
ATOM H64 HGA3 0.090
ATOM H65 HGA3 0.090
ATOM H66 HGA3 0.090
ATOM H67 HGA3 0.090
ATOM H68 HGA2 0.090
ATOM H69 HGA2 0.090
ATOM H70 HGP1 0.319
!Bond order
BOND C1 C3
BOND C1 C10
BOND C1 H43
BOND C2 C4
BOND C2 C10
BOND C2 H44
BOND C3 C11
BOND C3 H45
BOND C4 C11
BOND C4 H46
BOND C5 C7
BOND C5 C12
BOND C5 H47
BOND C6 C8
BOND C6 C12
BOND C6 H48
BOND C7 C13
BOND C7 H49
BOND C8 C13
BOND C8 H50
BOND C9 C14
BOND C9 C15
BOND C9 H51
BOND C10 C25
BOND C11 O35
BOND C12 O36
BOND C13 S42
BOND C14 N27
BOND C14 N29

BOND C15 N28
BOND C15 O37
BOND C16 N27
BOND C16 N28
BOND C16 O38
BOND C17 C21
BOND C17 N31
BOND C17 O32
BOND C18 C19
BOND C18 N29
BOND C18 H52
BOND C18 H53
BOND C19 N30
BOND C19 H54
BOND C19 H55
BOND C20 C21
BOND C20 N29
BOND C20 H56
BOND C20 H57
BOND C21 N30
BOND C21 H58
BOND C22 O35
BOND C22 H59
BOND C22 H60
BOND C22 H61
BOND C23 O37
BOND C23 H62
BOND C23 H63
BOND C23 H64
BOND C24 O38
BOND C24 H65
BOND C24 H66
BOND C24 H67
BOND C25 N31
BOND C25 H68
BOND C25 H69
BOND C26 O36
BOND C26 F39
BOND C26 F40
BOND C26 F41
BOND N30 S42
BOND N31 H70
BOND O33 S42
BOND O34 S42
IMPR C17 C21 N31 O32
END
read param card flex append

* Parameters generated by analogy by
* CHARMM General Force Field (CGenFF) program version 0.9.7 beta
*
! Penalties lower than 10 indicate the analogy is fair; penalties between 10
! and 50 mean some basic validation is recommended; penalties higher than
! 50 indicate poor analogy and mandate extensive validation/optimization.
BONDS
CG2R64 NG301 330.00 1.4000 ! 26FA , from CG2R61 NG311, penalty= 20
CG2R64 OG301 230.00 1.3820 ! 26FA , from CG2R61 OG301, penalty= 15
CG302 OG301 360.00 1.4150 ! 26FA , from CG321 OG301, penalty= 55
CG311 NG301 263.00 1.4740 ! 26FA , from CG321 NG311, penalty= 9
CG321 NG301 263.00 1.4740 ! 26FA , from CG321 NG311, penalty= 5
NG301 SG3O2 235.00 1.6500 ! 26FA , from NG311 SG3O2, penalty= 5
ANGLES
NG2R62 CG2R61 OG301 110.00 120.00 ! 26FA , from CG2R61 CG2R61 OG301, penalty= 45
CG2R61 CG2R64 NG301 40.00 120.00 ! 26FA , from CG2R61 CG2R61 NG311, penalty= 15.6
NG2R62 CG2R64 NG301 45.80 122.30 ! 26FA , from CG321 CG2R61 NG2R60, penalty= 55
NG2R62 CG2R64 OG301 110.00 120.00 ! 26FA , from CG2R61 CG2R61 OG301, penalty= 60
OG301 CG302 FGA3 118.00 107.00 30.00 2.15500 ! 26FA , from FGA3 CG302 FGA3, penalty= 36.3
CG2O1 CG311 NG301 43.70 110.00 ! 26FA , from CG2O2 CG321 NG321, penalty= 7.3
CG321 CG311 NG301 43.70 112.20 ! 26FA , from CG331 CG321 NG311, penalty= 5.5
NG301 CG311 HGA1 32.40 109.50 50.00 2.13000 ! 26FA , from NG311 CG321 HGA2, penalty= 4.6
CG2R61 CG321 NG2S1 50.00 107.00 ! 26FA , from CG2O2 CG321 NG2S1, penalty= 28.5
CG311 CG321 NG301 43.70 112.20 ! 26FA , from CG331 CG321 NG311, penalty= 2.1
CG321 CG321 NG301 43.70 112.20 ! 26FA , from CG331 CG321 NG311, penalty= 1.5
NG301 CG321 HGA2 32.40 109.50 50.00 2.13000 ! 26FA , from NG311 CG321 HGA2, penalty= 0.6
CG2R64 NG301 CG321 55.00 113.00 ! 26FA , from CG2R61 NG311 CG321, penalty= 8
CG311 NG301 CG321 53.00 110.90 ! 26FA , from CG3AM0 NG301 CG3AM0, penalty= 36.6
CG311 NG301 SG3O2 60.00 115.00 ! 26FA , from CG321 NG311 SG3O2, penalty= 5.6
CG321 NG301 CG321 53.00 110.90 ! 26FA , from CG3AM0 NG301 CG3AM0, penalty= 36.6
CG321 NG301 SG3O2 60.00 115.00 ! 26FA , from CG321 NG311 SG3O2, penalty= 5
CG2R61 OG301 CG302 65.00 108.00 ! 26FA , from CG2R61 OG301 CG321, penalty= 12

