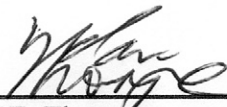


APPROVAL SHEET

Title of Dissertation: Using Molecular Simulations to Explore Conformational Transitions in the Hepatitis C Virus Polymerase.

Name of Candidate: Ester Sesmero Diaz
Doctor of Philosophy, 2017

Dissertation and Abstract Approved: _____


Dr. Ian F. Thorpe
Assistant Professor
Department of Chemistry & Biochemistry

Date Approved: 4-24-2017

ABSTRACT

Title of Document:

**USING MOLECULAR SIMULATIONS TO
EXPLORE CONFORMATIONAL
TRANSITIONS IN THE HEPATITIS C
VIRUS POLYMERASE.**

Ester Sesmero Diaz, Ph.D., 2017

Directed by:

Dr. Ian F. Thorpe, Department of Chemistry &
Biochemistry

Hepatitis C virus (HCV) is a global health concern for which there is no vaccine available. The HCV polymerase is responsible for the critical function of replicating the RNA genome of the virus. Transitions between its conformations are necessary for the appropriate functioning of the enzyme in the replication of the enzyme's RNA. We have shown that one can generate the free energy landscape (FEL) of the free enzyme by combining the results from Molecular Dynamic (MD) simulations initiated from a collection of X-ray structures of NS5B without inhibitor bound. MD is a suitable tool for this purpose since it provides structural and dynamic information regarding the conformational space sampled by the enzyme. From these results, we have elucidated key interactions that stabilize each conformation and that modulate the transitions between them. This new knowledge is essential to gain insight on the role of these conformations in replication and may be also valuable in suggesting novel targets for inhibition of the enzyme. Furthermore, by comparing these observations with the results of MD simulations of inhibitor bound systems, we gained a deeper understanding on the allosteric mechanisms of currently reported inhibitors. As a result, the knowledge garnered from this computational study may be instrumental for improving HCV therapies and those of other related viruses.

USING MOLECULAR SIMULATIONS TO EXPLORE CONFORMATIONAL
TRANSITIONS IN THE HEPATITIS C VIRUS POLYMERASE.

By

Ester Sesmero Diaz

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
2017

Acknowledgements

In first place I would like to THANK GOD. As Jodian and I have said many times without Him we would have not endured. I would like to THANK Our Lord Jesus Christ for sustaining me every moment through this journey. I would have not been able to reach this point without His assistance and the strength He has given me to persevere. There is a passage from Scripture that has come to my mind often over the last couple of years is “those who sow in tears will reap in joy” (Psalm 126:5). THANK YOU, Lord, for being at my side in the moments of tears and accompanying me all the way to this moment of JOY!

As I have attested throughout my grad school journey a big component to getting to finish your dissertation is having a good supportive network. I have not gotten here alone or solely by my own efforts, there has been an enormous number of people who have lent me a hand throughout this journey and I’m AMAZINGLY GRATEFUL to all of them!. There have been so many people helping me that it would be impossible for me to name them all but I want to try to highlight a few.

I would like to thank my mentor Dr. Ian Thorpe and the Thorpe lab. It has been very special to grow together and to learn from each other throughout these years. Dr. Thorpe, I will always admire and remember your open-minded approach to research and life in general and your thirst for learning new things and always striving for improvement. A special THANK YOU to Jodian Brown, Brittny Davis, Nicole Carbonaro and to our students Marie Espiritu and Ye Jin Bae. Jodian, thank you for being always there for me. As I told you many times you were like my “co-mentor”. Thank you for always have the “unsolicited” advise for me. Thank you for reviewing my papers, for thinking with me through my code errors, my proposals, my papers... I will always value and admire your honest and authentic way to live your faith and your life. I am grateful to be considered your friend and to have had you at my side through this time.

I would also like to thank my Thesis Committee: Dr. Marie-Christine Daniel who accompanied me in my journey since I came and did my first rotation with her, Dr. Ivan Erill, Dr. Lisa Kelly and Dr. Ryan White. Thank you for your time and helpful insights on my project.

A very important source of support for me during this time has been my spiritual family, embodied mainly by the Crusaders of Mary, Catholic Retrievers, St. Peter’s Catholic Church and Casa Sacri Cuori. From Catholic Retrievers, I would like to THANK in a special way Fr. John Rapisarda. Thank you, Father, for your unconditional support. Thank you for showing me what unconditional fatherly love truly means. A very special thank you goes too to the Women’s group (Marie Espiritu, Nicole Racine, Allison Opitz, Allie Fenney, Natasha Wilson, Val Schultz, Rachel Morin, Anna Haley, Elizabeth Chavez, Greta Curley... and many more). Thank you, ladies, it has been wonderful having you as sisters in my journey, sharing our experiences and supporting each other!. From St. Peter’s, I would like to thank in a special way Susan Timoney and everyone in the Sunday School program (especially Caitlin Jolly, Kristina Wolf and Martin and April Waugh). Susan, thank you for always making me feel welcome, thank you for your unconditional support and for being an open forum for me to share anything I have in my heart. From the Casa Sacri Cuori, I would like to thank in a special way Sr. Dede Byrne, for her generous heart that always finds ways to serve others, Mary Vigil, for her patience and for being

always a source of understanding and peace among peoples and Martha McNeill for being such a loving presence, always a reminder that love is what is truly important in life.

Although from a farther distance, from Spain, my family according to the flesh, have always supported me and believed in me. I would like to thank my dad, Eugenio Sesmero Granados, my mom, Esperanza Diaz Cremades and my sister, Emma Sesmero Diaz. My dad always said “I’m a prophet, I know you’ll be able to do it”. THANK YOU for always believing in me.

I would also like to thank the Meyerhoff, Chemistry-Biology Interface and Promise programs of which I am part of. Thank you in a special way to Dr. Michael Summers and Justine Johnson, Dr. Katherine Seley-Radtke and Dr. Renetta Tull. It has been a pleasure and a source of growth for me to be part of those programs. THANK YOU for all you do for us!.

Here in our Chemistry and Biochemistry Department I would also like to say a special thank you to the faculty and staff. In special way I would like to thank Dr. Mark Perks, with whom I have TAed CHEM 351L for numerous years and with whom I share my passion for teaching, Frank Tyminsky, always a caring presence among us, and Patricia Gagne, whom I feel is the embodiment of kindness. I would like to thank the graduate students with whom I have shared lab, classes, TA assignments or any activities here and that we have supported each other in the struggle. I would like to specially thank Janae Baptiste, Miji Jeon and Denise Williams. Thank you all for being a kind hand and a friendly smile for me, for making me feel home and helping me in time of need.

I would also like to thank those who have helped me in my struggles with research: Dr. Patrick O’Neill for sharing his knowledge with me and for having the patience to explain it to me step by step. Christopher Porter, for his help with data analysis but in a very special way helping me in building my confidence that programming was not impossible for me to learn. Dr. Alexander Sodt for always answering my emails asking how to solve my errors and for his assistance in writing my Independent proposal, I truly enjoyed our scientific discussion, and it renewed in me the passion for research and learning. Dr. Charles Bieberich, always an encouraging presence in the science world, a reminder that is possible to be a successful scientist and a man of deep faith. Farhan Augustine, thank you for sharing your passion for learning with me and your humble approach to daily life, you opened my eyes to realities that were new and unexplored to me.

Although I say it last they are not least important, other friends that have supported me. I would like to thank the friends that I met in my commute and with whom I have shared the experiences I was going through during our train rides for all these years: Irene Chan, Alexandra St.Pe’, Nkiru Nnawulezi, Jennifer Hosler, Sonia Bansal and many others. Alex, thanks for always being so encouraging, I never left your side without feeling more confident and ready to try again.

And THANK YOU to the many more that I did not get to mention by name here, thank you to everyone that has helped me along the way, from the bottom of my heart THANK YOU!

Table of contents

Chapter 1: Introduction: Using the Hepatitis C Virus RNA-Dependent RNA Polymerase as a Model to Understand Viral Polymerase Structure, Function and Dynamics

1.1 Dissertation Outline	1
1.1.1 Specific Aims	2
1.1.2 Significance	2
1.2 Chapter Overview	3
1.3 Introduction	4
1.4 General Structural Features of Viral Polymerases	6
1.5 Conserved Structural Motifs of Viral Polymerases	8
1.6 Structural Features of RdRps	9
1.7 Catalytic Mechanism and Polymerase Reaction Steps	13
1.8 NS5B Conformational Changes during the Replication Cycle	19
1.9 NS5B Inhibitors and Mechanisms of Action	20
1.10 Summary	28
1.11 Acknowledgements	29
1.12 References	29

Chapter 2: Molecular Simulations to Delineate Functional Conformational Transitions in the HCV Polymerase

2.1 Introduction	35
2.2 Methods	36
2.2.1 Structure Preparation and Simulation Conditions	36
2.2.2 Metrics of Conformational Sampling	38
2.2.2.1 Inter-domain Angle	39
2.2.2.2 Principal Components Analysis (PCA)	39
2.2.3 Root Mean Squared Fluctuations	41
2.2.4 Free Energy Landscapes	42
2.2.5 Weighted Histogram Analysis Method (WHAM)	42
2.3 Results and Discussion	46
2.3.1 The free enzyme samples a broad conformational distribution	46
2.3.2 Conformational states exhibit distinct characteristics	50
2.3.2.1 Structural differences	51
2.3.2.2 Root Mean Squared Fluctuations	54
2.3.2.3 Covariance Analysis	55
2.3.2.4 Large amplitude motions	56
2.3.2.5 Hydrogen bonding patterns	57
2.3.3 Potential relevance for therapeutics	59
2.3.4 Insights into the RNA replication cycle	61
2.4 Summary	64
2.5 Acknowledgements	66
2.6 References	67

Chapter 3. Effects of structural perturbations on conformational sampling in the HCV polymerase

3.1 Introduction	70
3.2 Methods	72
3.2.1 Structure Preparation and Simulation Conditions	72
3.3 Results and Discussion	77
3.3.1 The FEL of the enzyme maintains its essential features despite the presence of perturbations	77
3.3.2 Inhibitors modify the Free energy landscape	82
3.3.3 Inhibitors disrupt communication between domains and preferentially stabilize specific conformations	82
3.3.4 Mutations destabilize enzyme conformations	83
3.4 Conclusions	84
3.5 Acknowledgements	85
3.6 References	86

Chapter 4: Summary, Significance and Future Directions

4.1 Dissertation Summary: Major Findings of this Dissertation	88
4.1.1 Characterization of the NS5B conformations and the transitions between them (Chapter 2)	89
4.1.2 The effect of mutations and inhibitors show that insights from NS5B may be transferable to other viral polymerases (Chapter 3)	90
4.2 Significance	91
4.3 Future Directions	92
4.4 Acknowledgements	94
4.5 References	94

List of tables

Table 1.1. Baltimore classification of viruses compared with the classification of viral polymerases based on their targeted genetic material.	9
Table 1.2. Characteristic sequence motifs in polymerases from prototypical viruses: hepatitis C virus (HCV), poliovirus (PV) and foot-and-mouth-disease virus (FMDV)	10
Table 1.3. RdRps virus families and species (1, 22).	11
Table 1.4. Genera and species of the Flaviviridae family.	13
Table 2.1. Simulated systems.	45
Table 3.1 Simulated systems.	76
Table A1: WHAM weights and number of snapshots in each substate for individual trajectories.	96
Table A2: Inter-domain angles computed for a random sample of HCV polymerase structures from the PDB.	98
Table A3. Characteristic hydrogen bonds in each conformational state	99
Table A4. Characteristic hydrogen bond networks in each conformational state	103
Table B1. Characteristic hydrogen bonds in each conformational state for the mutated systems. They are color coded following the same criteria that in appendix A.	107
Table B2. Characteristic hydrogen bonds in each conformational state for the inhibitor bound systems. They are color coded following the same criteria that in appendix A.	111
Table B3. RMSD comparison of the PDBs representative of the different minima.	114

List of figures

Figure 1.1. (a) Right-hand structure of HCV polymerase (NS5B). Palm, fingers and thumb domains are shown in red, blue and green, respectively; (b) Duplex channel in NS5B (front of the enzyme); (c) NTP channel in NS5B (back of the enzyme); (d) Motifs and functional regions of NS5B. Motif A in red, B in orange, C in yellow, D in bright green, E in pink, F in purple and G in cyan. Functional regions: I in light green, II in violet and III in tan; (e) Template channel (top view of the enzyme). 7

Figure 1.2. Schematic describing de novo initiation in Hepaciviruses and Pestiviruses. Note that Flaviviruses do not anchor their C-terminus in the Endoplasmic Reticulum (ER). This figure was generated by incorporating the descriptions provided by both Appleby *et al.* [35] and Choi [1]. The linker and C-terminal anchor are shown in orange as one contiguous element. The β -flap is colored red (as in Figure 1.3), the template strand in purple, the growing strand in green, the stabilizing GTP in blue and the Endoplasmic Reticulum (ER) in brown. The “N” and “P” indicate where the N-site (nucleotide-site) and P-site (priming-site) are. These correspond to the positions of the growing strand that bind to residues “ n ” and “ $n + 1$ ” of the template strand, respectively. 16

Figure 1.3. NS5B structure with characteristic elements highlighted. (a) Front view and (b) top view. The linker is shown in orange, the β -flap in red. The fingertips are shown in blue (the delta 1 loop) and green (the delta 2 loop). 17

Figure 1.4. Two metals ions mechanism in RdRps. The squares represent the bases that are part of the nucleotides. This figure is inspired by a similar figure from Choi *et al.* (1). 18

Figure 1.5. NS5B inhibitors. (a) The three allosteric sites of NS5B are highlighted with space filling representations of inhibitors that bind in these locations. Thumb site 1 (NNI-1) in yellow, thumb site 2 (NNI-2) in green, palm sites (NNI-3/4) in purple; (b) chemical structures of NIs and NNIs that are in clinical trials or have already been approved (52-59). 22

Figure 1.6. Mechanisms of inhibition for NNIs. NS5B must transition between open and closed states to perform replication (upper left). NNI-1 inhibitors have been observed to reduce enzyme stability. NNI-2 inhibitors have been shown to reduce conformational sampling, confining the enzyme in closed conformations. NNI-3 inhibitors mainly block access of the RNA template but also induce some restriction of conformational sampling. The RNA template is represented as a black rectangle and the inhibitor as an orange ellipse. 23

Figure 2.1. a) Structure of HCV polymerase showing the fingers, palm, and thumb subdomains colored blue, red and green respectively. The inter-domain angle formed at the intersection of the subdomains is shown in black. The location of the template channel is indicated by the translucent rectangle. b) Schematic describing the role of polymerase conformational states in the RNA replication cycle. 39

Figure 2.2. Free energy landscape obtained from combining all trajectories and projected unto the inter-domain angle (x-axis) and the largest amplitude PC from the 2BRL simulation (y-axis). Contours in the plot are displayed at intervals of $k_B T / 2$ 49

Figure 2.3. One-dimensional free energy landscapes obtained from combining all trajectories and projected unto a) the inter-domain angle and b) the largest amplitude PC from 2BRL (PC_{2BRL}). In both plots the left side corresponds to more closed states while the right corresponds to more open states. 50

Figure 2.4. Motifs, functional regions and characteristic elements of NS5B affected in the conformational transitions. Motif G is shown in green, functional region I in yellow, functional region II in blue, functional region III in pink, delta 1 loop in purple, delta 2 loop (which

overlaps with motif F) in orange, beta flap in red. (a) front view, (b) top view.(c) Zoomed image of regions within rectangle in a) showing relative positions of fingers and thumb subdomains in closed, intermediate and open conformations.	52
Figure 2.5. RMSF (left) and covariance (right) plots generated from conformational ensembles in the open, intermediate and closed states.	53
Figure 3.1. Location of mutations in the NS5B polymerase. In the figure we show inhibitors in the allosteric sites so that the locations of mutations with respect to the inhibitors are easily recognizable. NNI-1 is in yellow, NNI-2 in green and NNI-3 in purple.	77
Figure 3.2. 1-D (right) and 2-D (left) Free energy landscapes of free systems and ligand bound systems.	78
Figure 3.3. FEL of genotype 2a systems.	80
Figure A1. Free energy landscapes computed for individual trajectories.	95
Figure A2. Location of the original PDB coordinates with respect to free energy landscapes presented in the main text. Symbols corresponding to PDB IDs for each structure are provided in the legend of each panel. Note that all of the coordinates display conformations in the vicinity of the free energy minimum in the intermediate region of the landscapes ($69^\circ < \text{inter-domain angle} < 74^\circ$). It is possible that conditions amenable to crystallization may preferentially stabilize structures with these domain angle values.	97
Figure A3. Schematic descriptions of H-bond categories.	103
Figure A4. Schematic depictions of hydrogen bond networks in open and closed states. Text in parentheses refers to the enzyme domain or structural element in which a given residue is located as follows: fingers (f), delta 1 loop (d1), delta 2 loop (d2).	104
Figure A5. Molecular view of residues involved in H-bond network #1 that is schematically represented above. The polymerase is depicted in cartoon representation. The delta 1 and delta 2 loops are shown in purple and orange while residues in the fingers domain are shown in black. This extensive network occurs only in the closed conformation and involves twelve residues from the delta 1 loop, delta 2 loop and part of the adjoining fingers. The occurrence of this network emphasizes the key role of the delta 1 and 2 loops in stabilizing the closed conformation.	105
Figure A6. Length of hydrogen bond connecting ASP 444 and ARG 109 (blue) and inter-domain angle (orange) as a function of simulation time in the 2WHO trajectory.	106
Figure B1. Individual Free Energy landscapes of the inhibitor bound systems.	107
Figure B2. Individual Free Energy landscapes of the mutated systems.	118

Chapter 1. Introduction: Using the Hepatitis C Virus RNA-Dependent RNA Polymerase as a Model to Understand Viral Polymerase Structure, Function and Dynamics

*This chapter is reproduced in part from: Sesmero, E. and Thorpe, I.F, Viruses **2015**, 7, 3974-3994*

1.1 Dissertation Outline

The conformations sampled by the NS5B enzyme have a key role in replicating the RNA genome of HCV. Thus, characterizing the conformations will supply more detailed knowledge on how the enzyme works and how to inhibit it. The results generated by the Molecular Dynamics (MD) simulations of NS5B structures will provide a deeper insight into the structure, function and dynamics of each conformation. These simulations will also provide valuable information about the interactions that are key for stabilizing each conformation and facilitating transition between them. This knowledge may suggest new targets or novel strategies to inhibit the enzyme. In first place, the results of MD simulations of NS5B structures without inhibitor bound allowed us to construct the free energy landscape (FEL) of the conformational space sampled by NS5B as a free enzyme and elucidate the key interactions that stabilize each conformation. Once these conformations were characterized, they were compared with the results of MD simulations of NS5B bound to inhibitors and with mutations to allow further understanding of the fundamental role that conformational changes play in the function of viral polymerases.