CG2R64 OG301 CG331 65.00 108.00 ! 26FA , from CG2R61 OG301 CG331, penalty= 3
CG2R61 SG3O2 NG301 70.00 97.00 ! 26FA , from CG2R61 SG3O2 NG311, penalty= 0.6
NG301 SG3O2 OG2P1 75.00 110.50 ! 26FA , from NG311 SG3O2 OG2P1, penalty= 0.6
DIHEDRALS
NG2S1 CG2O1 CG311 NG301 0.6000 1 0.00 ! 26FA , from NG2S1 CG2O1 CG314 NG3P3, penalty= 37.9
OG2D1 CG2O1 CG311 NG301 0.0000 1 0.00 ! 26FA , from OG2D1 CG2O1 CG314 NG3P3, penalty= 37.9
CG2R64 CG2R61 CG2R61 OG301 3.1000 2 180.00 ! 26FA , from CG2R61 CG2R61 CG2R61 OG301, penalty= 3
CG2R61 CG2R61 CG2R64 NG301 3.1000 2 180.00 ! 26FA , from CG2R61 CG2R61 CG2R61 NG311, penalty= 15.6
HGR61 CG2R61 CG2R64 NG2R62 2.8000 2 180.00 ! 26FA , from HGR61 CG2R61 CG2R64 NG2R60, penalty= 1
HGR61 CG2R61 CG2R64 NG301 2.4000 2 180.00 ! 26FA , from NG311 CG2R61 CG2R61 HGR61, penalty= 15.6
CG2R61 CG2R61 CG321 NG2S1 0.2300 2 180.00 ! 26FA , from CG2R61 CG2R61 CG321 CG2R61, penalty= 47
OG301 CG2R61 NG2R62 CG2R64 3.1000 2 180.00 ! 26FA , from CG331 CG2R61 NG2R60 CG2R64, penalty= 55.9
CG2R61 CG2R61 OG301 CG302 1.6200 2 180.00 ! 26FA , from CG2R61 CG2R61 OG301 CG321, penalty= 12
CG2R61 CG2R61 OG301 CG302 0.1900 4 180.00 ! 26FA , from CG2R61 CG2R61 OG301 CG321, penalty= 12
NG2R62 CG2R61 OG301 CG331 1.7400 2 180.00 ! 26FA , from CG2R61 CG2R61 OG301 CG331, penalty= 45
CG2R61 CG2R61 SG3O2 NG301 0.2200 2 0.00 ! 26FA , from CG2R61 CG2R61 SG3O2 NG311, penalty= 0.6
NG301 CG2R64 NG2R62 CG2R64 3.1000 2 180.00 ! 26FA , from CG331 CG2R61 NG2R60 CG2R64, penalty= 64.9
OG301 CG2R64 NG2R62 CG2R61 3.1000 2 180.00 ! 26FA , from CG321 CG2R61 NG2R60 CG2R61, penalty= 70
OG301 CG2R64 NG2R62 CG2R64 1.8000 2 180.00 ! 26FA , from NG2S3 CG2R64 NG2R62 CG2R64, penalty= 68
CG2R61 CG2R64 NG301 CG321 1.2000 2 180.00 ! 26FA , from CG2R61 CG2R61 NG311 CG321, penalty= 20
CG2R61 CG2R64 NG301 CG321 0.3000 4 0.00 ! 26FA , from CG2R61 CG2R61 NG311 CG321, penalty= 20
NG2R62 CG2R64 NG301 CG321 1.2000 2 180.00 ! 26FA , from CG2R61 CG2R61 NG311 CG321, penalty= 65
NG2R62 CG2R64 NG301 CG321 0.3000 4 0.00 ! 26FA , from CG2R61 CG2R61 NG311 CG321, penalty= 65
NG2R62 CG2R64 OG301 CG331 1.7400 2 180.00 ! 26FA , from CG2R61 CG2R61 OG301 CG331, penalty= 60
FGA3 CG302 OG301 CG2R61 0.2400 1 0.00 ! 26FA , from CG321 CG321 OG301 CG2R61, penalty= 98