1.1.1 Specific Aims:

- 1. Determine the structural and dynamic properties of the conformations sampled by NS5B as a free enzyme and understand what enzyme properties govern the transitions between them.***

We have shown that combining the results of several MD simulations of NS5B structures without inhibitor bound allowed us to generate the FEL of the conformational space sampled by the enzyme. Furthermore, from the simulation data we were able to characterize the structure, dynamics and key interactions of each conformation and their role in mediating transitions between them.

- 2. Determine how the structural and dynamic properties of the NS5B conformations are perturbed by mutations and the binding of inhibitors.***

Comparing our findings from Aim 1 with the results of MD simulations of NS5B bound to inhibitors or subject to mutations enabled us to understand how these conditions alter the conformational sampling and dynamics of the enzyme. This information illuminates the fundamental roles of conformational changes in RNA replication.

1.1.2 Significance

Despite a wealth of enzymatic and structural information regarding the HCV polymerase there are gaps in our understanding of the mechanism of RNA replication and the role that enzyme conformational changes and dynamics play in this process. In our research, we

employed extensive MD simulations to examine the conformational space explored by the enzyme. The FEL reveals features of the enzyme conformational space not apparent from the initial X-ray structures. This knowledge advanced our understanding of the molecular mechanism that governs the conformational transitions during the RNA replication cycle and allowed us to better understand the mechanisms of action of HCV polymerase inhibitors. This contribution is significant because it provides a deeper understanding of the role that conformational changes play in the RNA replication process. This knowledge may also be applied to other less studied viruses for which fewer treatments are available. In addition, this information could be transferred to understand the process of allosteric inhibition in other viral polymerases such as HIV-RT, facilitating the identification of novel inhibitors for these enzymes. In general, understanding the molecular basis of replication in viral polymerases will have a broad impact because these enzymes are critical for a number of viral infections. Moreover, greater knowledge of allostery is important because this process plays an important role in many biochemical processes such as enzyme kinetics and cell signaling.

1.2 Chapter Overview

The primary goals of this dissertation are to characterize the conformations of NS5B as a free enzyme and determine how they are perturbed by mutations and the binding of inhibitors. In Chapter 1, we give a general outlook about the knowledge up-to-date related to HCV, NS5B, the RNA replication process that it performs and the current inhibitors for it. In Chapter 2, we employ molecular dynamics (MD) simulations to understand the specific structural and dynamic characteristics of each conformation of NS5B as a free enzyme. We were able to generate the FEL of NS5B as a free enzyme, which was unknown before. This FEL shows the existence of

an intermediate conformation that was not previously reported. We characterized each one of the conformations (open, intermediate and closed) and the transitions between them, suggesting new targets and strategies for therapeutics. Chapter 3 discusses the effect of mutations and allosteric inhibitors on the NS5B conformations and how the latter are perturbed by those. The results from this study showed that the model we proposed for NS5B as a free enzyme may be widely applicable (to other genotypes and maybe other viral polymerases). It also confirmed our hypothesis that one mechanism of inhibition used by reported inhibitors is modifying the FEL. Furthermore, based on our H-bonds study, we found two strategies that the inhibitors use to prevent the enzyme from working that are: i) breaking the communication between domains, and ii) stabilizing preferentially one of the conformations. Lastly, Chapter 4 summarizes the implications of our findings and describes future research efforts.

1.3 Introduction

Polymerases are crucial in the viral life cycle. They have an essential role in replicating and transcribing the viral genome and as a result are key targets for therapies to treat viral infection. A virus may not need to encode its own polymerase depending on where it spends most of its life cycle. Some small DNA viruses that spend all their time in the cell nucleus can make use of the host cell's polymerases. However, viruses that remain in the cytoplasm do need to encode their own (1).

For viruses that require their own polymerase, most of these enzymes display detectable activity *in vitro* without accessory factors. This is primarily because the sizes of genomes that can be packaged in the viral capsid are limited (1, 2). In addition, some polymerases perform other functions related to viral genome transcription and replication. Examples include the RNA-

dependent RNA polymerases from the Flavivirus genus of the Flaviviridae family, retrovirus reverse transcriptases and some viral DNA-dependent polymerases. Flavivirus polymerases have a methyltransferase domain that catalyzes methylations of a 5'-RNA cap (3). The retrovirus reverse transcriptase has an additional ribonuclease H domain that catalyzes degradation of the RNA strand in the RNA-DNA hybrid during genome replication (4). Some viral DNA-dependent polymerases have a nuclease domain with proof-reading activity to correct nucleotides incorrectly incorporated during genome synthesis (5).

With regard to copying the viral genome, distinct replication mechanisms are used by different types of viral polymerases. A number of functions must be orchestrated depending on the specific virus in question (1):

- (1) Recognition of the nucleic acid binding site
- (2) Coordination of the chemical steps of nucleic acid synthesis
- (3) Conformational rearrangement to allow for processive elongation
- (4) Termination of replication at the end of the genome

Viral polymerases are often classified into four main categories based on the nature of the genetic material of the virus as follows: RNA-dependent RNA polymerases (RdRps), RNA-dependent DNA polymerases (RdDps), DNA-dependent RNA polymerases (DdRps), and DNA-dependent DNA polymerases (DdDps) (1). DdDps and DdRps are used for the replication and transcription, respectively, of DNA for both viruses and eukaryotic cells. In contrast, RdDps and RdRps are mainly used by viruses since the host cell does not require reverse transcription or RNA replication. RdDps are employed by retroviruses such as the human immunodeficiency virus (HIV). RdRps are employed by viruses such as Hepatitis C virus (HCV), poliovirus (PV), human rhinovirus (HRV), foot-and-mouth-disease virus (FMDV) and coxsackie viruses (CV),

among others. We will primarily focus on RdRps in this review since they are crucial in the replication process of viruses that are important global pathogens.

There are seven classes of viruses according to the Baltimore classification (6) based on the genome type and method of mRNA synthesis. These are associated with the four classes of polymerases specified in the previous paragraph as shown in Table 1.1.

1.4 General Structural Features of Viral Polymerases

The structure of all polymerases resembles a cupped right hand and is divided into three domains referred to as the palm, fingers and thumb (see Figure 1.1 a) (1, 7). This nomenclature is based on an analogy to the structure of the Klenow fragment of DNA polymerase (8). The palm domain is the most highly conserved domain across different polymerases and is the location of the active site. In contrast, the thumb domain is the most variable. Finger and thumb domains vary significantly in both size and secondary structure depending on the specific requirements for replication in a given virus (*i.e.*, replicating single- or double-stranded RNA/DNA genomes). The finger and thumb domains of different polymerases have similar positions with respect to the palm, which contains the active site in which catalytic addition of nucleotides occurs. Changes in the relative positions of the finger and thumb domains are associated with conformational changes of the polymerase at different stages of replication (7). Three well-defined channels have been identified on the polymerase, serving as the entry path for template and NTPs (*i.e.*, the template and NTP channels) and exit path for double stranded RNA (dsRNA) product (*i.e.*, the duplex channel) (9, 10) (see Figure 1.1 b,c,e).

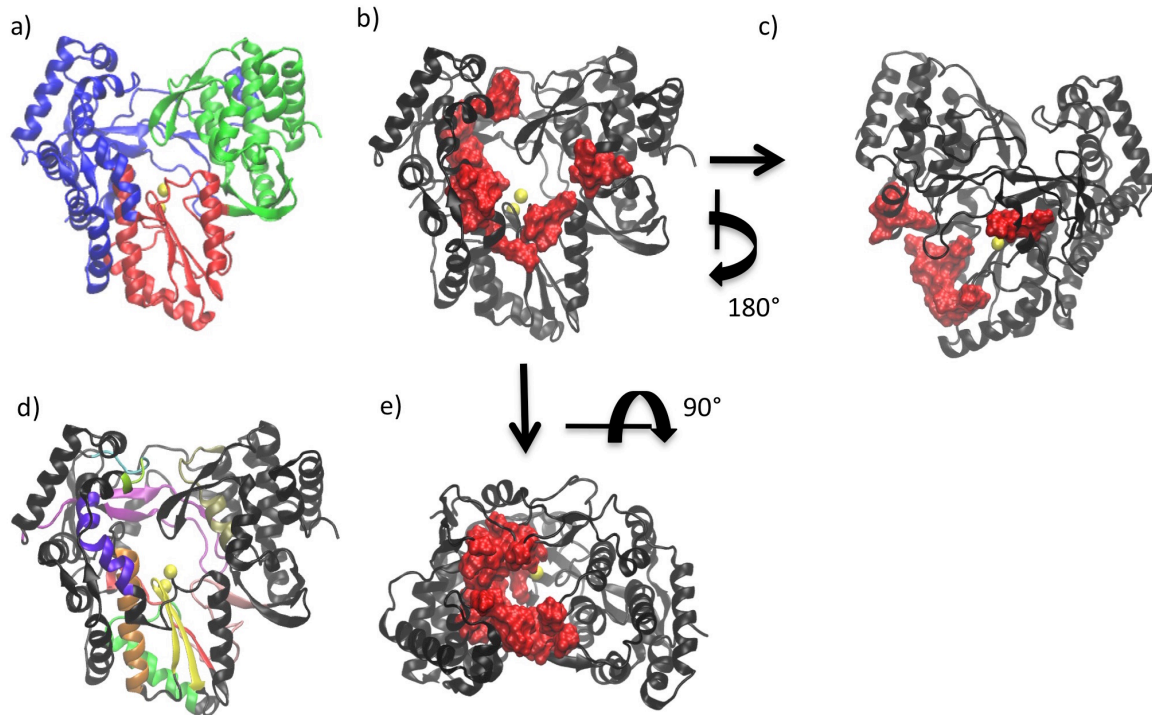


Figure 1.1. (a) Right-hand structure of HCV polymerase (NS5B). Palm, fingers and thumb domains are shown in red, blue and green, respectively; (b) Duplex channel in NS5B (front of the enzyme); (c) NTP channel in NS5B (back of the enzyme); (d) Motifs and functional regions of NS5B. Motif A in red, B in orange, C in yellow, D in bright green, E in pink, F in purple and G in cyan. Functional regions: I in light green, II in violet and III in tan; (e) Template channel (top view of the enzyme).

In the active site, the correct NTP to be added to the daughter strand is selected by Watson-Crick base-pairing with the template base. The selectivity for ribose (rNTP) vs. deoxyribose NTPs (dNTP) is regulated by the interaction of the polymerase with the 2'-OH of the NTP. In general, DNA polymerases that incorporate dNTP in the growing daughter strand have a large side chain that prevents binding of an rNTP with a 2'-OH. However, RNA polymerases utilize amino acids with a small side chain and form H-bonds with the 2'-OH of the rNTP. The polymerase active site often binds the correct NTP with 10–1000-fold higher affinity than incorrect NTPs (11). While viral polymerases often have domains in addition to the fingers, palm

and thumb that carry out functions related to other aspects of viral genome transcription and replication (see Introduction), this is not the case for the HCV polymerase

1.5 Conserved Structural Motifs of Viral Polymerases

There are several structural motifs (designated A through G, see Figure 1.1 d) that display varying levels of conservation among the different viral polymerases. Some motifs have been shown to be conserved across all viral polymerases (motifs A to E) while others (motifs F and G) have only been shown to be conserved for the RdRps. High levels of conservation, despite the low sequence similarity among polymerases, suggests that these motifs have functions that are vital for the action of these enzymes (1, 7, 9, 12, 13).

Motifs A and C have been closely studied because they are located in the active site. Motif C includes the GDD amino acid sequence that is the hallmark of RdRps. These conserved residues are bound to the metal ions (Mg^{2+} or Mn^{2+}) necessary for catalysis. Motif B contains a consensus sequence of SGxxxT and is located at the junction of the fingers and palm domain (7). Motif F binds to incoming NTPs and RNA and is situated near the entrance of the RNA template channel. The sequence of this motif is not conserved in *de novo* initiating RdRps such as that present in HCV (14). These polymerase sequence motifs have also been used to identify new polymerase genes in newly sequenced virus genomes (1). Further details about the roles of each motif are shown in Table 1.2 . Other regions have also been shown to have fundamental importance in the RdRps function and have been named “functional regions”. See Table 1.3 for a list of the residues included in these regions and the functional roles of each.

Table 1.1. Baltimore classification of viruses compared with the classification of viral polymerases based on their targeted genetic material.

Genetic material	Baltimore Classification	Polymerase Classes	Examples
DNA	ssDNA viruses	DNA dependent DNA polymerases	Human parvovirus B19
	dsDNA viruses	DNA dependent RNA polymerases	Bacteriophage ϕ 29
RNA	(+) ssRNA viruses	RNA dependent RNA polymerases	HCV, PV, West Nile virus
	(-) ssRNA viruses		Influenza
	dsRNA		Bacteriophage ϕ 6
RT	ssRNA-rt viruses	RNA dependent DNA polymerases	Retrovirus
	dsDNA-rt viruses		Hepatitis B

1.6 Structural Features of RdRps

RdRps replicate the genomic material in RNA viruses. Many of these viruses are significant public health concerns including HCV, Dengue virus, Japanese encephalitis and Yellow fever. For this reason, RdRps are key targets for new drugs and it is crucial to understand the mechanisms by which they replicate viral genomes. The fact that there are no mammalian homologs of RdRps (17, 18) makes them an optimal drug target because potential therapeutics would tend to selectively affect the viral polymerases without interfering with the function of host polymerases.

Within RdRp encoding viruses, there are ssRNA viruses (both + and – sense) and dsRNA viruses (see Table 1.3.). Genome replication in (+) ssRNA viruses takes place in a membrane-bound replication complex (9, 19, 20). (+) RNA serves as mRNA and can be translated immediately

after entering the cell (1). Thus, unlike the (–) RNA viruses, (+) RNA viruses do not need to package an RdRp within the virion (21).

Table 1.2. Characteristic sequence motifs in polymerases from prototypical viruses: hepatitis C virus (HCV), poliovirus (PV) and foot-and-mouth-disease virus (FMDV) (7, 9, 12, 15, 16).

Conserved elements		Role	Location	Residues		
				HCV	PV	FMDV
Motifs	A	Coordinates Magnesium and selects type of nucleic acid (RNA vs. DNA)	Palm	216–227	229–240	236–247
	B	Determines nucleotide choice (rNTP or dNTP)	Palm	287–306	293–312	303–322
	C	Coordinates Magnesium	Palm	312–325	322–335	332–345
	D	Helps accommodate active site NTPs	Palm	332–353	338–362	348–373
	E	Maintains rigidity of secondary structure that is required for relative positioning of thumb and palm domains	Palm	354–372	363–380	374–392
	F	Binds incoming NTPs and RNA	Fingers	132–162	153–178	158–183
	G	Binds primer and template	Fingers	95–99	113–120	114–121
Functional regions	I	Binds template	Fingers	91–94	107–112	108–113
	II	Binds template	Fingers	168–183	184–200	189–205
	III	Binds nascent RNA duplex	Thumb	401–414	405–420	416–430

Table 1.3. RdRps virus families and species (1, 22).

	Virus family	Representative Species
(+) ssRNA	Picornaviridae	Poliovirus (PV) Human rhinovirus (HRV) Foot-and-mouth-disease virus (FMDV) Coxsackie viruses (CV) Hepatitis A virus (HAV)
	Caliciviridae	Rabbit hemorrhagic disease virus (RHDV) Norwalk virus (NV) Sapporo virus
	Togaviridae	Sindbis virus
	Flaviviridae	West Nile virus (WNV) Yellow fever virus Dengue virus (DENV) Japanese encephalitis disease virus (JEV) Hepatitis C virus (HCV) Bovine viral diarrhea virus (BVDV)
(–) ssRNA	Orthomyxoviridae	Influenza virus
	Paramyxoviridae	Measles and mumps viruses
	Bunyaviridae	Hantavirus
	Rhabdoviridae	Rabies virus
	Filoviridae	Ebola and Marburg virus
	Bornaviridae	Borna disease virus
dsRNA	Cystoviridae	Bacteriophage $\phi 6$
	Reoviridae	Reovirus
	Birnaviridae	Fish infectious pancreatic necrosis virus (IPNV) Infectious bursal disease virus (IBDV)

The first X-ray structure of an RdRp was generated for Poliovirus (PV) polymerase in 1997 (23). X-ray structures are currently available from seven families of RdRps. These include (+) RNA viruses: Picornaviridae (PV, HRV, FMDV, CV and HAV), Caliciviridae (RHDV, NV and Sapporo virus) and Flaviviridae (HCV and BVDV) as well as (–) RNA viruses: Orthomyxoviridae (Influenza virus) and dsRNA viruses: Cystoviridae (Bacteriophage $\phi 6$), Reoviridae (Reovirus and Rotavirus) and Birnaviridae (IBDV). A table listing each NS5B structure currently available in the PDB is included as supporting information. The PDB IDs, a description of each structure and their resolution is provided. Similar information for other viral polymerases is presented in Tables 1 and 2 of Subissi *et al.* (14).

A characteristic trait of RdRps is the extensive interaction between finger and palm domains (24). RdRps have an extension of the finger domain called the fingertips that connects the finger and thumb domains to form a fully enclosed active site. The fingertips also contribute to the formation of well-defined template and NTP channels in the front and back of the polymerase, respectively.

RdRps were originally thought to be found uniquely in viruses. However, in 1971 the first eukaryotic RdRp was found in Chinese Cabbage (25). Later on, cellular RdRps were also found in plants, fungi and nematodes (26-28). Cellular RdRps play important roles in both transcriptional and post-transcriptional gene silencing (29). Although viral and cellular RdRps show little sequence homology, both share the “right hand” shape containing palm, thumb and finger domains. The palm domain of cellular RdRps is particularly well-conserved and contains four motifs maintained in all polymerases. These facts make it likely that the cellular RdRps share some of the basic mechanistic principles of viral RdRps and that knowledge obtained for viral RdRps may be transferable to cellular RdRps (13).

The Flaviviridae family has been widely studied because many members of this family cause diseases in humans. Within this family, there are three genera: Flaviviruses, Hepaciviruses and Pestiviruses (see Table 1.4.). HCV is part of the Hepaciviruses genus and is an important pathogen for which no vaccine is currently available. In explaining recent insights regarding the mechanism by which the HCV RdRp (gene product NS5B) replicates the viral genome, we will make comparisons with other members of the Flaviviridae family. However, we note that some differences may exist, particularly if the other family members are part of a different genus.

1.7 Catalytic Mechanism and Polymerase Reaction Steps

All known polymerases synthesize nucleic acid in the 5' to 3' direction (9). Thus, replication in positive-stranded RNA viruses occurs via a negative-stranded intermediate. The polymerase reaction has three stages: initiation, elongation and termination. For this cycle to take place the polymerase needs to have binding sites for: (a) the template strand; (b) the primer strand or initiating NTP (P-site) and (c) incoming NTP (N-site). The 3'-nucleotide defining the site of initiation is designated “*n*”. Residues at the “*n*” and “*n* + 1” positions of the template define the P-site and N-site.

Table 1.4. Genera and species of the Flaviviridae family.

Virus Family	Genus	Species
Flaviviridae	Flaviviruses	West Nile virus Yellow fever virus Dengue virus Japanese encephalitis disease virus
	Hepaciviruses	Hepatitis C virus (HCV)
	Pestiviruses	Bovine viral diarrhea virus (BVDV)

At the initiation stage, the formation of the first phosphodiester bond is key for polymerization of the nucleotides to begin. To form this phosphodiester bond, a hydroxyl group corresponding to a nucleotide 3'-OH is needed. Depending on how this 3'-OH is supplied, two mechanisms are differentiated: primer dependent in the case that a primer provides the required hydroxyl group, or primer independent (also called *de novo*) if this hydroxyl group is provided by the first NTP (2). The variety of mechanisms reflect the adaptation of the viruses to the host

cell (1). The size of the thumb domain seems to define whether a polymerase uses the primer-dependent or *de novo* mechanism. Most viruses in Picornaviridae and Caliciviridae families utilize a primer-dependent mechanism, but exceptions are found, such as noroviruses in the Caliciviridae family, that synthesize the (–) strand *de novo* (30). In general, these enzymes have a small thumb domain that provides a wider template channel to accommodate both template and primer. For this mechanism, different primers such as polypeptides, capped mRNAs or oligonucleotides may be used. In contrast, the Flaviviridae family that employs the *de novo* mechanism has a large thumb domain and narrower template channel suited to accommodate only the ssRNA and NTP (1, 14). However, we note that under certain conditions *de novo* polymerases can be induced to become primer dependent (31).

When the *de novo* mechanism is used, initiation takes place exactly at the 3'-terminus of the template RNA, so the initiating NTP (the first NTP of the growing strand) is dictated by the template. Both HCV and BVDV from the Flaviviridae family have been observed to require high concentrations of GTP for the initiation of RNA synthesis regardless of the RNA template nucleotide, (32, 33) which led to the suggestion that GTP may be needed for structural support of the initiating NTP. Harrus *et al.* (34) also suggested that GTP may act as the “initiation platform” and D'Abramo *et al.* (35) pointed out that this GTP may stabilize the interaction between the 3'-end of the template and the priming nucleotide. This stabilizing GTP binds inside the template channel, 6 Å from the catalytic site. It is not incorporated into the nascent RNA strand and is thought to be released from the active site during the elongation stage (9). We note that another GTP molecule has been reported to bind at the rear of the thumb domain near the fingertips in NS5B. This GTP has been suggested to play a role in activating *de novo* initiation or in allosterically regulating the conformational changes needed for replication (36).

Because base-pairing alone is insufficient to stabilize the dinucleotide product in the “P-site”, specialized structural elements are employed (13). Besides the stabilizing GTP, there is also a polymerase structural initiation platform, the so-called β -flap (residues 443–454). This β -flap likely supports the stabilizing GTP but would need to move out of the way in the elongation phase to allow the dsRNA product to exit (1, 9, 34). Other researchers have suggested the C-terminal linker (residues 531 to 570) also plays a regulatory role in the initiation stage of replication by acting as a buttress during the initiation stage and moving out of the template channel in a similar way as the β -flap in order to allow egress of the double stranded RNA (14, 37) (see Figure 1.2).

An advantage of the primer-dependent mechanism is that a stable elongation complex is formed more easily. There is limited abortive cycling, if any, and no requirement for large conformational rearrangements (13). In contrast, for the *de novo* mechanism, the first dinucleotide is not sufficiently stable, and an initiation platform is needed to provide additional stabilization. This reduced stability sometimes results in abortive cycling for the *de novo* mechanism. However, an advantage of the *de novo* mechanism is that no additional enzymes are needed to generate the primer (38).

After the template and primer or initiating NTP are bound to the enzyme, the steps required for single-nucleotide addition are (1):

- (1) incorporation of the incoming NTP into the growing daughter strand by formation of the phosphodiester bond
- (2) release of pyrophosphate
- (3) translocation along the template.

These three steps are repeated cyclically during elongation until the full RNA strand is replicated.

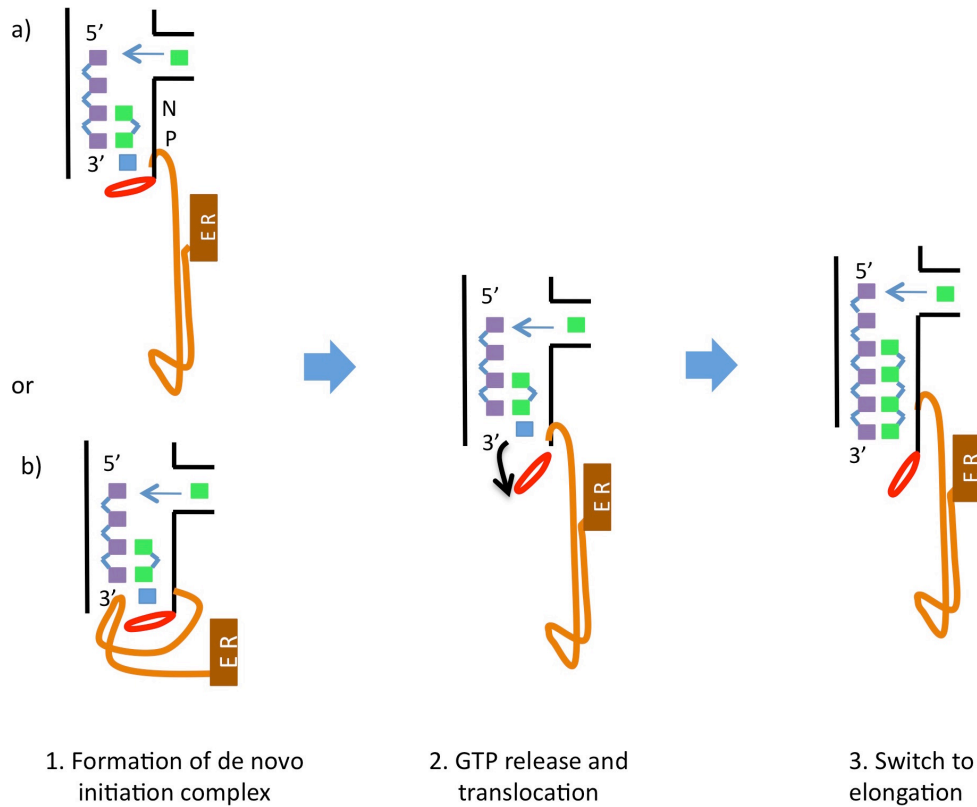


Figure 1.2. Schematic describing de novo initiation in Hepaciviruses and Pestiviruses. Note that Flaviviruses do not anchor their C-terminus in the Endoplasmic Reticulum (ER). This figure was generated by incorporating the descriptions provided by both Appleby *et al.* [35] and Choi [1]. The linker and C-terminal anchor are shown in orange as one contiguous element. The β -flap is colored red (as in Figure 1.3), the template strand in purple, the growing strand in green, the stabilizing GTP in blue and the Endoplasmic Reticulum (ER) in brown. The “N” and “P” indicate where the N-site (nucleotide-site) and P-site (priming-site) are. These correspond to the positions of the growing strand that bind to residues “ n ” and “ $n + 1$ ” of the template strand, respectively.

In order to facilitate nucleotide addition, all polymerases have two metal ions (Mg^{2+} or Mn^{2+}) in the active site bound to two conserved aspartic acid residues. These metal ions have been shown to be essential for catalysis via the so-called “two metal ions” mechanism. This

mechanism was proposed by Steiz in 1998 (39) and is as follows: the incoming NTP binds to metal ion B that orients the NTP in the active site and that may contribute to charge neutralization during catalysis. Metal ion B coordinates to the β - and γ -phosphate groups of the incoming NTP as well as the aspartic acid residue in motif A. Once the nucleotide is in place, the second divalent cation (Metal ion A) coordinates to the initiating NTP, lowering the pKa of the 3'-OH and facilitating nucleophilic attack on the α -phosphate. This then leads to formation of the phosphodiester bond and the release of pyrophosphate (PPi). Metal ion A coordinates to the α -phosphate group of the incoming NTP, the 3'-OH of the priming NTP and the aspartic acid residue in motif C (see Figure 1.4). Both metal ions stabilize the charge and geometry of the phosphorane pentavalent transition state during the nucleotidyl transfer reaction (1, 13, 38).

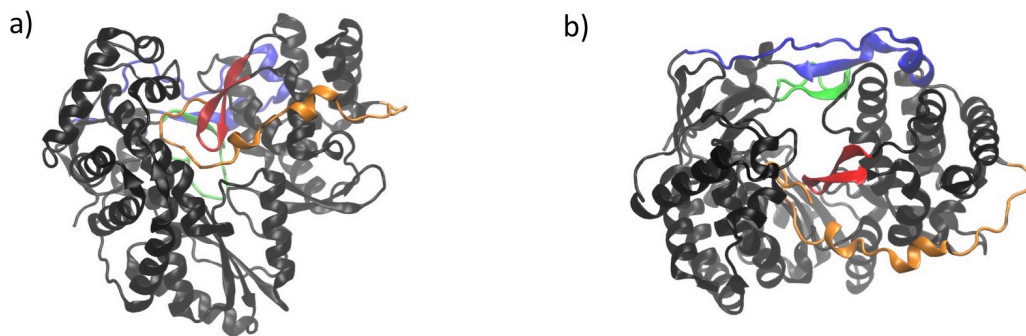


Figure 1.3. NS5B structure with characteristic elements highlighted. (a) Front view and (b) top view. The linker is shown in orange, the β -flap in red. The fingertips are shown in blue (the delta 1 loop) and green (the delta 2 loop).

The switch to elongation requires a major conformational change in the polymerase structure. Both the β -flap and the linker need to be displaced and an opening of the enzymatic core occurs. This open conformation may be one of the factors that enable a higher processivity in the

elongation stage compared to the initiation stage (for more detailed information about this change in conformation see Section 5 below).

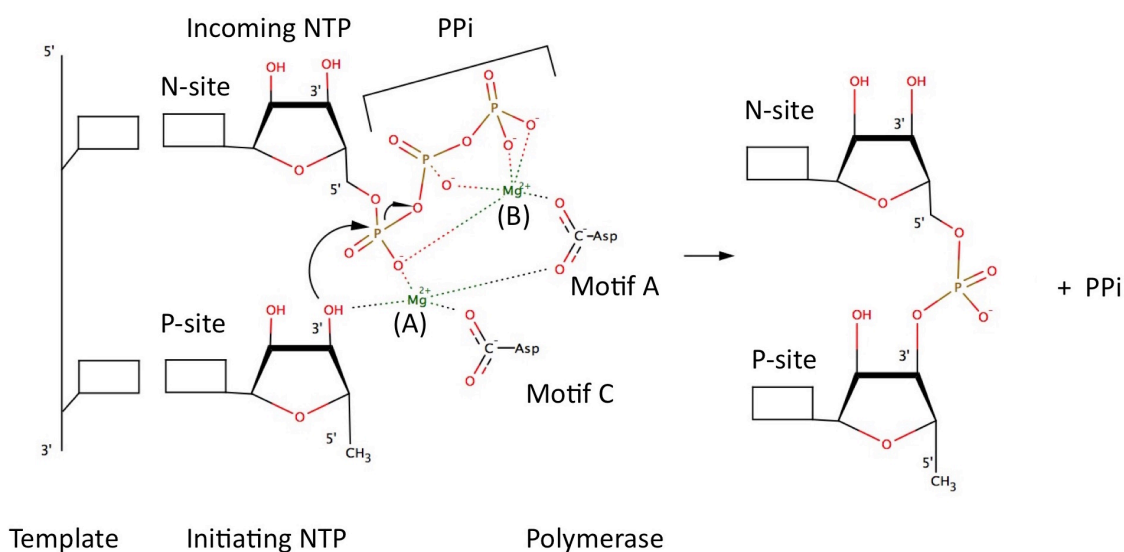


Figure 1.4. Two metals ions mechanism in RdRps. The squares represent the bases that are part of the nucleotides. This figure is inspired by a similar figure from Choi *et al.* (1).

Little is known about the termination of RNA synthesis. It has been suggested that the polymerase may simply fall off the end of the template once the complementary strand has been synthesized (14, 40). It is important to note that RNA synthesis by NS5B is error-prone due to the lack of proofreading activity of the RdRp enzymes. The mutation rates are estimated to be on the order of one mutation per 10^3 – 10^7 nucleotides resulting in approximately one error per replicated genome (1, 41). In contrast, the mutation rate in *E. coli*, where cellular polymerases benefit from error-correcting mechanisms, is on the order of one mutation per 10^9 – 10^{10} nucleotides (1). The large error rate results in the high genetic variability of the HCV viruses and provides a molecular basis for the rapid development of resistance to therapies.

1.8 NS5B Conformational Changes during the Replication Cycle

One characteristic unique to viral RdRps is their “closed-hand” shape. This terminology started to be used because their X-ray structures appear to be more closed than the previously characterized DdDps, DdRps and RTs (called “open-hand”) (7, 14, 38, 42). This “closed-hand” shape is characterized by the fingertips region, a hallmark of RdRps, that connects the finger and palm domains on the back of the enzyme as well as by the so-called β -flap on the front of the enzyme (see Figure 1.3). The latter is specific to the Flaviviridae RdRps while the linker, or a variation of it, is common to most of the *de novo* initiating RdRps (40) (note, however, that it is not found in the Flavivirus RdRps). The linker (residues 531 to 570) connects the NS5B catalytic core (residues 1 to 530) with the C-terminus transmembrane anchor (residues 571 to 591). These last twenty-one C-terminal residues seem not to influence RNA synthesis *in vitro* (40). Given that these residues are very hydrophobic, their removal facilitates expression and purification of the enzyme. Thus, most biochemical and all structural studies have been carried out with the so-called NS5B Δ 21 enzyme variant in which these residues have been removed.