FGA3 CG302 OG301 CG2R61	0.2900	2	0.00 ! 26FA , from CG321 CG321 OG301 CG2R61, penalty= 98
FGA3 CG302 OG301 CG2R61	0.0200	3	0.00 ! 26FA , from CG321 CG321 OG301 CG2R61, penalty= 98
CG2O1 CG311 CG321 NG301	0.2000	3	0.00 ! 26FA , from CG2O1 CG311 CG321 OG311, penalty= 38
NG301 CG311 CG321 NG301	0.2000	3	0.00 ! 26FA , from OG302 CG311 CG321 OG302, penalty= 67
NG301 CG311 CG321 HGA2	0.1600	3	0.00 ! 26FA , from NG311 CG321 CG331 HGA3, penalty= 10.6
HGA1 CG311 CG321 NG301	0.1600	3	0.00 ! 26FA , from NG311 CG321 CG331 HGA3, penalty= 11.6
CG2O1 CG311 NG301 CG321	2.5000	1	180.00 ! 26FA , from NG311 CG321 NG311 CG2R61, penalty= 150.5
CG2O1 CG311 NG301 CG321	1.5000	2	0.00 ! 26FA , from NG311 CG321 NG311 CG2R61, penalty= 150.5
CG2O1 CG311 NG301 CG321	0.5000	3	0.00 ! 26FA , from NG311 CG321 NG311 CG2R61, penalty= 150.5
CG2O1 CG311 NG301 SG3O2	1.4000	1	180.00 ! 26FA , from NG311 CG321 NG311 SG3O2, penalty= 77
CG2O1 CG311 NG301 SG3O2	0.5000	2	0.00 ! 26FA , from NG311 CG321 NG311 SG3O2, penalty= 77
CG2O1 CG311 NG301 SG3O2	0.1000	3	0.00 ! 26FA , from NG311 CG321 NG311 SG3O2, penalty= 77
CG321 CG311 NG301 CG321	0.1000	3	0.00 ! 26FA , from CG321 CG314 NG3P2 CG324, penalty= 121.5
CG321 CG311 NG301 SG3O2	0.1000	1	0.00 ! 26FA , from CG331 CG321 NG311 SG3O2, penalty= 9.9
CG321 CG311 NG301 SG3O2	0.7000	2	0.00 ! 26FA , from CG331 CG321 NG311 SG3O2, penalty= 9.9
CG321 CG311 NG301 SG3O2	0.1000	3	0.00 ! 26FA , from CG331 CG321 NG311 SG3O2, penalty= 9.9
HGA1 CG311 NG301 CG321	0.0000	3	180.00 ! 26FA , from HGA2 CG321 NG311 CG2R61, penalty= 82.5
HGA1 CG311 NG301 SG3O2	0.1000	3	0.00 ! 26FA , from HGA2 CG321 NG311 SG3O2, penalty= 9
NG301 CG321 CG321 NG301	0.2500	1	180.00 ! 26FA , from OG301 CG321 CG321 OG301, penalty= 66
NG301 CG321 CG321 NG301	1.2400	2	0.00 ! 26FA , from OG301 CG321 CG321 OG301, penalty= 66
NG301 CG321 CG321 HGA2	0.1600	3	0.00 ! 26FA , from NG311 CG321 CG331 HGA3, penalty= 6.6
CG2R61 CG321 NG2S1 CG2O1	0.2000	1	180.00 ! 26FA , from CG2O2 CG321 NG2S1 CG2O1, penalty= 28.5
CG2R61 CG321 NG2S1 HGP1	0.0000	1	0.00 ! 26FA , from CG2O2 CG321 NG2S1 HGP1, penalty= 28.5
CG311 CG321 NG301 CG2R64	2.5000	1	180.00 ! 26FA , from NG311 CG321 NG311 CG2R61, penalty= 47.5
CG311 CG321 NG301 CG2R64	1.5000	2	0.00 ! 26FA , from NG311 CG321 NG311 CG2R61, penalty= 47.5

CG311 CG321 NG301 CG2R64	0.5000	3	0.00 ! 26FA , from NG311 CG321 NG311 CG2R61, penalty= 47.5
CG311 CG321 NG301 CG321	2.5000	1	180.00 ! 26FA , from NG311 CG321 NG311 CG2R61, penalty= 118
CG311 CG321 NG301 CG321	1.5000	2	0.00 ! 26FA , from NG311 CG321 NG311 CG2R61, penalty= 118
CG311 CG321 NG301 CG321	0.5000	3	0.00 ! 26FA , from NG311 CG321 NG311 CG2R61, penalty= 118
CG321 CG321 NG301 CG2R64	2.5000	1	180.00 ! 26FA , from NG311 CG321 NG311 CG2R61, penalty= 47.5
CG321 CG321 NG301 CG2R64	1.5000	2	0.00 ! 26FA , from NG311 CG321 NG311 CG2R61, penalty= 47.5
CG321 CG321 NG301 CG2R64	0.5000	3	0.00 ! 26FA , from NG311 CG321 NG311 CG2R61, penalty= 47.5
CG321 CG321 NG301 CG311	2.5000	1	180.00 ! 26FA , from NG311 CG321 NG311 CG2R61, penalty= 118
CG321 CG321 NG301 CG311	1.5000	2	0.00 ! 26FA , from NG311 CG321 NG311 CG2R61, penalty= 118
CG321 CG321 NG301 CG311	0.5000	3	0.00 ! 26FA , from NG311 CG321 NG311 CG2R61, penalty= 118
CG321 CG321 NG301 CG321	2.5000	1	180.00 ! 26FA , from NG311 CG321 NG311 CG2R61, penalty= 118
CG321 CG321 NG301 CG321	1.5000	2	0.00 ! 26FA , from NG311 CG321 NG311 CG2R61, penalty= 118
CG321 CG321 NG301 CG321	0.5000	3	0.00 ! 26FA , from NG311 CG321 NG311 CG2R61, penalty= 118
CG321 CG321 NG301 SG3O2	0.1000	1	0.00 ! 26FA , from CG331 CG321 NG311 SG3O2, penalty= 5.9
CG321 CG321 NG301 SG3O2	0.7000	2	0.00 ! 26FA , from CG331 CG321 NG311 SG3O2, penalty= 5.9
CG321 CG321 NG301 SG3O2	0.1000	3	0.00 ! 26FA , from CG331 CG321 NG311 SG3O2, penalty= 5.9
HGA2 CG321 NG301 CG2R64	0.0000	3	180.00 ! 26FA , from HGA2 CG321 NG311 CG2R61, penalty= 8
HGA2 CG321 NG301 CG311	0.0000	3	180.00 ! 26FA , from HGA2 CG321 NG311 CG2R61, penalty= 78.5
HGA2 CG321 NG301 CG321	0.0000	3	180.00 ! 26FA , from HGA2 CG321 NG311 CG2R61, penalty= 78.5
HGA2 CG321 NG301 SG3O2	0.1000	3	0.00 ! 26FA , from HGA2 CG321 NG311 SG3O2, penalty= 5
HGA3 CG331 OG301 CG2R64	0.0850	3	0.00 ! 26FA , from HGA3 CG331 OG301 CG2R61, penalty= 3
CG311 NG301 SG3O2 CG2R61	1.5000	2	0.00 ! 26FA , from CG321 NG311 SG3O2 CG2R61, penalty= 5.6
CG311 NG301 SG3O2 CG2R61	0.5000	3	0.00 ! 26FA , from CG321 NG311 SG3O2 CG2R61, penalty= 5.6
CG311 NG301 SG3O2 OG2P1	0.2000	3	0.00 ! 26FA , from CG321 NG311 SG3O2 OG2P1, penalty= 5.6
CG321 NG301 SG3O2 CG2R61	1.5000	2	0.00 ! 26FA , from CG321 NG311 SG3O2 CG2R61, penalty= 5