Most of the NS5B structures that have been reported are thought to be in the closed conformation. However, it has been observed that *de novo* initiation by NS5B *in vitro* does not only occur at the 3' end of the template but also can take place at internal template sites (43, 44) and on circular templates (31). These facts suggest that in solution there is an equilibrium between the closed and open conformations. The existence of the open conformation is supported by the structure of NS5B from genotype 2a NS5B, (45) as well as the structure recently published by Mosley *et al.* (46). The latter contains a variant of NS5B that lacks the β -flap in complex with primer-template RNA. Molecular dynamics simulations of Davis *et al.* (47) also indicate the occurrence of open NS5B conformations.

The closed conformation is thought to represent the initiation state of the polymerase. In this conformation, the catalytic core only provides sufficient space for a single-stranded RNA template and the nucleotides required for *de novo* initiation of RNA synthesis, but is not wide enough to accommodate double-stranded RNA (40). To transition to elongation, a major conformational change is needed so that the nascent RNA can egress. Primer-dependent RdRps undergo less dramatic conformational changes than *de novo*-initiating RdRps (14) because the thumb domain of primer-dependent RdRps is smaller, leaving enough room for the dsRNA product to exit. Transitioning to elongation in RdRps thus requires the adoption of an open conformation (34, 40, 46, 48). To arrive at the open conformation, the β -flap would need to be moved out of the way, the stabilizing GTP should unbind and also a rotation of the thumb domain should take place. This would position it further from the center of the enzyme, increasing the size of the template and duplex channels so the dsRNA can exit the enzyme (34, 48). If the C-terminal linker does act as an initiation platform together with the β -flap, this element would also need to move away from the template channel in the transition to elongation as described by Appleby *et al.* (37) (see Figure 1.2). It is worth noting that conformational changes have been reported in several RdRp structures (45, 46, 49). These findings suggest that these enzymes exhibit considerable conformational variability, which is similar to observations made for other polymerases (50, 51).

1.9 NS5B Inhibitors and Mechanisms of Action

There are two main classes of NS5B inhibitors: nucleoside inhibitors (NIs) and non-nucleoside inhibitors (NNIs) (see Figure 1.5). NIs bind in the active site and generally act as non-

obligate terminators of RNA synthesis after being incorporated into the newly produced RNA strand. The advantages of NIs are that they have shown stronger antiviral activity, are able to inhibit multiple HCV genotypes and have a higher barrier to the emergence of drug resistance (48). However, they have the potential to also affect host polymerases since they interact with an active site that has similar features among diverse types of polymerases. Sofosbuvir, the drug most recently approved for HCV treatment is in this group. NNIs are allosteric inhibitors that bind to sites other than the active site. NNIs are also promising, though they have not yet been used in a clinical setting. NNIs are attractive for use in future anti-HCV therapies due to the decreased likelihood that they will exhibit nonspecific side effects compared to NIs. However, HCV is more likely to become resistant to these inhibitors because there is typically not strong evolutionary pressure to maintain the amino acid sequence of NNI binding sites. We focus on NNIs in this review because the role of NIs as terminators of RNA synthesis is well understood. In contrast, although many structures with NNIs bound have been solved, their mechanism of action still remains to be elucidated.

Four NNI sites have been identified: two in the thumb (NNI-1 and NNI-2) and two in the palm (NNI-3 and NNI-4) (see Figure 1.5). Brown *et al.* [74] provide evidence that NNI-3 and NNI-4 are likely to be distinct regions within a single large pocket rather than two individual pockets. For this reason, we use the nomenclature NNI-3/4 to denote both of these partially overlapping sites. Due to the fact that there are multiple distinct allosteric sites, it may be possible to use multiple NNIs in combination with each other or with NIs in the effort to overcome resistance. NNIs are thought to inhibit NS5B by affecting the equilibrium distribution of conformational states required for normal catalytic activity of the enzyme (13, 47). Most of

the NNIs that bind to the palm domain have been found to stabilize the β -flap via critical interactions with Tyr448 (60), fixing it in the closed, initiation-appropriate conformation and preventing these residues from moving out to allow the RNA double helix to egress (46). NNI-2

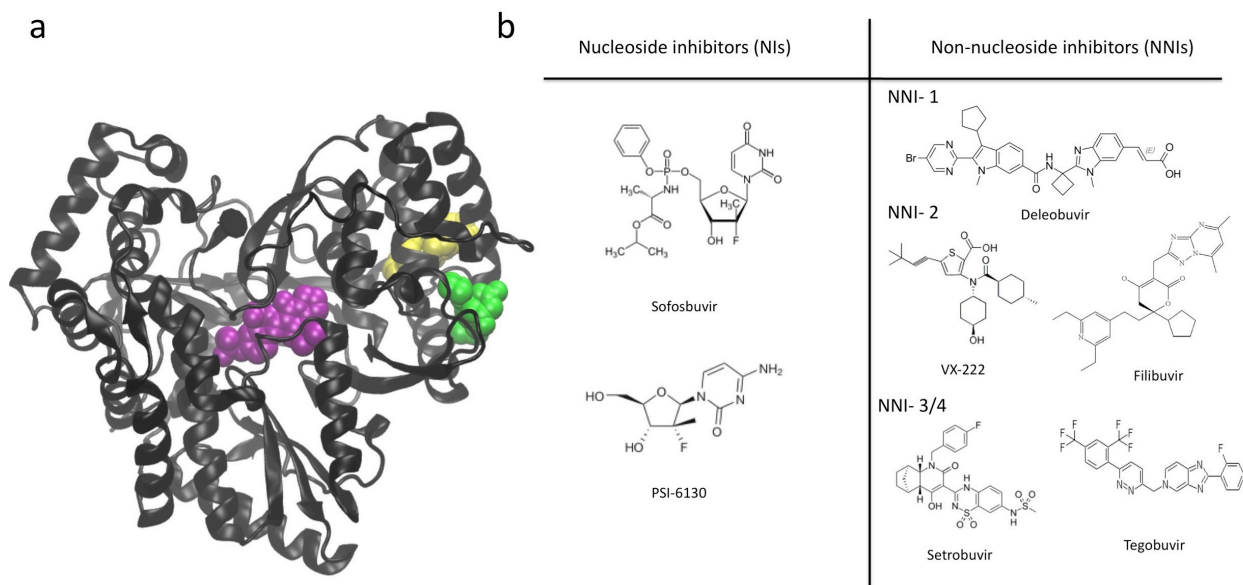


Figure 1.5. NS5B inhibitors. **(a)** The three allosteric sites of NS5B are highlighted with space filling representations of inhibitors that bind in these locations. Thumb site 1 (NNI-1) in yellow, thumb site 2 (NNI-2) in green, palm sites (NNI-3/4) in purple; **(b)** chemical structures of NIs and NNIs that are in clinical trials or have already been approved (52-59).

ligands have also been suggested to prevent the occurrence of important conformational changes in NS5B (16, 45, 61). Some studies have suggested that palm NNIs inhibit initiation while thumb NNIs inhibit an early phase of replication that occurs after initiation but before elongation starts (62-65). Thus, the different allosteric sites may display distinct modes of action. Davis *et al.* [46] studied the mechanism of inhibition of allosteric inhibitors in the different allosteric sites. They found that inhibitors in the NNI-1 pocket seem to prevent enzyme function by reducing its overall stability and preventing it from stably adopting functional conformations. In contrast, NNI-2 inhibitors seem to reduce conformational sampling, preventing the transitions

between conformational states that are required for NS5B to function. NNI-3 inhibitors were also observed to restrict conformational sampling, though the dominant mode of action of these molecules was predicted to result from blocking access of the RNA template (see Figure 1.6).

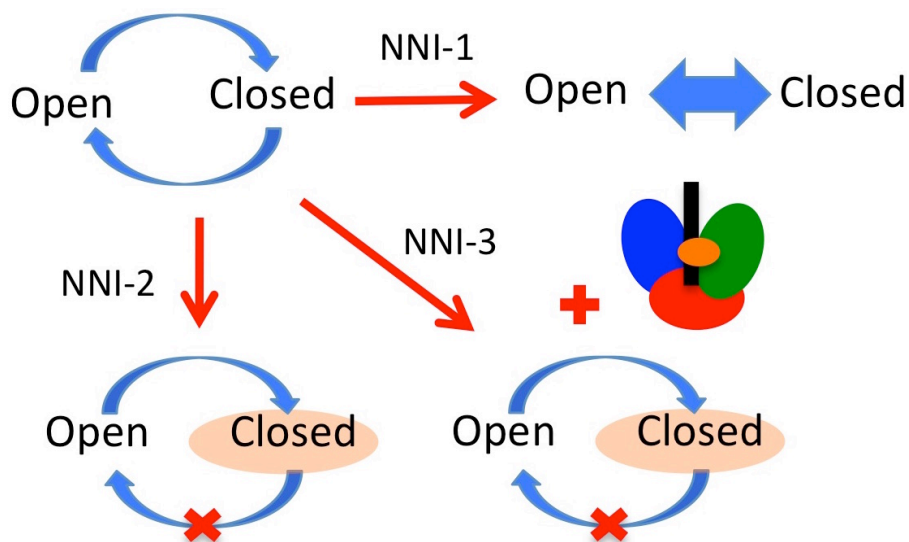


Figure 1.6. Mechanisms of inhibition for NNIs. NS5B must transition between open and closed states to perform replication (upper left). NNI-1 inhibitors have been observed to reduce enzyme stability. NNI-2 inhibitors have been shown to reduce conformational sampling, confining the enzyme in closed conformations. NNI-3 inhibitors mainly block access of the RNA template but also induce some restriction of conformational sampling. The RNA template is represented as a black rectangle and the inhibitor as an orange ellipse.

This may facilitate their use in combination therapies by degrading complementary functionalities in the enzyme. Understanding the molecular mechanisms by which small molecules in general and NNIs in particular inhibit the function of NS5B is essential for rationally design NS5B inhibitors. Such molecules may ultimately serve as a basis for more efficacious or cost-effective HCV therapies, either individually or in combination.

One informative example that illustrates the useful interplay between determining the roles of structural and functional elements of NS5B and understanding the efficacy of NNIs is provided

by recent studies of Gilead pharmaceuticals. Boyce *et al.* (66) assessed the activities and biophysical properties of a number of NS5B variants using mutations and deletions in the enzyme C-terminus and β -flap, in concert with challenging the enzyme using diverse NNIs. Their observations suggest that ligands which bind to NNI-2 exhibit a unique inhibitory mechanism relative to other NNIs. Boyce *et al.* (66) discovered that NNI-2 ligands are most effective when both the C-terminus and β -flap of the enzyme are present. These inhibitors were found to stabilize NS5B in a closed conformation, consistent with simulation studies by Davis *et al.* (16, 47). Boyce *et al.* (67) found that interactions between the C-terminus and the β -flap were required for inhibition, but not for ligand binding. These authors determined that NNI-2 inhibitors exhibited decreased efficacy for truncated NS5B variants and suggested that while the C-terminus and β -flap do not alter the intrinsic interactions of NNI-2 ligands with the enzyme, they do play an important role in propagating the allosteric effects that result from inhibitors binding to distant enzyme locations. This finding is consistent with mutational data for NNI-2 inhibitors, which map viral resistance mutations to areas around the β -flap(67).

Simulation studies by Davis and Thorpe suggest that the enzyme C-terminus reduces conformational sampling in NS5B, likely eliminating transitions between the closed and open conformations necessary for the initiation and elongation phases of replication respectively (68). These observations predict that enzymes without C-terminal residues should display increased activity, consistent with the findings of Boyce *et al.* (67). Other studies from Davis *et al.* (16, 47) indicate that an NNI-2 ligand can restrict conformational sampling even if the C-terminus is absent, stabilizing the enzyme in a very closed state. One might expect that this property could account for the inhibitory action of NNI-2 ligands without needing to invoke a role for the C-terminus as suggested by Boyce *et al.* (67). However, there are several important considerations

to be noted. First, the simulation studies examine the impact of binding a ligand to the enzyme and do not directly probe inhibition. The studies of Boyce *et al.* (67) indicate that binding affinities of NNI-2 ligands are not a good proxy for inhibition efficacy. Thus, observations in the simulation studies may not be explicitly linked to allosteric inhibition. Another consideration is that the simulation studies were not carried out with both the inhibitor and the C-terminus present. It is possible that conformational restriction of the enzyme in the presence of both entities would be even more dramatic, consistent with the enhanced inhibition in Boyce *et al.* (67) measured in the presence of the C-terminus. Finally, different ligands were employed in each study and it could be that distinct inhibitors elicit different effects even though they bind to the same location. There is evidence that different NNI-2 ligands are able to alter the conformational distribution of NS5B to different extents (47). Thus, it is possible that the enzyme C-terminus is only required for observing the inhibitory effects of certain NNI-2 ligands.

In contrast to NNI-2, Boyce *et al.* (67) observed that the potency of NNI-1 ligands was not affected by the presence of the C-terminus or β -flap. This observation suggests that these ligands possess a completely different mechanism of action compared to NNI-2 inhibitors. These authors noted that the presence of NNI-1 ligands lowered the melting temperature of NS5B, consistent with decreased stability of the enzyme. The decrease of NS5B stability in the presence of NNI-1 ligands was noted as well in other studies (69, 70). This finding is also consistent with results from simulations of Davis *et al.* (47) that suggest NNI-1 and NNI-2 ligands have distinct modes of action. In contrast to the stabilizing effect of NNI-2 ligands, it was observed that an NNI-1 ligand destabilized conformational sampling in NS5B, preventing the enzyme from stably occupying functional conformational states.

With regard to palm inhibitors, Boyce *et al.* (47) observed that such ligands display larger dissociation constants in NS5B constructs for which C-terminal residues were deleted, suggesting that the C-terminus facilitates binding to palm sites. Palm site inhibitors also demonstrate decreased potency in these deletion constructs, indicating that the C-terminus is needed for both binding and inhibition. In simulation studies, Davis *et al.* (47) observed that NNI-3 ligands were able to bind to the enzyme without the C-terminus present and also restricted conformational sampling of NS5B in a similar manner to NNI-2 ligands. However, the conformations sampled when ligands were bound to NNI-3 tended to be more open in general than those induced by an NNI-2 ligand. These conformations may perturb the replication cycle to a reduced extent compared to NNI-2 or NNI-1 ligands. It is possible that in the presence of the C-terminus NNI-3 ligands elicit more dramatic changes in conformational sampling. Nonetheless, the authors concluded that the dominant inhibitory effect of palm ligands is likely due to direct obstruction of the RNA template channel (thus preventing the template from accessing the active site) rather than conformational restriction. This observation is consistent with previous predictions (71).

The findings of Boyce *et al.* (67) are important because they indicate the enzyme C-terminus plays a crucial role in modulating the efficacy of NNIs. The likely molecular basis of this observation can be readily understood by considering the schematic shown in Figure 1.2 . In this figure, it is apparent that the C-terminus acts as a “stopper” in the template channel, preventing elongation of the nascent RNA strand. Thus, both the C-terminus and β -flap need to be removed from the template channel before elongation can proceed. If the C-terminus is not present, the template channel cannot be effectively blocked and replication is less likely to be affected by presence of the inhibitor. This is a quite interesting result, as it points to the limitations of some

inhibitor studies that may have been carried out *in vitro* using enzyme variants without the C-terminus. It is likely that any ligands employing the inhibitory mechanisms described by Boyce *et al.* (47) would not be identified in such studies. Thus, the role of NS5B regulatory elements in strongly modulating the efficacy of inhibitors must be taken into account when assessing ligand potency.

Studies such as those of Boyce *et al.* (66) or Davis and colleagues (16, 47, 68) may be useful to understand the differing susceptibility of different NS5B variants (and thus different HCV strains or genotypes) to the presence of diverse inhibitors. For example, in some viral genotypes, the C-terminus might interact more strongly with the template channel than in others. One would anticipate that NNI-2 ligands would be more effective in inhibiting such enzyme variants. The studies reviewed in this article indicate that understanding the structure and function of NS5B provides powerful insight into the molecular mechanisms governing inhibition of this enzyme and the functional properties of other RdRps. For example, recent structural studies of the Influenza virus polymerase reveal a β -flap element similar to that which modulates the activity of NS5B and which may adopt a similarly important role in these enzymes (72, 73). We note that simulation studies are particularly helpful in this regard by allowing molecular mechanisms underlying the observed structure-function relationships to be elucidated (16, 47, 68).

Understanding the molecular mechanisms involved in inhibition by NNIs could facilitate the design and deployment of these molecules. The insights acquired may also be transferable to other polymerases to better understand the relationship between structure, function and dynamics in these enzymes. Due to the fact that individual NNIs can have distinct sites of binding, it should be possible to combine multiple NNIs such that their total inhibitory effect is enhanced relative to applying any given inhibitor on its own (74). It may be beneficial to target

complementary activities or distinct conformational states of the enzyme with an array of small molecules to degrade a wide spectrum of NS5B functionality in a therapeutic context. For example, it is possible that a large fraction of NS5B exists within the host cell in an auto-inhibited state with the C-terminus occupying the template channel. In this way, the virus can avoid negatively perturbing the host cell and facilitate evasion of the host immune response. One could envision targeting both actively replicating and auto-inhibited NS5B molecules with different inhibitors in order to more effectively degrade intracellular enzyme activity.

1.10 Summary

Flaviviridae viruses are (+) RNA viruses with RdRp polymerases that utilize the *de novo* mechanism for initiation. While *Flaviviridae* polymerases possess elements common to other RdRps such as the fingertips region, they are also unique in possessing the β -flap that may be used as an initiation platform during genome replication.

The important pathogen HCV is a member of the *Flaviviridae* family within the *Hepacivirus* genus and employs NS5B as the RdRp that replicates its genome. There are two key steps involved in the replication process: (1) the formation of the initial dinucleotide and (2) the transition from initiation to processive elongation. Structural elements of NS5B that likely have a crucial role in these steps are the C-terminal linker and the β -flap (see Figure 1.3). Initiation is also facilitated by the so-called “stabilizing GTP” in the active site (see Figure 1.2). Finally, a conformational change involving movement of the thumb and finger domains to position them further apart has been observed to accompany the transition from initiation to elongation, resulting in an open-hand conformation. The linker and the β -flap may have dual roles: (1) acting

as initiation platforms to stabilize formation of the first dinucleotide and (2) regulating the transition to elongation. These structural elements can prevent the enzyme from moving to the elongation stage and must be displaced to allow for processive elongation to take place.

Thus, the available evidence suggests that NS5B possesses an intrinsic capacity to be regulated via allosteric effectors including NNIs, the β -flap and the C-terminal linker. In addition, the role of these effectors seems to be strongly modulated by the specific context of the interaction. Understanding how these structural elements govern enzyme activity and how they interface with inhibitors is important for understanding the molecular mechanisms of allosteric inhibition in NS5B. Such knowledge paves the way for rational design of inhibitors and combination therapies both for NS5B and for the polymerases to which these insights can be generalized. This information may also be useful in designing enzymes with attenuated activity, as would be required if one sought to develop a strain of HCV that could serve as the basis for a vaccine. Attenuating HCV by degrading the activity of NS5B is one strategy that could prove useful in this regard. One potential drawback to such efforts is the high mutation rate of HCV that results from the error-prone nature of NS5B. However, it is possible that one could circumvent this issue by generating a polymerase that not only possesses reduced efficacy, but also displays increased fidelity and thus faithfully replicates the viral genome.

Viral polymerases and, specifically, RdRps share many common structural, functional and dynamic features. Thus, the knowledge obtained in understanding how NS5B functions may be transferable to polymerases from closely related viruses such as Dengue or West Nile virus, or even to other more distantly related polymerases such as reverse transcriptase from HIV and 3D-pol from Poliovirus.

1.11 Acknowledgements

I would like to thank Dr. David Gonzalez Calatayud for his help providing the program to make figure 1.4 .

1.12 References

1. Choi, K.H. Viral polymerases. In *Viral Molecular Machines*; Springer Science: New York, USA, 2012.
2. Ortin, J.; Parra, F. Structure and function of rna replication. *Annu. Rev. Microbiol.* **2006**, *60*, 305–326.
3. Zhou, Y.; Ray, D.; Zhao, Y.; Dong, H.; Ren, S.; Li, Z.; Guo, Y.; Bernard, K.A.; Shi, P.-Y.; Li, H.; *et al.* Structure and function of flavivirus ns5 methyltransferase. *J. Virol.* **2007**, *81*, 3891–3903.
4. Zhou, D.; Chung, S.; Miller, M.; Grice, S.F.J.L.; Wlodawer, A. Crystal structures of the reverse transcriptase-associated ribonuclease h domain of xenotropic murine leukemia-virus related virus. *J. Struct. Biol.* **2012**, *177*, 638–645.
5. Knopf, C. Evolution of viral DNA-dependent DNA polymerases. *Virus Genes* **1998**, *16*, 47–58.
6. Baltimore, D. Expression of animal virus genomes. *Bacteriol. Rev.* **1971**, *35*, 235–241.
7. Shatskaya, G.S. Structural organization of viral RNA-dependent RNA polymerases. *Biochemistry* **2013**, *78*, 231–235.
8. Ollis, D.L.; Brick, P.; Hamlin, R.; Xuong, N.G.; Steitz, T.A. Structure of large fragment of *Escherichia coli* DNA polymerase i complexed with dtmp. *Nature* **1985**, *313*, 762–766.
9. McDonald, S.M. RNA synthetic mechanisms employed by diverse families of RNA viruses. *WIREs RNA* **2013**, *4*, 351–367.
10. Ferrer-Orta, C.; Verdaguer, N. RNA virus polymerases. In *Viral Genome Replication*; Cameron, C., Gotte, M., Raney, K.D., Eds.; Springer Science: New York, USA, 2009.
11. Gao, G.; Orlova, M.; Georgiadis, M.M.; Hendrickson, W.A.; Goff, S.P. Conferring RNA polymerase activity to a DNA polymerase: A single residue in reverse transcriptase controls substrate selection. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 407–411.
12. Cameron, C.E.; Moustafa, I.M.; Arnold, J.J. Dynamics: The missing link between structure and function of the viral RNA-dependent RNA polymerase? *Curr. Opin. Struct. Biol.* **2009**, *19*, 768–774.
13. Ng, K.K.-S.; Arnold, J.J.; Cameron, C.E. *Structure and Function Relationships Among RNA-Dependent RNA Polymerases*; Springer-Verlag: Berlin, Germany; Heidelberg, Germany, 2008; Volume 320.
14. Subissi, L.; Decroly, E.; Selisko, B.; Canard, B.; Imbert, I. A closed-handed affair: Positive-strand RNA virus polymerases. *Future Virol.* **2014**, *9*, 769–784.
15. Moustafa, I.M.; Shen, H.; Morton, B.; Colina, C.M.; Cameron, C.E. Molecular dynamics simulations of viral RNA polymerases link conserved and correlated motions of functional elements to fidelity. *J. Mol. Biol.* **2011**, *410*, 159–181.
16. Davis, B.; Thorpe, I.F. Thumb inhibitor binding eliminates functionally important dynamics in the hepatitis c virus RNA polymerase. *Proteins Struct. Funct. Bioinform.* **2013**, *81*, 40–52.

17. International Committee on Taxonomy of Viruses. Available online: <http://www.Ictvonline.Org> (accessed on February 15, 2015).
18. Gong, J.; Fang, H.; Li, M.; Liu, Y.; Yang, K.; Xu, W. Potential targets and their relevant inhibitors in anti-influenza fields. *Curr. Med. Chem.* **2009**, *16*, 3716–3739.
19. Malet, H.; Masse, N.; Selisko, B.; Romette, J.L.; Alvarez, K.; Guillemot, J.C.; Tolou, H.; Yap, T.L.; Vasudevan, S.; Lescar, J.; *et al.* The flavivirus polymerase as a target for drug discovery. *Antivir. Res.* **2008**, *2008*, 23–35.
20. Welsch, S.; Miller, S.; Romero-Brey, I. Composition and three-dimensional architecture of the dengue virus replication and assembly sites. *Cell Host Microbe* **2009**, *5*, 365–375.
21. Hsu, N.Y.; Ilnytska, O.; Belov, G. Viral reorganization of the secretory pathway generates distinct organelles for RNA replication. *Cell* **2010**, *141*, 799–811.
22. Zuckerman, A.J. Hepatitis viruses. In *Medical Microbiology*; Baron, S., Ed.; The University of Texas Medical Branch: Galveston, TX, USA, 1996.
23. Hansen, J.L.; Long, A.M.; Schultz, S.C. Structure of the RNA-dependent RNA polymerase of poliovirus. *Structure* **1997**, *5*, 1109–1122.
24. Lindenbach, B.D.; Tellinghuisen, T.L. Hepatitis c virus genome replication. In *Viral Genome Replication*; Cameron, C., Gotte, M., Raney, K.D., Eds.; Springer Science: New York, USA, 2009.
25. Astier-Manifacier, S.; Cornuet, P. RNA-dependent RNA polymerase in chinese cabbage. *Biochim. Biophys. Acta* **1971**, *232*, 484–493.
26. Boege, F.; Heinz, L.S. RNA-dependent RNA polymerase from healthy tomato leaf tissue. *FEBS Lett.* **1980**, *121*, 91–96.
27. Cogoni, C.; Macino, G. Gene silencing in neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase. *Nature* **1999**, *399*, 166–169.
28. Smardon, A.; Spoerke, J.M.; Stacey, S.C.; Klein, M.E.; Mackin, N.; Maine, E.M. Ego-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in c. Elegans. *Curr. Biol.* **2000**, *10*, 169–178.
29. Maida, Y.; Masutomi, K. RNA-dependent RNA polymerases in RNA silencing. *Biol. Chem.* **2011**, *392*, 299–304.
30. Rohayem, J.; Robel, I.; Jager, K.; Scheffler, U.; Rudolph, W. Protein-primed and *de novo* initiation of RNA synthesis by norovirus 3dpol. *J. Virol.* **2006**, *80*, 7060–7069.
31. Ranjith-Kumar, C.T.; Kao, C.C. Recombinant viral rdrps can initiate RNA synthesis from circular templates. *RNA* **2006**, *12*, 303–312.
32. Luo, G.; Hamatake, R.K.; Mathis, D.M.; Racela, J.; Rigat, K.L.; Lemm, J.; Colonno, R.J. *De novo* initiation of RNA synthesis by the RNA-dependent RNA polymerase (ns5b) of hepatitis C virus. *J. Virol.* **2000**, *74*, 851–863.
33. Kao, C.C.; Vecchio, A.M.D.; Zhong, W. *De novo* initiation of RNA synthesis by a recombinant flaviviridae RNA-dependent RNA polymerase. *Virology* **1999**, *253*, 1–7.
34. Harrus, D. Further insights into the roles of GTP and the C terminus of the hepatitis C virus polymerase in the initiation of RNA synthesis. *J. Biol. Chem.* **2010**, *285*, 32906–32918.
35. D'Abramo, C.M.; Deval, J.; Cameron, C.E.; Cellai, L.; Gotte, M. Control of template positioning during *de novo* initiation of RNA synthesis by the bovine viral diarrhea virus NS5B polymerase. *J. Biol. Chem.* **2006**, *281*, 24991–24998.
36. Bressanelli, S. Structural analysis of the hepatitis C virus RNA polymerase in complex with ribonucleotides. *J. Virol.* **2002**, *76*, 3482–3492.

37. Appleby, T.C.; Perry, J.K.; Murakami, E.; Barauskas, O.; Feng, J.; Cho, A.; Fox, D., III; Wetmore, D.R.; McGrath, M.E.; Ray, A.S.; *et al.* Structural basis for RNA replication by the hepatitis C virus polymerase. *Science* **2015**, *347*, 771–775.
38. Van Dijk, A.A.; Makeyev, E.V.; Bamford, D.H. Initiation of viral RNA-dependent RNA polymerization. *J. Gen. Virol.* **2004**, *85*, 1077–1093.
39. Steitz, T. A mechanism for all polymerases. *Nature* **1998**, *391*, 231–232.
40. Lohmann, V. *Hepatitis C Virus: From Molecular Virology to Antiviral Therapy*; Springer-Verlag: Berlin, Germany; Heidelberg, Germany, 2013; Volume 369.
41. Drake, J.W. A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 7160–7164.
42. Ferrer-Orta, C.; Arias, A.; Escarmi, C.; Verdaguer, N. A comparison of viral RNA-dependent RNA polymerases. *Curr. Opin. Struct. Biol.* **2006**, *16*, 27–34.
43. Binder, M.; Quinckert, D.; Bochkarova, O.; Klein, R.; Kezmic, N.; Bartenschlager, R.; Lohmann, V. Identification of determinants involved in initiation of hepatitis c virus RNA synthesis by using intergenotypic chimeras. *J. Virol.* **2007**, *81*, 5270–5283.
44. Shim, J.H.; Larson, G.; Hong, J.Z. Selection of 3' template bases and initiating nucleotides by hepatitis c virus RNA by and ago2-miR-122 complex. *Proc. Natl. Acad. Sci. USA* **2002**, *109*, 941–946.
45. Biswal, B.K.; Cherney, M.M.; Wang, M.; Chan, L.; Yannopoulos, C.G.; Bilimoria, D.; Nicolas, O.; Bedard, J.; James, M.N. Crystal structures of the RNA-dependent RNA polymerase genotype 2A of hepatitis C virus reveal two conformations and suggest mechanisms of inhibition by non-nucleoside inhibitors. *J. Biol. Chem.* **2005**, *280*, 18202–18210.
46. Mosley, R.T. Structure of hepatitis C virus polymerase in complex with primer- template RNA. *J. Virol.* **2012**, *86*, 6503–6511.
47. Davis, B.C.; Brown, J.A.; Thorpe, I.F. Allosteric inhibitors have distinct effects, but also common modes of action, in the hcv polymerase. *Biophys. J.* **2015**, *108*, 1785–1795.
48. Caillet-Saguy, C.; Lim, S.P.; Shi, P.-Y.; Lescar, J.; Bressanelli, S. Polymerases of hepatitis C viruses and flaviviruses: Structural and mechanistic insights and drug development. *Antivir. Res.* **2014**, *105*, 8–16.
49. Choi, K.H.; Groarke, J.M.; Young, D.C.; Kuhn, R.J.; Smith, J.L.; Pevear, D.C.; Rossmann, M.G. The structure of the RNA-dependent RNA polymerase from bovine viral diarrhea virus establishes the role of GTP in de novo initiation. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 4425–4430.
50. Rothwell, P.J.; Waksman, G. Structure and mechanism of DNA polymerases. *Adv. Protein Chem.* **2005**, *71*, 401–440.
51. Doublié, S.; Sawaya, M.R.; Ellenberger, T. An open and closed case for all polymerases. *Structure* **1999**, *7*, R31-R35.
52. Wendt, A.; Adhoute, X.; Castellani, P.; Oules, V.; Ansaldi, C.; Benali, S.; Bourliere, M. Chronic hepatitis c: Future treatment. *Clin. Pharmacol.* **2014**, *6*, 1–17.
53. Larrey, D.; Lohse, A.W.; de Ledingham, V.; Trepo, C.; Gerlach, T.; Zarski, J.P.; Tran, A.; Mathurin, P.; Thimme, R.; Arasteh, K.; *et al.* Rapid and strong antiviral activity of the non-nucleosidic NS5B polymerase inhibitor BI 207127 in combination with peginterferon α 2a and ribavirin. *J. Hepatol.* **2012**, *57*, 39–46.

54. Jacobson, I.; Pockros, P.J.; Lalezari, J.; Lawitz, E.; Rodriguez-Torres, M.; DeJesus, E.; Haas, F.; Martorell, C.; Pruitt, R.; Purohit, V.; *et al.* Virologic response rates following 4 weeks of fildesovir in combination with pegylated interferon α -2a and ribavirin in chronically-infected HCV genotype-1 patients. *J. Hepatol.* **2010**, *52*, S465–S465.
55. Rodriguez-Torres, M.; Lawitz, E.; Conway, B.; Kaita, K.; Sheikh, A.M.; Ghalib, R.; Adrover, R.; Cooper, C.; Silva, M.; Rosario, M.; *et al.* Safety antiviral activity of the HCV non-nucleoside polymerase inhibitor VX-222 in treatment-naive genotype 1 HCV-infected patients. *J. Hepatol.* **2010**, *52*, S14–S14.
56. Lawitz, E.; Rodriguez-Torres, M.; Rustgi, V.K. Safety and antiviral activity of ana 598 in combination with pegylated interferon α -2a plus ribavirin in treatment-naive genotype 1 chronic HCV patients. *J. Hepatol.* **2010**, *52*, 334A–335A.
57. Lawitz, E.; Jacobson, I.; Godofsky, E.; Foster, G.R.; Flisiak, R.; Bennett, M.; Ryan, M.; Hinkle, J.; Simpson, J.; McHutchison, J.; *et al.* A phase 2b trial comparing 24 to 48 weeks treatment with tegobuvir (GS-9190)/PEG/RBV to 48 weeks treatment with PEG/RBV for chronic genotype 1 HCV infection. *J. Hepatol.* **2011**, *54*, S181–S181.
58. Gane, E.J.; Stedman, C.A.; Hyland, R.H. Nucleotide polymerase inhibitor sofosbuvir plus ribavirin for hepatitis C. *N. Engl. J. Med.* **2013**, *368*, 34–44.
59. Wedemeyer, H.; Jensen, D.; Herring, R., Jr. Efficacy and safety of mericitabine in combination with PEG-IFN α -2a/RBV in G1/4 treatment naive HCV patients: Final analysis from the propel study. *J. Hepatol.* **2012**, *56*, S481–S482.
60. Pfefferkorn, J.A. Inhibitors of hcv ns5b polymerase. Part 1: Evaluation of the southern region of (2Z)-2-(benzoylamino)-3-(5-phenyl-2-furyl)acrylic acid. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2481–2486.
61. Wang, M. Non-nucleoside analogue inhibitors bind to an allosteric site on hcv ns5b polymerase. Crystal structures and mechanism of inhibition. *J. Biol. Chem.* **2003**, *278*, 9489–9495.
62. Ontoria, J.M.; Rydberg, E.H.; Carfi, A. Identification and biological evaluation of a series of 1*H*-benzo[*de*]isoquinoline-1,3(2*H*)-diones as hepatitis C virus NS5B polymerase inhibitors. *J. Med. Chem.* **2009**, *52*, 5217–5227.
63. Nyanguile, O.; Pauwels, F.; van den Broeck, W.; Boutton, C.W.; Quirynen, L.; Ivens, T.; van der Helm, L.; Vandercruyssen, G.; Mostmans, W.; Delouvroy, F.; *et al.* 1,5-Benzodiazepines, a novel class of hepatitis C virus polymerase nonnucleoside inhibitors. *Antimicrob. Agents Chemother.* **2008**, *52*, 4420–4431.
64. Nyanguile, O.; Devogelaere, B.; Fanning, G.C. 1a/1bsubtype profiling of nonnucleoside polymerase inhibitors of hepatitis C virus. *J. Virol.* **2010**, *84*, 2923–2934.
65. Tomei, L.; Altamura, S.; Migliaccio, G. Mechanism of action and antiviral activity of benzimidazole-based allosteric inhibitors of the hepatitis C virus RNA-dependent RNA polymerase. *J. Virol.* **2003**, *77*, 13225–13231.
66. Boyce, S.E.; Tirunagari, N.; Niedziela-Majka, A.; Perry, J.; Wong, M.; Kan, E.; Lagpacan, L.; Barauskas, O.; Hung, M.; Fenaux, M.; *et al.* Structural and regulatory elements of HCV NS5B polymerase—B-Loop and C-terminal tail—Are required for activity of allosteric thumb site II inhibitors. *PLoS ONE* **2014**, *9*, e84808.
67. Howe, A.Y.; Cheng, H.; Thompson, I.; Chunduru, S.K.; Herrmann, S. Molecular mechanism of a thumb domain hepatitis C virus nonnucleoside RNA-dependent RNA polymerase inhibitor. *Antimicrob. Agents Chemother.* **2006**, *50*, 4103–4113.