CG321 NG301 SG3O2 CG2R61 0.5000 3 0.00 ! 26FA , from CG321 NG311 SG3O2 CG2R61, penalty= 5
CG321 NG301 SG3O2 OG2P1 0.2000 3 0.00 ! 26FA , from CG321 NG311 SG3O2 OG2P1, penalty= 5

Table A2.4: Ligand Parameters for the 23E Molecule

GROUP		!CHARGE
ATOM N1	NG2S1	-0.520
ATOM C2	CG3C50	0.064
ATOM C3	CG2O1	0.556
ATOM N4	NG2S1	-0.486
ATOM C5	CG2R61	0.132
ATOM O6	OG2D1	-0.430
ATOM C7	CG2R61	-0.120
ATOM C8	CG2R61	-0.109
ATOM C9	CG2R61	0.001
ATOM C10	CG2R61	-0.109
ATOM C11	CG2R61	-0.120
ATOM C12	CG2DC1	-0.149
ATOM C13	CG2DC1	-0.069
ATOM C14	CG2O2	0.638
ATOM O15	OG2D1	-0.540
ATOM C16	CG3C52	-0.172
ATOM C17	CG3C52	-0.172
ATOM C18	CG3C52	-0.180
ATOM C19	CG3C52	-0.180
ATOM O20	OG311	-0.600
ATOM C21	CG2O1	0.394
ATOM C22	CG2R61	-0.108
ATOM C23	CG2R61	-0.186
ATOM C24	CG2R61	-0.241
ATOM C25	CG2R61	-0.247
ATOM N26	NG2R51	-0.321
ATOM C27	CG2RC0	0.434
ATOM C28	CG2RC0	0.123
ATOM C29	CG2R51	-0.050
ATOM C30	CG2R51	-0.062
ATOM C31	CG311	-0.082
ATOM C32	CG2R61	0.153
ATOM C33	CG321	-0.191
ATOM C34	CG321	-0.196
ATOM C35	CG321	-0.146
ATOM C36	CG321	-0.196
ATOM C37	CG321	-0.191
ATOM O38	OG2D1	-0.426
ATOM C39	CG2R61	-0.110

ATOM C40 CG2R61 -0.115
ATOM C41 CG2R61 -0.109
ATOM C42 CG2R61 -0.120
ATOM C43 CG2R61 0.031
ATOM C44 CG321 -0.125
ATOM N45 NG2S1 -0.391
ATOM C46 CG2O1 0.499
ATOM O47 OG2D1 -0.484
ATOM H48 HGP1 0.240
ATOM H49 HGP1 0.335
ATOM H50 HGA4 0.150
ATOM H51 HGA4 0.150
ATOM H52 HGA2 0.090
ATOM H53 HGA2 0.090
ATOM H54 HGA2 0.090
ATOM H55 HGA2 0.090
ATOM H56 HGA2 0.090
ATOM H57 HGA2 0.090
ATOM H58 HGA2 0.090
ATOM H59 HGA2 0.090
ATOM H60 HGP1 0.426
ATOM H61 HGA1 0.090
ATOM H62 HGA2 0.090
ATOM H63 HGA2 0.090
ATOM H64 HGA2 0.090
ATOM H65 HGA2 0.090
ATOM H66 HGA2 0.090
ATOM H67 HGA2 0.090
ATOM H68 HGA2 0.090
ATOM H69 HGA2 0.090
ATOM H70 HGA2 0.090
ATOM H71 HGA2 0.090
ATOM H72 HGA2 0.090
ATOM H73 HGA2 0.090
ATOM H74 HGP1 0.327
ATOM H75 HGR61 0.115
ATOM H76 HGR61 0.115
ATOM H77 HGR61 0.115
ATOM H78 HGR61 0.115
ATOM H79 HGR61 0.200
ATOM H80 HGR61 0.197
ATOM H81 HGR61 0.193
ATOM H82 HGR61 0.115
ATOM H83 HGR61 0.115
ATOM H84 HGR61 0.115
ATOM H85 HGR61 0.115
! Bond order
BOND N1 C21
BOND N1 C2