68. Davis, B.; Thorpe, I.F. Molecular simulations illuminate the role of regulatory components of the RNA polymerase from the hepatitis C virus in influencing protein structure and dynamics. *Biochemistry* **2013**, *52*, 4541–4552.
69. Ando, I.; Adachi, T.; Ogura, N.; Toyonaga, Y.; Sugimoto, K. Preclinical characterization of JTK-853, a novel nonnucleoside inhibitor of the hepatitis C virus RNA-dependent RNA polymerase. *Antimicrob. Agents Chemother.* **2012**, *56*, 4250–4256.
70. Caillet-Saguy, C.; Simister, P.C.; Bressanelli, S. An objective assessment of conformational variability in complexes of hepatitis C virus polymerase with non-nucleoside inhibitors. *J. Mol. Biol.* **2011**, *414*, 370–384.
71. Beaulieu, P. Recent advances in the development of NS5B polymerase inhibitors for the treatment of hepatitis C virus infection. *Expert Opin. Ther. Pat.* **2009**, *49*, 145–164.
72. Pflug, A.; Guilligay, D.; Reich, S.; Cusack, S. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* **2014**, *516*, 355–360.
73. Reich, S.; Guilligay, D.; Pflug, A.; Malet, H.; Berger, I.; Crepin, T.; Hart, D.; Lunardi, T.; Nanao, M.; Ruigrok, R.W.; *et al.* Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* **2014**, *516*, 361–366.
74. Brown, J.A.; Thorpe, I.F. Dual allosteric inhibitors jointly modulate protein structure and dynamics in the hepatitis C virus polymerase. *Biochemistry* **2015**, *54*, 4131–4141.

Chapter 2. Molecular Simulations to Delineate Functional Conformational Transitions in the HCV Polymerase

This chapter is reproduced in part from: Sesmero, E., Brown, J.A. and Thorpe, I.F, Journal of Computational Chemistry, accepted, DOI: 10.1002/jcc.24662

2.1 Introduction

Despite a wealth of enzymatic and structural information regarding the HCV polymerase there are gaps in our understanding of the mechanism of RNA replication and the role that enzyme conformational changes and dynamics play in this process (1). In this study we employed extensive molecular dynamics (MD) simulations to examine the conformational space explored by the free enzyme. Simulations were initiated from five different crystal structures representing HCV genotype 1b (see Table 2.1.). This genotype is estimated to be most prevalent worldwide, comprising 83.4 million cases and 46.2% of all HCV cases, as well as being the most prevalent source of infection in the United States (2). The results from the independent simulations were combined using the Weighted Histogram Analysis Method (WHAM) to generate Free Energy Landscapes (FELs) describing conformational transitions in the enzyme. The FELs reveal the location of distinct conformational states that likely have functional significance as well as the molecular interactions associated with these states. This knowledge will advance our understanding of the molecular mechanism that govern the conformational transitions during the RNA replication cycle and may ultimately allow us to better understand the mechanisms of action of HCV polymerase inhibitors.

2.2 Methods

2.2.1 Structure Preparation and Simulation Conditions

The PDB IDs for structures employed in this study are specified in Table 2.1. . Bound ligands were removed from 2WHO,(3) 3HHK,(4) 3CO9(5) and 2BRL(6) so that all five structures were simulated as free enzymes. Simulations of 2WHO were previously performed by Davis and Thorpe;(7) simulations of 3HHK, 3CO9 and 2BRL by Davis, Brown and Thorpe (8) and simulations of 1QUV(9) by Brown and Thorpe (10). These data were subjected to further analysis in this study. Most of our original studies were designed to probe the impact of ligands on the enzyme, and so we focused on PDB structures of enzyme-ligand complexes. Among this subset we chose to use structures that include ligands bound to each of the three distinct allosteric sites in the enzyme. 2WHO and 2BRL contain ligands bound to the NNI-2 and NNI-1 sites respectively while 3HHK and 3CO9 contain ligands at the NNI-3 site. Finally, 1QUV was selected because it is one of the few structures in the PDB of the HCV polymerase not bound to any ligand and we desired to see whether this would influence the results.

Structure 2BRL does not contain residues 22-35 or 148-152. Consequently, coordinates for these residues were taken from 2WHO, placed into the 2BRL structure and energy minimized before simulations were initiated as described previously (10). The C-terminal residues of the enzyme are not necessary for its function in vitro and are thought to be membrane associated in vivo (11). Thus, the C-terminal residues of all systems were truncated so that each contained a total of 531 residues.

Simulations were performed using the CHARMM (12) protein force field and NAMD (13) version 2.7 or later as detailed in our previous studies (7, 8, 10). Minimization of each solvated enzyme was carried out using NAMD for 1000 steps using the conjugate gradient method and

applying periodic boundary conditions. All bonds to hydrogen atoms were constrained using the SHAKE algorithm, and electrostatic interactions were calculated by the Particle Mesh Ewald method. A distance of 9 Å was used for the Coulomb cutoff and 11.4 Å for the non-bonded pair list. Minimization was followed by 2,500,000 steps of simulation in the NVT ensemble at 300 K using a 2 fs integration time step. Temperature was maintained via velocity reassignment every 100 steps, during which the positions of alpha carbons was restrained using force constants of 10 kcal(mol⁻¹ Å⁻²). Restraints were removed and simulations performed in the NPT ensemble at 300 K for 2,500,000 steps using Berendsen pressure coupling at 1.01 bar with a relaxation time of 100 fs. The temperature was maintained via velocity reassignment every 100 steps. Finally, production NVT simulations were performed using the same conditions as the original NVT simulations, except that temperature was maintained using a Langevin thermostat with a damping coefficient of 1 ps⁻¹ applied to heavy atoms and restraints were not employed. The pressure was checked every 10 ns to confirm that it remained within reasonable range of 1.01 bar. Table 2.1. provides a description of each system and lists how much MD data was generated for each. Simulations were carried out for each system until the temporal evolution of various structural metrics indicated that the different systems began to sample stable, stationary probability distributions. This procedure has been detailed in our previous studies (7, 8, 10). The evolution of molecular dynamics trajectories is subject to some degree of randomness, in particular for systems that employ a stochastic thermostat such as the Langevin thermostat employed in these studies. In addition to their different initial configurations, the independent trajectories are subject to different random fluctuations. Consequently some trajectories attained the criteria noted above at an earlier time point than others, leading to different total trajectory lengths.

For each simulation the final 200 ns of data was analyzed. This choice was made for two reasons. First, we wanted to use data that was close to the end of each trajectory to ensure that the probability distribution being sampled is likely to be stationary. Second, the underlying probability distributions reconstructed from the independent trajectories will depend on which data is included. Inclusion of more data from a given trajectory could in principle skew the results to be more similar to the distribution in that trajectory. Without *a priori* information that would cause us to have more faith in the data coming from a given trajectory than any other, we treat all the data on equal footing by including the same amounts of data for each. We chose 200 ns as a standard interval because this is the largest amount of equilibrated data available from the shortest trajectory (2WHO).

2.2.2 Metrics of Conformational Sampling

As stated in the Introduction, the enzyme is thought to adopt at least two distinct conformational states: closed and open. Consequently, it is essential to detect and analyze distinct states in the protein conformational ensemble. To achieve this goal we employed two different structural metrics. As one metric we selected the angle formed between enzyme subdomains, as this structural coordinate is directly related to the extent to which the enzyme is open or closed. The other metric is based on the largest amplitude mode obtained from Principal Component Analysis (PCA) of the 2BRL trajectory. This mode was used as a coordinate unto which we projected snapshots from each trajectory. Both metrics are described in more detail below.

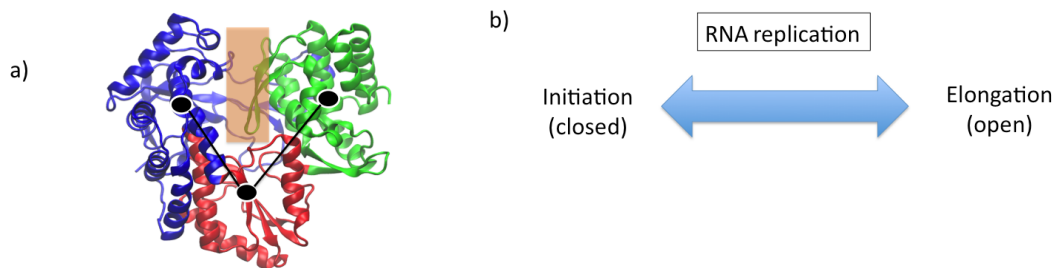


Figure 2.1. a) Structure of HCV polymerase showing the fingers, palm, and thumb subdomains colored blue, red and green respectively. The inter-domain angle formed at the intersection of the subdomains is shown in black. The location of the template channel is indicated by the translucent rectangle. b) Schematic describing the role of polymerase conformational states in the RNA replication cycle.

2.2.2.1 Inter-domain Angle

The inter-domain angle occurs at the intersection of the fingers and thumb with the palm (see figure 2.1) and is computed by determining the vectors connecting the centers of mass of each subdomain. The residue composition of each subdomain is defined according to Lesburg et al.:(14) the fingers contains residues 1 to 188 and 227 to 287, the palm contains residues 189 to 226 and 288 to 370, and the thumb contains residues 371 to 531. The centers of mass for each subdomain were computed according to these definitions and used to evaluate the inter-domain angle for each protein snapshot. Larger values of this angle are associated with more open conformations while smaller values are associated with more closed conformations.

2.2.2.2 Principal Components Analysis (PCA)

In order to perform PCA, we replaced each amino acid residue in the enzyme by the coordinates of its center of mass (COM) in every snapshot. The resulting trajectories of residue

COMs were then employed to compute covariance matrices for the protein fluctuations. C_{ij} , the covariance value between the COMs for residues i and j , is defined as:

$$C_{ij} = \frac{\langle (r_i - \langle r_i \rangle)(r_j - \langle r_j \rangle) \rangle}{|r_i - \langle r_i \rangle| |r_j - \langle r_j \rangle|} \quad (1)$$

where r_i is the instantaneous position of the COM of residue i , $\langle r_i \rangle$ is the average position of this site (obtained after least squares fitting of the coordinates in each snapshot to the initial structure) and the angle brackets denote ensemble averages. The covariance matrix is normalized as shown in equation (1) so that sites with completely positively correlated motion display C_{ij} values of 1, while sites with completely negatively correlated motions display C_{ij} values of -1. If two sites are not correlated or move in orthogonal directions, their C_{ij} values will be zero. Calculation of the covariance matrices was performed using the CORREL module of CHARMM.

To calculate the principal components (PCs), we diagonalized the covariance matrix using the CHARMM VIBRAN module: $\mathbf{V}^T \mathbf{C} \mathbf{V} = \mathbf{\Lambda}$. Each PC is defined by an eigenvalue λ and an orthonormal eigenvector \mathbf{V} . The eigenvectors describe the character (i.e. shape) of structural fluctuations sampled during the trajectories, while the eigenvalues describe the magnitude of these fluctuations. Typically the largest eigenvalue accounts for the majority of the overall fluctuations in the system, the second for the next largest contribution to the fluctuations and so on.

In order to define a PC-based coordinate appropriate for construction of free energy landscapes, we first identified 2BRL as the system with the greatest overall fluctuations among all the protein simulations. We then selected the 2BRL eigenvector associated with the largest eigenvalue. One reason for these choices is that 2BRL is most likely to encompass fluctuations

present in the other systems since it has the largest degree of overall fluctuations. In addition, the largest amplitude PC from 2BRL is likely to be representative of fluctuations occurring in each of the different protein trajectories. This PC corresponds to a motion primarily associated with opening and closing of the enzyme, in which the fingers and thumb move towards and away from one another while the palm acts as a fulcrum (see movie in supplemental material).

Individual trajectory snapshots were projected unto this PC, with the numerical values resulting from the projection defining our desired structural metric. The projection P_{lk} of a specific snapshot R_k unto PC eigenvector V_l is computed by evaluating the dot product between the eigenvector and the normalized coordinates: $P_{lk} = V_l \cdot R_k / |R_k|$, where R_k represents the difference between the instantaneous value of the coordinates for snapshot k and the average structure. The result of this calculation is a measure of the degree of overlap between the conformational space represented by a specific snapshot and that described by the PC eigenvector. Larger values of P_{lk} indicate greater overlap between the “direction” inherent in a trajectory snapshot and the PC eigenvector. Thus, projections of the trajectories unto this PC provide an additional coordinate that is useful in characterizing the protein conformational space.

2.2.3 Root Mean Squared Fluctuations

The root mean square fluctuation (RMSF) for the center of mass of each residue was calculated to determine the variability of the enzyme structure with respect to its average in a particular ensemble. The RMSF for the i th residue was computed by subtracting the average position of the center of mass of the residue \bar{r}_i from its instantaneous position r_i as shown below:

$$RMSF_i = \sqrt{\frac{\sum_i^N (r_i - \langle r_i \rangle)^2}{N}}$$

(2)

Above, N denotes the total number of data points. RMSF provides a measure of the local flexibility or conformational diversity of the enzyme.

2.2.4 Free Energy Landscapes

In order to describe the conformational space adopted by the enzyme, we took the probability distribution $P(\eta_1, \eta_2)$ obtained from our trajectories, inverted it to generate free energy profiles via $\Delta F = -k_B T \ln P(\eta_1, \eta_2)$ and projected the free energy ΔF unto the coordinates described above. Although the inter-domain angle is an intuitive structural coordinate, the lowest frequency PC from 2BRL provides an unbiased way to delimit the enzyme conformational space that only depends on the dominant fluctuations occurring in the trajectory. Projection of structural data unto this mode may provide additional information because PCs innately represent the intrinsic structural fluctuations of the enzyme. Thus, this mode may incorporate fluctuations that are not obviously related to the conformational transition, but nonetheless play a role in it. To combine the results of the five independent trajectories we used the Weighted Histogram Analysis Method (WHAM) as described below.

2.2.5 Weighted Histogram Analysis Method (WHAM)

WHAM is a well-known approach for incorporating data from multiple simulations in order to optimally describe a probability distribution, under the assumption that each simulation

provides an independent estimate of this distribution. One of the underlying goals for performing WHAM is to reproduce the potential of mean force (PMF) from a probability distribution obtained from molecular simulations. WHAM is an evolution of single histogramming techniques developed by Ferrenberg and Swendsen(15) and extended to the case of multiple histograms by these authors in later studies.(16, 17) WHAM defines a weighting function to combine information from independent trajectories such that the statistical error is minimized among the different data sets. Other approaches have been introduced more recently to optimally combine data from independent simulations, such as the lambda windows employed for free energy perturbation calculations (18, 19). However, an advantage of WHAM is that it can be easily extended to systems that do not involve alchemical perturbation or for cases in which one wishes to match probability distributions across multiple dimensions. The method only requires that one can define a probability distribution (i.e. as a histogram) along the coordinates of interest. The essential equations of WHAM are as follows:(20, 21)

$$\begin{aligned}
\rho_o(\eta) &= C \sum_{i=1}^N \frac{n_i}{\sum_{j=1}^N n_j e^{f_j - \beta W_j(\eta)}} \rho_i^b(\eta) \\
e^{-f_j} &= \int d\eta e^{-\beta W_j(\eta)} \rho_o(\eta) \\
&= C \int d\eta \sum_{i=1}^N \frac{n_i e^{-\beta W_j(\eta)}}{\sum_{j=1}^N n_j e^{f_j - \beta W_j(\eta)}} \rho_i^b(\eta)
\end{aligned}
\tag{3}$$

Above, $\rho_o(\eta)$ represents the underlying unbiased probability distribution as a function of some coordinate η , ρ_i^b is the (possibly biased) distribution from simulation i , N is the number of simulations, n_i is the number of data points in the i th simulation, f_j is the free energy coefficient (or thermodynamic weight) of the j th simulation and W_j is the biasing potential applied to the j th

simulation. C is an undetermined constant that does not impact free energy differences evaluated using this approach. The aim of WHAM is to shift the probability distributions of individual simulations so that they agree with one another as closely as possible, while minimizing the error in the underlying probability distribution $\rho_o(\eta)$. This goal is accomplished via the f_j values shown above. The reader will note that $\rho_o(\eta)$ depends on the f_j values, but that these coefficients also depend on $\rho_o(\eta)$. Thus, the equations in (3) must be solved iteratively in order to determine a solution. First, a value is assumed for each f_j (these are typically set to zero initially) and an estimate made for $\rho_o(\eta)$. The computed $\rho_o(\eta)$ is then employed to update values of each f_j and the cycle is continued until the f_j values no longer change appreciably, at which point the process is halted.

While it has become common practice to use WHAM to combine simulations subject to biasing potentials,(20-22) it is not widely recognized that use of such energy functions is optional when applying the method (16, 17, 23). Biasing potentials only serve to encourage a system to visit regions of the FEL that may be higher in energy and thus not well sampled (24, 25). Because there were no explicit biasing potentials applied to our simulations, the W_j in equation 3 were set to zero. The resulting probability distribution and free energy coefficients are then written as:(16, 17, 23)

$$\begin{aligned}\rho_o(\eta) &= C \sum_{i=1}^N \frac{n_i}{\sum_{j=1}^N n_j e^{f_j}} \rho_i(\eta) \\ e^{-f_i} &= C \int d\eta \sum_{i=1}^N \frac{n_i}{\sum_{j=1}^N n_j e^{f_j}} \rho_i(\eta)\end{aligned}\tag{4}$$

In writing these equations we represent the probability distributions from the independent MD simulations as being distinct $\rho_i(\eta)$. Our reason for doing so is that any realistic simulation is

finite and thus: i) subject to statistical errors and/or ii) preferentially explores conformations that are concentrated in the vicinity of the initial point. These effects may result in disparities between the probability distributions evaluated from independent trajectories. Such effects can ultimately be ascribed to insufficient sampling, since systems that are truly ergodic will display probability distributions that match exactly. The result of these effects is a shift in the probability distribution obtained in one simulation relative to that obtained in another. Such effects can thus be corrected via WHAM to ensure that the probability distributions in independent simulations match as closely as possible (15, 16, 23). One can thus use WHAM to correct for limited sampling and distinct initial conditions in independent simulations of a given system.

Table 2.1. Simulated systems.						
PDB ID	Number of residues in PDB structure	Ligand location in PDB structure [a]	Simulation time (ns)	Total No. of simulated atoms	Inter-domain angle in PDB structure	$PC_{2BRL} \times 10^{-1}$ [b]
2WHO	531	NNI-2	400	60,304	71.3	1.82
1QUV	556	None	1000	60,391	70.1	1.55
3HHK	563	NNI-3	900	68,550	69.9	1.56
3CO9	562	NNI-3	710	62,397	69.4	1.52
2BRL	531	NNI-1	800	60,265	72.2	1.88
[a] NNI: Non-nucleoside inhibitor site						
[b] Projection of original PDB structure on PC_{2BRL}						

2.3 Results and Discussion

2.3.1 The free enzyme samples a broad conformational distribution

The free energy landscape obtained using the combined data from all five simulations is shown in figure 2.2 . Landscapes generated for each individual trajectory are provided as supporting information in appendix A. It is apparent that the free enzyme can explore diverse conformations. The conformational states adopted include those that are very open as well as states that are quite closed, spanning a range of approximately 16 degrees with respect to the inter-domain angle. These diverse conformations occur despite the absence of other components of the replication complex, including nucleotides and the RNA template, which might be anticipated to facilitate adoption of these states. In particular, one might expect that open states required for processive elongation would be more likely in the presence of the bound template and nascent RNA duplex. The observed conformational diversity suggests that the structural transitions required for enzyme function occur via a conformational selection mechanism. Consequently, it is likely that other components of the replication complex will preferentially stabilize one or more of the observed conformations at the appropriate point in the replication cycle, but are not necessarily required for these states to be adopted.

The conformational space in the landscape can be divided into three distinct zones. These are classified based on the inter-domain angle as follows: closed (inter-domain angle $< 69^\circ$), intermediate ($69^\circ < \text{inter-domain angle} < 74^\circ$) and open (inter-domain angle $> 74^\circ$). Each zone describes a population in which the enzyme primarily adopts a closed, intermediate or open conformation respectively. As noted previously, the closed and open conformations are thought to be required for the initiation and elongation stages, respectively, of RNA replication (26-31). The intermediate state appears to be necessary in order to convert between these two

conformations. It is interesting to note that all of the initial crystallographic coordinates for the simulations are located in the vicinity of the local minimum found in the intermediate state. This can be clearly seen if the locations of the crystal structures are plotted with respect to the FEL coordinates shown in figures 2.2 and 2.3 (see supporting material in appendix A). To more closely examine this issue, we computed inter-domain angles for a random sample of HCV polymerase structures currently available in the PDB (see appendix A). In every case we observe that the inter-domain angle for these structures also falls within the intermediate interval described above. It is possible that conditions amenable to crystallization may preferentially stabilize structures with these domain angle values, thus leading to a preponderance of crystal structures occurring in the local minimum represented by the intermediate region. While previous studies have suggested the existence of open and closed states in the HCV polymerase,(26, 27, 32) our work is the first to show the relationship between different states in the context of the free energy landscape and to provide detailed structural information regarding the identity of these states.

The conformational distribution in figure 2.2 is also apparent if one examines one-dimensional (1D) free energy profiles along the individual structural coordinates (see figure 2.3). The 1D profile for the inter-domain angle clearly shows the presence of a minimum corresponding to each of the three states identified above, with a higher barrier separating the intermediate from the open state. This barrier is also reflected in the 1D profile along PC_{2BRL}. However the latter profile also shows that, along this coordinate, the intermediate state can be further subdivided into two substates. Thus, there are a total of four distinguishable states that occur in the enzyme. The presence of an obligate intermediate between the closed and open states suggests a sequential order of transitions between closed, intermediate and open that may

reflect the sequence of conformational states that occurs during the RNA replication cycle. Thus, these free energy landscapes are likely to provide information about the molecular forces and interactions that facilitate transitions between functional conformational states. These observations are consistent with kinetic studies that indicate the initiation and elongation stages of replication are associated with conformational states that display distinct kinetic signatures.

Studies by Tomei et al.(33) suggest that a slow conformational change in the enzyme precedes the elongation stage of RNA replication. This process may represent the transition from initiation to the elongation stage, but the precise nature of this phenomenon is unknown (33). Later work by Reich et al.(34) indicate that a large-scale tertiary structural change is required to convert from initiation to elongation. The results of both studies are consistent with the larger barrier we observe between the intermediate and open states on the landscape representing the structural change required to convert from initiation to elongation.

Reich et al. also observed that product release by the polymerase proceeds in a stepwise fashion. This was attributed to the adoption of a series of four states that can be differentiated based on their distinct contributions to binding progress curves. At saturating nucleotide concentrations, Reich et al. observed decreased rate constants associated with the first three of these states. These authors ascribed this observation to reduced fluctuations of the template due to stabilization induced by the presence of nucleotides. This proposal is consistent with the possibility that nucleotides stabilize preferred locations for template occupancy and thus induce reduced positional fluctuations.

In contrast to the first three states, the rate constant associated with the fourth state increased in a hyperbolic manner at saturating nucleotide concentrations. Reich et al. interpreted

these observations to indicate that the three initial states are directly involved in RNA replication via template repositioning, while the fourth reflects product release.

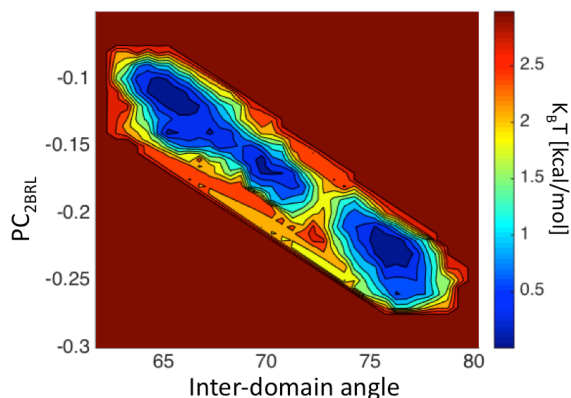


Figure 2.2. Free energy landscape obtained from combining all trajectories and projected unto the inter-domain angle (x-axis) and the largest amplitude PC from the 2BRL simulation (y-axis). Contours in the plot are displayed at intervals of $k_B T / 2$

The observations of Reich et al. are consistent with the free energy landscapes presented above, and in particular the 1D profile along PC_{2BRL} shown in figure 2.3 . The three initial states identified by Reich et al. are consistent with the three minima towards the left-most (i.e. more closed) region of the x-axis in the lower panel of figure 2.3 . Thus, these minima may correspond to the states associated with repositioning of the RNA template as suggested by Reich et al. It is conceivable that relatively low barriers between these states allow for facile repositioning of the template once it is bound.

Consequently, the fourth state reported by Reich et al. likely corresponds to the right-most (i.e. most open) minimum shown in this figure. This situation is consistent with product release occurring from the most open state. This relationship seems reasonable given that newly synthesized RNA is most likely to be released from an open enzyme conformation since this state is required for processive elongation. However, the studies of Reich et al. are unable to

supply details about the relative stability of these states or the molecular interactions that occur within them as is provided in the present study.

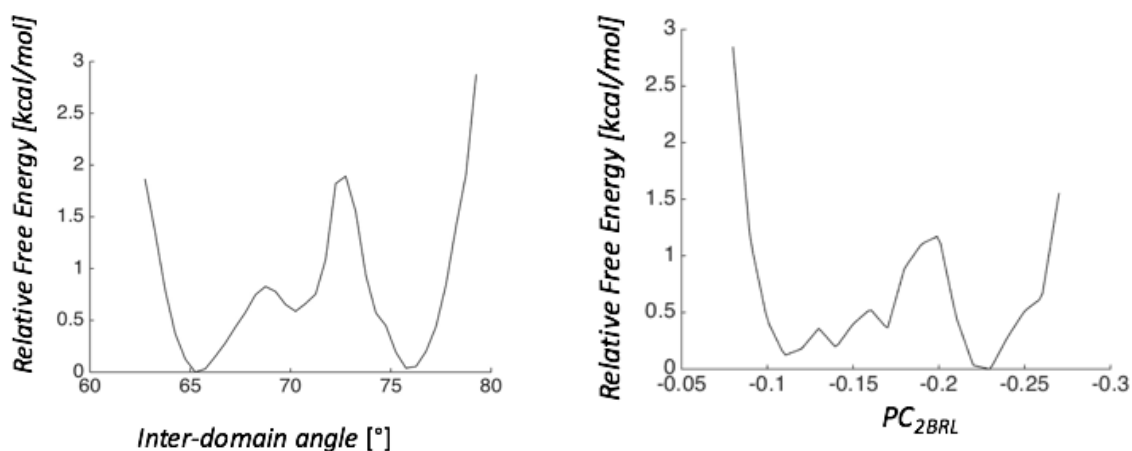


Figure 2.3. One-dimensional free energy landscapes obtained from combining all trajectories and projected unto a) the inter-domain angle and b) the largest amplitude PC from 2BRL (PC_{2BRL}). In both plots the left side corresponds to more closed states while the right corresponds to more open states.

Moreover, the congruence between the results presented in this work and the studies described above suggests that our simulations describe conformational ensembles that are relevant to understanding global conformational changes that mediate the RNA replication cycle in HCV polymerase and the molecular signatures associated with these conformations. These topics are described in greater detail below.

2.3.2 Conformational states exhibit distinct characteristics

It is apparent that the closed and intermediate states shown in figure 2.2 are quite close in energy and would appear to be part of the same broad minimum. However, we provide evidence below that all three states are distinct with their own unique structural fluctuations, correlated motions and patterns of hydrogen bonding. In order to examine these properties, we constructed

trajectories only comprised of the data within a given zone for analysis. Below, we note that the center of mass trajectories were employed for RMSF and covariance analyses.

2.3.2.1 Structural differences

The primary structural differences among the three conformational states lie within the thumb and fingers. The most marked difference is movement of the thumb away from the fingers in the open conformation relative to the closed conformation such that the inter-domain angle increases. Another less pronounced change occurs within the fingers involving the outward motion of regions within the fingers in the open conformation. Structural alignment of the open and closed conformations (see figure 2.4) reveals the altered positions of numerous components of the thumb and fingers. The displaced components include a helix that encompasses functional region III (residues 401-404) in the thumb, which has been noted to have a role in binding nascent RNA duplex. Thus, movement of this region would be needed to convert from the initiation stage of RNA synthesis to the elongation stage (which requires an open conformation to allow the exit of the nascent RNA duplex). Displaced regions within the fingers include functional regions I and II (residues 91-94 and 168-183, respectively) and motif G (residues 95-99). All three regions are thought to be involved in binding RNA template.

A noticeable displacement was also seen in the beta flap region (residues 443-454) that is implicated in correct positioning of the template (35). As we consider the closed, intermediate and open conformations in turn we observe that the beta flap moves further from the center of the enzyme. This observation is consistent with its role as a platform that helps to stabilize the initiation stage of RNA replication (28, 36, 37). The beta flap is likely to interact with a GTP needed to form the initiation complex in the closed conformation.

However, the beta flap would need to move out of the way of the nascent RNA duplex when transitioning to the intermediate and open conformations for elongation, consistent with the shifted positions we observe in the intermediate and open states.

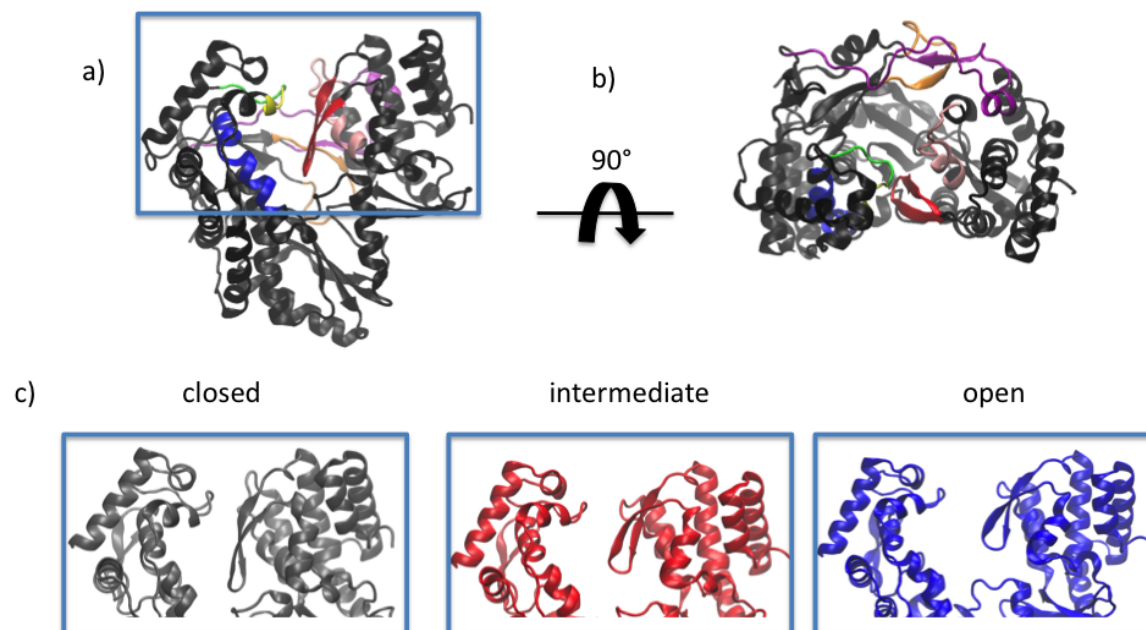


Figure 2.4. Motifs, functional regions and characteristic elements of NS5B affected in the conformational transitions. Motif G is shown in green, functional region I in yellow, functional region II in blue, functional region III in pink, delta 1 loop in purple, delta 2 loop (which overlaps with motif F) in orange, beta flap in red. (a) front view, (b) top view. (c) Zoomed image of regions within rectangle in a) showing relative positions of fingers and thumb subdomains in closed, intermediate and open conformations.