BOND N1 H48
BOND C2 C16
BOND C2 C3
BOND C2 C17
BOND C3 N4
BOND C3 O6
BOND N4 C5
BOND N4 H49
BOND C5 C11
BOND C5 C7
BOND C7 C8
BOND C7 H82
BOND C8 C9
BOND C8 H83
BOND C9 C10
BOND C9 C12
BOND C10 C11
BOND C10 H85
BOND C11 H84
BOND C12 C13
BOND C12 H50
BOND C13 C14
BOND C13 H51
BOND C14 O15
BOND C14 O20
BOND C16 C18
BOND C16 H52
BOND C16 H53
BOND C17 C19
BOND C17 H54
BOND C17 H55
BOND C18 C19
BOND C18 H56
BOND C18 H57
BOND C19 H58
BOND C19 H59
BOND O20 H60
BOND C21 C22
BOND C21 O38
BOND C22 C23
BOND C22 C25
BOND C23 C24
BOND C23 H80
BOND C24 C28
BOND C24 H79
BOND C25 C27
BOND C25 H81
BOND N26 C30

```
BOND N26  C44
BOND N26  C27
BOND C27  C28
BOND C28  C29
BOND C29  C31
BOND C29  C30
BOND C30  C32
BOND C31  C33
BOND C31  C37
BOND C31  H61
BOND C32  C39
BOND C32  C43
BOND C33  C34
BOND C33  H62
BOND C33  H63
BOND C34  C35
BOND C34  H64
BOND C34  H65
BOND C35  C36
BOND C35  H66
BOND C35  H67
BOND C36  C37
BOND C36  H68
BOND C36  H69
BOND C37  H70
BOND C37  H71
BOND C39  C40
BOND C39  H78
BOND C40  C41
BOND C40  H77
BOND C41  C42
BOND C41  H76
BOND C42  C43
BOND C42  H75
BOND C43  N45
BOND C44  C46
BOND C44  H72
BOND C44  H73
BOND N45  C46
BOND N45  H74
BOND C46  O47
IMPR C3    C2    N4    O6
IMPR C14   C13   O15   O20
IMPR C21   C22   N1    O38
IMPR C46   C44   N45   O47

END

```

read param card flex append
* Parameters generated by analogy by
* CHARMM General Force Field (CGenFF) program version 0.9.7 beta
*
! Penalties lower than 10 indicate the analogy is fair; penalties between 10
! and 50 mean some basic validation is recommended; penalties higher than
! 50 indicate poor analogy and mandate extensive validation/optimization.
BONDS
CG2DC1 CG2O2 300.00 1.4800 ! 23EA , from CG2DC2 CG2O5, penalty= 10
CG2O1 CG3C50 250.00 1.4900 ! 23EA , from CG2O1 CG3C51, penalty= 6
CG2R51 CG2R61 300.00 1.4750 ! 23EA , from CG2O1 CG2R61, penalty= 145
CG2R51 CG311 229.63 1.5000 ! 23EA , from CG2R51 CG321, penalty= 4
CG321 NG2R51 400.00 1.4580 ! 23EA , from CG331 NG2R51, penalty= 6
CG3C50 CG3C52 195.00 1.5180 ! 23EA , from CG3C51 CG3C52, penalty= 6
CG3C50 NG2S1 320.00 1.4340 ! 23EA , from CG3C51 NG2S0, penalty= 26
ANGLES
CG2DC1 CG2DC1 CG2O2 60.00 120.00 ! 23EA , from CG2DC1 CG2DC1 CG2O4, penalty= 1.5
CG2DC1 CG2DC1 CG2R61 29.00 122.00 ! 23EA , from CG2DC3 CG2DC1 CG2R61, penalty= 3.5
CG2O2 CG2DC1 HGA4 32.00 122.00 ! 23EA , from CG2O4 CG2DC2 HGA4, penalty= 1.5
CG3C50 CG2O1 NG2S1 80.00 116.50 ! 23EA , from CG3C51 CG2O1 NG2S1, penalty= 0.8
CG3C50 CG2O1 OG2D1 80.00 118.00 ! 23EA , from CG3C51 CG2O1 OG2D1, penalty= 0.8
CG2DC1 CG2O2 OG2D1 70.00 121.80 ! 23EA , from CG2DC2 CG2O5 OG2D3, penalty= 10.5
CG2DC1 CG2O2 OG311 40.00 113.90 30.00 2.37000 ! 23EA , from CG2R61 CG2O2 OG311, penalty= 21
CG2R51 CG2R51 CG2R61 45.80 130.00 ! 23EA , from CG2R51 CG2R51 CG321, penalty= 71
CG2R51 CG2R51 CG311 45.80 130.00 ! 23EA , from CG2R51 CG2R51 CG321, penalty= 0.6
CG2R61 CG2R51 NG2R51 45.80 124.00 ! 23EA , from CG321 CG2R51 NG2R51, penalty= 71
CG2RC0 CG2R51 CG311 30.00 126.70 ! 23EA , from CG2RC0 CG2R51 CG321, penalty= 0.6
CG2R51 CG2R61 CG2R61 36.00 120.00 ! 23EA , from CG2DC1 CG2R61 CG2R61, penalty= 22.5
CG2R51 CG311 CG321 58.35 114.00 ! 23EA , from CG2R51 CG321 CG311, penalty= 4.6
CG2R51 CG311 HGA1 55.00 109.50 ! 23EA , from CG2R51 CG321 HGA2, penalty= 4
CG2O1 CG321 NG2R51 50.00 107.00 ! 23EA , from CG2O1 CG321 NG2S1, penalty= 11

NG2R51 CG321 HGA2 33.43 110.10 ! 23EA , from NG2R51 CG331 HGA3, penalty= 6
CG2O1 CG3C50 CG3C52 52.00 112.30 ! 23EA , from CG2O1 CG3C51 CG3C52, penalty= 6
CG2O1 CG3C50 NG2S1 50.00 108.20 ! 23EA , from CG2O1 CG3C51 NG2S0, penalty= 8
CG3C52 CG3C50 CG3C52 58.00 109.50 11.16 2.56100 ! 23EA , from CG3C52 CG3C51 CG3C52, penalty= 6
CG3C52 CG3C50 NG2S1 70.00 110.80 ! 23EA , from CG3C52 CG3C51 NG2S0, penalty= 8
CG3C50 CG3C52 CG3C52 58.00 109.50 11.16 2.56100 ! 23EA , from CG3C51 CG3C52 CG3C52, penalty= 0.8
CG3C50 CG3C52 HGA2 35.00 111.40 22.53 2.17900 ! 23EA , from CG3C51 CG3C52 HGA2, penalty= 0.8
CG2R51 NG2R51 CG321 70.00 127.80 ! 23EA , from CG2R53 NG2R51 CG331, penalty= 3.9
CG2RC0 NG2R51 CG321 70.00 125.90 ! 23EA , from CG2RC0 NG2R51 CG331, penalty= 0.9
CG2O1 NG2S1 CG3C50 50.00 120.00 ! 23EA , from CG2O1 NG2S1 CG321, penalty= 10
CG3C50 NG2S1 HGP1 35.00 117.00 ! 23EA , from CG321 NG2S1 HGP1, penalty= 10
DIHEDRALS
CG2O2 CG2DC1 CG2DC1 CG2R61 0.5600 1 180.00 ! 23EA , from CG2DC2 CG2DC1 CG2DC1 CG2O4, penalty= 24
CG2O2 CG2DC1 CG2DC1 CG2R61 7.0000 2 180.00 ! 23EA , from CG2DC2 CG2DC1 CG2DC1 CG2O4, penalty= 24
CG2O2 CG2DC1 CG2DC1 HGA4 5.2000 2 180.00 ! 23EA , from CG2DC2 CG2DC1 CG2DC1 HGA4, penalty= 28.5
CG2R61 CG2DC1 CG2DC1 HGA4 5.2000 2 180.00 ! 23EA , from CG2DC2 CG2DC1 CG2DC1 HGA4, penalty= 22.5
CG2DC1 CG2DC1 CG2O2 OG2D1 1.4000 2 180.00 ! 23EA , from CG2DC3 CG2DC1 CG2O5 OG2D3, penalty= 14
CG2DC1 CG2DC1 CG2O2 OG311 1.4000 2 180.00 ! 23EA , from CG2DC3 CG2DC1 CG2O5 CG331, penalty= 59.4
HGA4 CG2DC1 CG2O2 OG2D1 0.0000 2 180.00 ! 23EA , from HGA4 CG2DC1 CG2O5 OG2D3, penalty= 10.5
HGA4 CG2DC1 CG2O2 OG311 0.0000 2 180.00 ! 23EA , from HGA4 CG2DC1 CG2O5 CG331, penalty= 55.9
CG2DC1 CG2DC1 CG2R61 CG2R61 0.7500 2 180.00 ! 23EA , from CG2DC3 CG2DC1 CG2R61 CG2R61, penalty= 3.5
CG2DC1 CG2DC1 CG2R61 CG2R61 0.1900 4 0.00 ! 23EA , from CG2DC3 CG2DC1 CG2R61 CG2R61, penalty= 3.5
NG2S1 CG2O1 CG321 NG2R51 0.6000 1 0.00 ! 23EA , from NG2S1 CG2O1 CG321 NG2S1, penalty= 11
OG2D1 CG2O1 CG321 NG2R51 0.0000 1 0.00 ! 23EA , from OG2D1 CG2O1 CG321 NG2S1, penalty= 11
NG2S1 CG2O1 CG3C50 CG3C52 0.4000 1 0.00 ! 23EA , from NG2S1 CG2O1 CG3C51 CG3C52, penalty= 6
NG2S1 CG2O1 CG3C50 CG3C52 0.6000 2 0.00 ! 23EA , from NG2S1 CG2O1 CG3C51 CG3C52, penalty= 6

NG2S1 CG2O1 CG3C50 NG2S1 0.3000 1 0.00 ! 23EA , from NG2S1 CG2O1 CG3C51 NG2S0, penalty= 8
NG2S1 CG2O1 CG3C50 NG2S1 -0.3000 4 0.00 ! 23EA , from NG2S1 CG2O1 CG3C51 NG2S0, penalty= 8
OG2D1 CG2O1 CG3C50 CG3C52 0.4000 1 180.00 ! 23EA , from OG2D1 CG2O1 CG3C51 CG3C52, penalty= 6
OG2D1 CG2O1 CG3C50 CG3C52 0.6000 2 0.00 ! 23EA , from OG2D1 CG2O1 CG3C51 CG3C52, penalty= 6
OG2D1 CG2O1 CG3C50 NG2S1 -0.3000 4 0.00 ! 23EA , from OG2D1 CG2O1 CG3C51 NG2S0, penalty= 8
CG2R61 CG2O1 NG2S1 CG3C50 1.6000 1 0.00 ! 23EA , from CG2R61 CG2O1 NG2S1 CG321, penalty= 10
CG2R61 CG2O1 NG2S1 CG3C50 4.0000 2 180.00 ! 23EA , from CG2R61 CG2O1 NG2S1 CG321, penalty= 10
CG3C50 CG2O1 NG2S1 CG2R61 1.6000 1 0.00 ! 23EA , from CG321 CG2O1 NG2S1 CG2R61, penalty= 10
CG3C50 CG2O1 NG2S1 CG2R61 2.5000 2 180.00 ! 23EA , from CG321 CG2O1 NG2S1 CG2R61, penalty= 10
CG3C50 CG2O1 NG2S1 HGP1 2.5000 2 180.00 ! 23EA , from CG3C51 CG2O1 NG2S1 HGP1, penalty= 0.8
OG2D1 CG2O1 NG2S1 CG3C50 2.5000 2 180.00 ! 23EA , from OG2D1 CG2O1 NG2S1 CG321, penalty= 10
CG2DC1 CG2O2 OG311 HGP1 0.9750 1 180.00 ! 23EA , from CG2R61 CG2O2 OG311 HGP1, penalty= 21
CG2DC1 CG2O2 OG311 HGP1 2.7000 2 180.00 ! 23EA , from CG2R61 CG2O2 OG311 HGP1, penalty= 21
CG2DC1 CG2O2 OG311 HGP1 0.0500 3 180.00 ! 23EA , from CG2R61 CG2O2 OG311 HGP1, penalty= 21
CG2DC1 CG2O2 OG311 HGP1 0.2500 6 180.00 ! 23EA , from CG2R61 CG2O2 OG311 HGP1, penalty= 21
CG2R61 CG2R51 CG2R51 CG2RC0 2.0000 2 180.00 ! 23EA , from CG2R51 CG2R51 CG2R51 CG2RC0, penalty= 48.5
CG2R61 CG2R51 CG2R51 CG311 6.9000 2 180.00 ! 23EA , from CG2RC0 CG2R51 CG2R51 CG3C52, penalty= 78.5
CG311 CG2R51 CG2R51 NG2R51 3.0000 2 180.00 ! 23EA , from CG321 CG2R51 CG2R51 NG2R51, penalty= 0.6
CG2R51 CG2R51 CG2R61 CG2R61 1.0000 2 180.00 ! 23EA , from NG2S1 CG2O1 CG2R61 CG2R61, penalty= 228
NG2R51 CG2R51 CG2R61 CG2R61 1.0000 2 180.00 ! 23EA , from NG2S1 CG2O1 CG2R61 CG2R61, penalty= 196
CG311 CG2R51 CG2RC0 CG2R61 2.5000 2 180.00 ! 23EA , from CG321 CG2R51 CG2RC0 CG2R61, penalty= 0.6
CG311 CG2R51 CG2RC0 CG2RC0 3.0000 2 180.00 ! 23EA , from CG321 CG2R51 CG2RC0 CG2RC0, penalty= 0.6
CG2R51 CG2R51 CG311 CG321 0.2000 1 0.00 ! 23EA , from CG2R51 CG2R51 CG321 CG311, penalty= 4.6
CG2R51 CG2R51 CG311 CG321 0.2700 2 0.00 ! 23EA , from CG2R51 CG2R51 CG321 CG311, penalty= 4.6
CG2R51 CG2R51 CG311 CG321 0.0000 3 0.00 ! 23EA , from CG2R51 CG2R51

CG321 CG311, penalty= 4.6
CG2R51 CG2R51 CG311 HGA1 0.0000 3 0.00 ! 23EA , from CG2R51 CG2R51 CG321 HGA2, penalty= 4
CG2RC0 CG2R51 CG311 CG321 0.0900 2 180.00 ! 23EA , from CG2RC0 CG2R51 CG321 CG311, penalty= 4.6
CG2RC0 CG2R51 CG311 CG321 0.5700 3 0.00 ! 23EA , from CG2RC0 CG2R51 CG321 CG311, penalty= 4.6
CG2RC0 CG2R51 CG311 HGA1 0.2000 3 0.00 ! 23EA , from CG2RC0 CG2R51 CG321 HGA2, penalty= 4
CG2R51 CG2R51 NG2R51 CG321 0.0000 1 0.00 ! 23EA , from CG2R51 CG2R51 NG2R51 CG3C51, penalty= 31.4
CG2R61 CG2R51 NG2R51 CG2RC0 6.0000 2 180.00 ! 23EA , from CG2RC0 CG2R51 NG2R51 CG2R51, penalty= 48.5
CG2R61 CG2R51 NG2R51 CG321 0.0000 1 0.00 ! 23EA , from CG2R51 CG2R51 NG2R51 CG3C51, penalty= 80.4
CG2O1 CG2R61 CG2R61 CG2RC0 3.1000 2 180.00 ! 23EA , from CG2O1 CG2R61 CG2R61 CG2R61, penalty= 1.5
CG2R51 CG2R61 CG2R61 CG2R61 3.1000 2 180.00 ! 23EA , from CG2DC1 CG2R61 CG2R61 CG2R61, penalty= 22.5
CG2R51 CG2R61 CG2R61 NG2S1 3.1000 2 180.00 ! 23EA , from CG2R61 CG2R61 CG2R61 NG2S1, penalty= 28.5
CG2R51 CG2R61 CG2R61 HGR61 2.4000 2 180.00 ! 23EA , from CG2DC1 CG2R61 CG2R61 HGR61, penalty= 22.5
CG2R61 CG2RC0 NG2R51 CG321 11.0000 2 180.00 ! 23EA , from CG2RC0 CG2RC0 NG2R51 CG331, penalty= 48.4
CG2RC0 CG2RC0 NG2R51 CG321 11.0000 2 180.00 ! 23EA , from CG2RC0 CG2RC0 NG2R51 CG331, penalty= 0.9
CG2R51 CG311 CG321 CG321 0.0400 3 0.00 ! 23EA , from CG2R61 CG321 CG321 CG321, penalty= 12.5
CG2R51 CG311 CG321 HGA2 0.2000 3 0.00 ! 23EA , from HGA1 CG311 CG321 CG2R51, penalty= 8.1
CG2O1 CG321 NG2R51 CG2R51 0.0000 1 180.00 ! 23EA , from CG2O2 CG311 NG2R53 CG2R53, penalty= 68.5
CG2O1 CG321 NG2R51 CG2RC0 0.0000 1 180.00 ! 23EA , from CG2O2 CG311 NG2R53 CG2R53, penalty= 69.5
HGA2 CG321 NG2R51 CG2R51 0.0000 3 0.00 ! 23EA , from HGA3 CG331 NG2R51 CG2R53, penalty= 9
HGA2 CG321 NG2R51 CG2RC0 0.1900 3 0.00 ! 23EA , from HGA3 CG331 NG2R51 CG2RC0, penalty= 6
CG2O1 CG3C50 CG3C52 CG3C52 0.1400 3 0.00 ! 23EA , from CG2O1 CG3C51 CG3C52 CG3C52, penalty= 6
CG2O1 CG3C50 CG3C52 HGA2 0.1400 3 0.00 ! 23EA , from CG2O1 CG3C51 CG3C52 HGA2, penalty= 6
CG3C52 CG3C50 CG3C52 CG3C52 0.4100 3 180.00 ! 23EA , from CG3C52 CG3C51 CG3C52 CG3C52, penalty= 6
CG3C52 CG3C50 CG3C52 HGA2 0.1900 3 0.00 ! 23EA , from CG3C52 CG3C51 CG3C52 HGA2, penalty= 6
NG2S1 CG3C50 CG3C52 CG3C52 0.1400 3 0.00 ! 23EA , from NG2S0 CG3C51 CG3C52 CG3C52, penalty= 8

NG2S1 CG3C50 CG3C52 HGA2 0.1400 3 0.00 ! 23EA , from NG2S0 CG3C51 CG3C52 HGA2, penalty= 8
CG2O1 CG3C50 NG2S1 CG2O1 0.8000 3 0.00 ! 23EA , from CG2O1 CG3C51 NG2S0 CG2O1, penalty= 26
CG2O1 CG3C50 NG2S1 HGP1 0.0000 1 0.00 ! 23EA , from CG2O1 CG321 NG2S1 HGP1, penalty= 65
CG3C52 CG3C50 NG2S1 CG2O1 0.8000 3 0.00 ! 23EA , from CG3C52 CG3C51 NG2S0 CG2O1, penalty= 26
CG3C52 CG3C50 NG2S1 HGP1 0.3000 1 0.00 ! 23EA , from CG3C52 CG3C51 NG2S3 HGP1, penalty= 46
CG3C50 CG3C52 CG3C52 CG3C52 0.4100 3 180.00 ! 23EA , from CG3C51 CG3C52 CG3C52 CG3C52, penalty= 0.8
CG3C50 CG3C52 CG3C52 HGA2 0.1900 3 0.00 ! 23EA , from CG3C51 CG3C52 CG3C52 HGA2, penalty= 0.8
IMPROPERS
CG2O1 CG3C50 NG2S1 OG2D1 120.0000 0 0.00 ! 23EA , from CG2O1 CG3C51 NG2S1 OG2D1, penalty= 0.6
CG2O2 CG2DC1 OG2D1 OG311 53.0000 0 0.00 ! 23EA , from CG2O2 CG2R61 OG2D1 OG311, penalty= 14

Appendix B

Table B3.1: Cross Comparisons of Predicted Sites From LISE and Each Target Site.

Known Binding Site	LISE Percent Overlap of Predicted Sites for Each Query Structure		
	2BRL	2WHO	3CO9
NNI-1	81%	<50%	<50%
NNI-2	66%	66%	75%
NNI-3	56%	56%	63%

Table B3.2: Cross Comparisons of Predicted Sites From Qsite Finder and Each Target Site.

Known Binding Site	Qsite Finder Percent Overlap of Predicted Sites for Each Query Structure		
	2BRL	2WHO	3CO9
NNI-1	88%	<50%	<50%
NNI-2	<50%	75%	66%
NNI-3	56%	56%	<50%

Table B3.3: Cross Comparisons of Predicted Sites From FTSite and Each Target Site.

Known Binding Site:	FTSite Percent Overlap of Top Predicted Sites for Each Query Structure		
	2BRL	2WHO	3CO9
NNI-1	<50%	<50%	<50%
NNI-2	<50%	<50%	<50%
NNI-3	<50%	50%	50%

Table B3.4: Cross Comparisons of Predicted Sites From LIGSITE^{csc} and Each Target Site.

Known Binding Site	LIGSITE^{csc} Percent Overlap of Predicted Sites for Each Query Structure		
	2BRL	2WHO	3CO9
NNI-1	81%	<50%	<50%
NNI-2	66%	75%	50%
NNI-3	<50%	<50%	<50%