A displacement is also seen in the fingertips region formed by delta loops 1 and 2 (residues 9 to 41 and 142-157, respectively, see figure 2.4). The fingertips are characteristic of RNA-dependent RNA polymerases and contribute to create a fully enclosed active site. Allosteric inhibitors that bind to the NNI-1 pocket in the thumb (6) have been reported to destabilize the enzyme by disrupting the interactions between the delta 1 loop and the thumb, thus preventing the transition

to the open state (8). Consequently, stabilization of this loop is likely to be a key factor in allowing the enzyme to convert from the closed to the intermediate and open states.

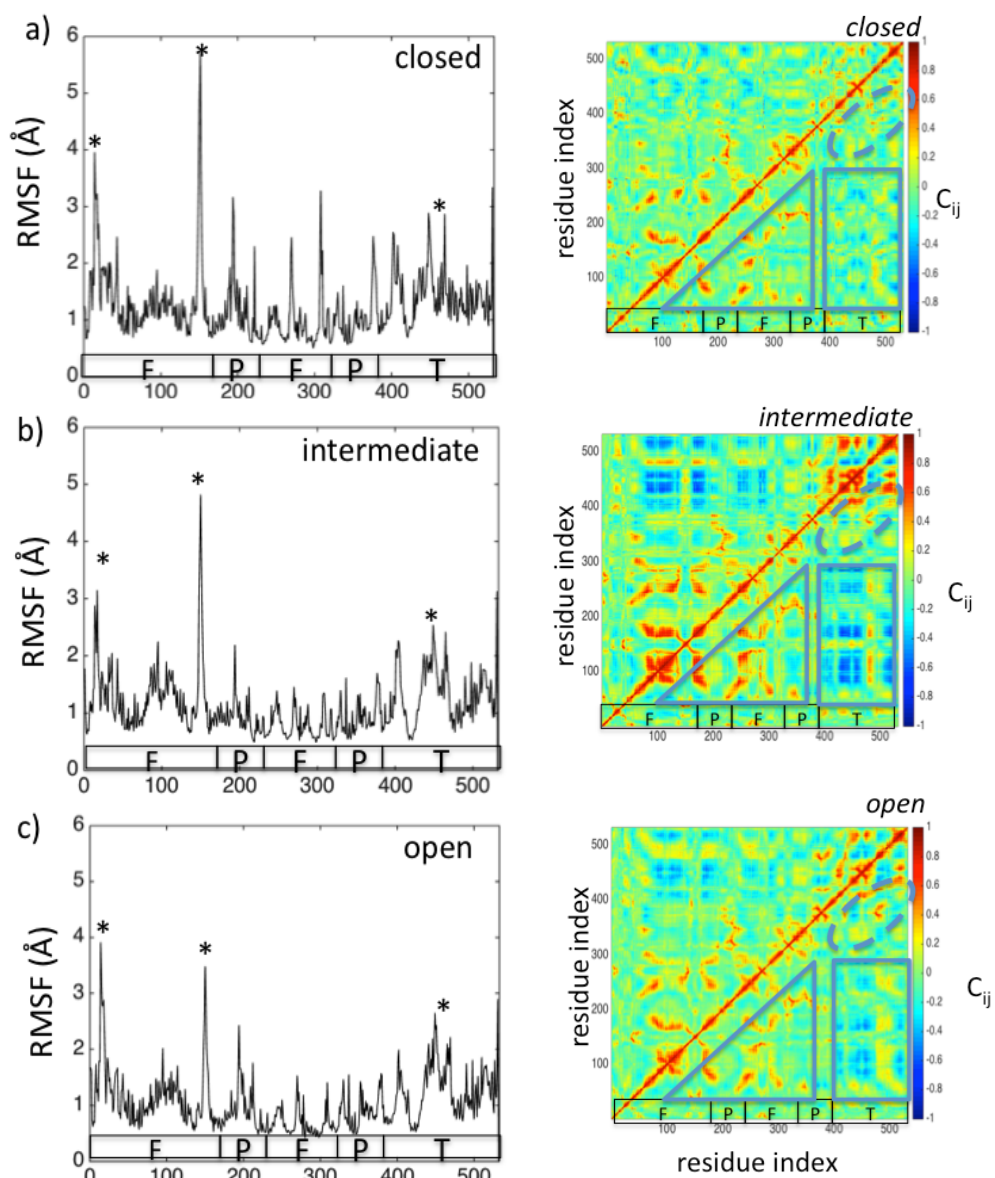


Figure 2.5. RMSF (left) and covariance (right) plots generated from conformational ensembles in the open, intermediate and closed states.

2.3.2.2 Root Mean Squared Fluctuations (RMSF)

The conformational diversity of individual protein residues within each zone can be assessed via evaluating the RMSF for each ensemble of structures. In order to compute the RMSF, each protein snapshot was first aligned to the average structure computed from the ensemble via least squares fitting of the coordinates. In the closed state high peaks of the RMSF are observed in each subdomain of NS5B, indicating that each subdomain exhibits a fairly broad distribution of conformations (see figure 2.5). These values are reduced significantly when the enzyme adopts the intermediate state. RMSF values in the intermediate and the open states are more similar to one another compared to those present in the closed state, although there appear to be additional slight decreases in the open state relative to the intermediate. In the open state, the enzyme is more rigid due to enhanced interactions between distant regions of the enzyme. Thus, there is a reduced propensity for enzyme fluctuations in the open state compare to the other two states (see also section 2.3.2.4 below).

The peaks in the RMSF plot for the closed state highlight key residues important to the initiation stage of the RNA replication cycle. The peak around residue 14 corresponds to the flexible delta 1 loop noted previously that plays a role in converting from the closed to the intermediate state.

The peak around residue 150 in the RMSF plot corresponds to the delta 2 loop. The delta 2 loop shows a decrease in flexibility going from closed to intermediate and open conformations. This loop contains motif F, which has a role in binding incoming NTPs and RNA (27, 36-39). Decreased flexibility in the intermediate and open states may reflect a decreased need for the enzyme to promote binding of NTPs and RNA, given that these species should have already

become bound to the enzyme during the initiation stage of the RNA replication cycle that occurs in the closed conformation.

A region that is highly flexible in all conformational states occurs in the vicinity of residue 448 that corresponds to the beta flap (residues 443-454). There are also peaks around residue 465 and 469 that are part of an alpha-helix spatially proximal to the beta-flap. As we note above, the beta-flap is required for correct positioning of the RNA template. Thus, a high degree of flexibility in this region may be required to allow the enzyme to modulate the template location throughout the RNA replication cycle.

2.3.2.3 Covariance Analysis

Covariance maps computed for the different conformations (see figure 2.6) indicate that there are three main regions with distinct correlation patterns in each state: 1) within the thumb (inside the dashed oval), 2) between the thumb and the fingers (inside the rectangle) and 3) between the fingers and palm (inside the triangle).

The rectangular outlined area displays many negative correlations in the intermediate and open state that are less evident in the closed state. Inspection of the PCs computed from the intermediate and open ensembles indicate this correlation results from the thumb and fingers moving in opposite directions with respect to the average structure, consistent with the opening and closing motion necessary for altering the conformational state of the polymerase (see movies in supporting material). The oval and triangular regions contain numerous residues that display positive correlations in all three states, though these are more intense for the intermediate state. The positive correlations in the oval region result from rigid-body motion of residues within the

thumb as it shifts position with respect to the fingers. Positive correlations in the triangle region are a result of residues in the palm and thumb moving in concert due to the proximity of the two subdomains. In general the patterns of correlated motion are consistent with the changes in flexibility reported above as well as with large-scale movement of the thumb and fingers relative to each other along the open to closed direction.

We note that the intermediate state displays the greatest intensity of correlations in all three regions. This is due to the fact that this state has enhanced fluctuations along the structural coordinates that define the open-close transition. When viewing the 1D version of the FEL constructed with respect to the PC coordinate (see figure 2.3), it is apparent that the intermediate state contains two separate minima (see discussion in section 3.1). The dominant modes in the intermediate state arise from fluctuations between these two minima (data not shown).

2.3.2.4 Large amplitude motions

We visualized the motion corresponding to the largest amplitude mode for each one of the conformational states (see movies in supplementary information). The largest amplitude PCs for the open and intermediate conformational states involve motions in which the thumb and fingers move towards and away from each other. In contrast to the open and intermediate states, the largest amplitude PC for the closed state shows more local movements, especially in the thumb and the fingertips (delta 1 and delta 2 loops). It is noteworthy that while the intermediate and open states are less flexible overall than the closed state, large scale motion in the former two states is concentrated along the direction of the open-closed coordinate. This observation suggests that in the open and intermediate states the overall enzyme motions are being funneled along the open-closed direction.

2.3.2.5 *Hydrogen bonding patterns*

Hydrogen bonds (H-bonds) play a crucial role in defining the structure and dynamics of biomolecules. Thus, it is not surprising that we see patterns of H-bonds that correspond to distinct conformational states. We observed two main categories of H-bonds that differ between the conformations: 1) H-bonds that are unique to one conformation and 2) H-bonds that are specific to two conformations. H-bonds in each group are likely responsible for stabilizing the conformation(s) in question. In the latter group we further observe the following subcategories: a) H-bonds that are found in the closed state, break in the intermediate and reform in the open (and are thus common to the closed and open states), b) H-bonds that are formed after converting from the closed to the intermediate and maintained in the open (and thus common to the intermediate and open states), c) H-bonds formed in converting from the open to the intermediate and maintained in the closed state (thus common to the intermediate and the closed). A full list of these H-bonds is provided in appendix A.

H-bonds that link the fingers and thumb at the front of the enzyme are of particular interest, as the lengths of these bonds are indicators of the enzyme conformation. For example, when the distance between the H-bond connecting ASP 444 and ARG 109 is plotted as a function of time, it is directly correlated to fluctuations in the inter-domain angle (see appendix A). Two of these H-bonds (ASP 444-ARG 109 and CYS 451-LEU 91) were previously reported by Davis et al.(8) to play a role in differentiating conformational states in the HCV polymerase. Our present results support this observation and allow us to identify an extensive list of H-bonds characteristic to each conformation.

The open state displays more unique H- bonds at the palm/fingers interface and the thumb/fingers interface than either of the other states. When the enzyme transitions to the open

state these additional H-bonds stabilize the enzyme structure, resulting in a more rigidified system. This occurrence is apparent in the decreased RMSF values obtained for the open state (see figure 2.5).

In the intermediate state, we see H-bonds formed involving residues 310, 312, 313, 316 and 325 that are part of motif C. This motif is in the palm and has a role in binding the magnesium ions required for the nucleotide transfer. The formation of these H-bonds help to rigidify this region of the enzyme (see figure 2.5), decreasing fluctuations of the magnesium ions in a way that may facilitate catalysis.

Two of the H-bonds formed during the transition from closed to intermediate and maintained in the open state (HSD 402 – GLU 398 and ARG 401 – GLU 18) include residues 401 and 402 that are part of functional region III. The presence of these H-bonds stabilizes this structural element, which plays a role in binding the nascent RNA duplex. Stabilization of functional region III likely facilitates interactions with the RNA duplex in the open state, allowing the enzyme to engage in processive elongation of the daughter strand.(40)

We find that the H-bonds are clustered into specific networks that are distinct to each classification. These often play a role in connecting the enzyme subdomains to one another (see appendix A). Perturbing H-bonds connecting subdomains would be expected to be particularly effective in inhibiting conformational transitions in the enzyme. This is because the movement of subdomains relative to one another is key in defining the global enzyme conformation. Of particular interest are H-bonds between the fingers and the thumb at the back of the enzyme (e.g. LYS 491 – HSD 33 or ALA 396- SER 27) as well as at the front of the enzyme (e.g. ASP 444 – ARG 109, CYS 451 – LEU 91 or ASN 406 – HSD 95). The former group includes interactions between the delta 1 loop and thumb that facilitate transitions to the open state (8). The latter

group includes H-bonds that we have shown to be highly correlated to fluctuations in the inter-domain angle (see discussion above and appendix A). This observation suggests that eliminating these H-bonds would be likely to reduce or eliminate fluctuations in the inter-domain angle. This would prevent the enzyme from making transitions between conformational states and result in cessation of the RNA replication cycle.

These H-bonds highlight key residues that are involved in modulating the conformational distribution present in the HCV polymerase. These locations could be targeted to induce desired conformational changes. For example, eliminating H-bonds unique to a given state is likely to specifically decrease the stability of that state. This occurrence may be relevant to understanding the mechanisms of action for inhibitors that target the enzyme. It is known that allosteric inhibitors that bind to different locations on the HCV polymerase are able to reduce enzyme activity despite being distant from the active site. Only recently have molecular mechanisms underlying the impact of these inhibitors been elucidated (7, 8, 10). It is possible that at least some of these inhibitors mediate their allosteric effects by altering the H-bonds as described above.

2.3.3. Potential relevance for therapeutics

The presence of multiple enzyme conformational states may be relevant to the search for therapeutics that target the HCV polymerase. Each conformational state could reflect a different target for therapeutic molecules such as polymerase inhibitors. It is well known that the polymerase has multiple locations at which diverse small molecule inhibitors are known to bind. These include the active site as well as at least three distinct allosteric sites on the enzyme (3-6).

The different conformational states may exhibit altered affinities for inhibitors at these distinct sites. For example, prior studies from our group indicate that inhibitors bound at the allosteric NNI-2 site in the thumb preferentially stabilize conformational states of the enzyme that are more closed.(7, 8, 10) Given that ligand binding can alter conformational preferences, it is conceivable that distinct conformational states can also modulate inhibitor affinity.

This result could be relevant to understanding one source of diminished inhibitor efficacy. Some allosteric inhibitors may be less effective than desired because they only target a subset of the available conformational states. This opens up the possibility of using combinations of small molecules targeting multiple conformational states in order to reduce enzyme activity to very low levels. In this context, even molecules that are not very potent on their own could be combined into an effective mixture if they target complementary conformational states. If such a strategy is employed, it is feasible that some HCV polymerase inhibitors that have displayed limited efficacy to date could be rescued as potential therapeutics. Thus, it may be possible to reduce enzyme activity effectively by targeting distinct conformations, thereby disrupting multiple points in the replication cycle.

Given their high degree of structural and functional similarity, the pattern of conformational fluctuations we observe for the HCV polymerase may occur in other viral polymerases as well. Our own studies strongly suggest that the structural similarities shared between polymerases from Foot-and-Mouth Disease Virus,(41) West Nile Virus (42) and Dengue Virus (43, 44) with the HCV polymerase extend to the presence of multiple allosteric inhibitor sites on the enzyme surface (45). Thus, it is conceivable that these enzymes can also adopt multiple conformational states along the RNA replication cycle. If this is the case, the ability of small molecules to stabilize specific conformational states (7, 8, 10) and consequently

block transitions between conformations may represent a general strategy that could be employed to reduce the activity of these enzymes.

2.3.4. Insights into the RNA replication cycle

When one compares the schematic of the RNA replication cycle shown in figure 2.1 with the 1D free energy landscapes shown in figure 2.3, it is apparent that the cycle represented in the schematic could be described by repeated traversals of the 1D landscapes: first in one direction and then in another. Thus, the polymerase could travel from the closed state to the open state via the intermediate and then reverse direction to adopt the closed state and begin the cycle anew. However, it is also possible that the cycle incorporates a different intermediate between the open and closed states that was not uncovered in our current analysis. This may occur, for example, if the presence of bound ligands such as nucleotides or RNA template significantly perturbs the free energy landscape.

Our work may also inform us about the conformational states that modulate RNA replication in other viral polymerases. Most of the studies done in this area to date have focused on detailed structural rearrangements of the active site rather than global conformational states important for the polymerase function. The most relevant work in this arena includes studies of the RNA polymerase from poliovirus by Gong and Peersen (46). These authors used crystal structures containing elongation complexes to identify distinct conformations of the polymerase active site occurring during RNA replication. These results were used to propose a cycle of conformational transitions for the active site that encompasses six distinct states the authors termed $S_1 - S_6$. The initial state S_1 displays an open, empty enzyme active site. In S_2 , the active

site remains open but includes bound nucleotide. A conformational change then occurs to lead to S_3 in which the priming nucleotide and substrate become aligned for nucleotidyl transfer as the active site adopts a catalytically competent conformation. While the authors did not observe S_3 directly, this state has been captured in previous crystallographic studies of the Norwalk virus RNA polymerase by Zamyatkin et al.(47) S_4 represents a conformation immediately after the nucleotidyl transfer has occurred, but before the geometry of the active site has changed appreciably. As the enzyme complex proceeds to S_5 the catalytic geometry becomes distorted. Finally, the hypothetical state S_6 represents a translocation intermediate that is required to restore the enzyme to S_1 in order to accommodate the next nucleotide.

Gong and Peersen describe these conformational transitions as involving an anti-correlated motion of the fingers and palm in which the two subdomains move towards and away from each other. They noted that this fluctuation is distinct from that which modulates the global open-close transition of the enzyme and which is also described in their work. The latter motion is consistent with the low frequency motions displayed in our own studies for the intermediate and open conformational states (see movies in appendix A). As described previously, it involves an anti-correlated movement of the thumb and fingers towards each other such that the inter-domain angle is modulated. Given that the states $S_1 - S_6$ identified by these authors involve only local structural rearrangements, it is likely that these states are encompassed within one or more of the conformational states identified in the present study since the latter involve global conformational reorganization. Since the sequence of events involved in RNA replication must form a closed loop in order for the process to occur repeatedly, we propose that the entire cycle of active site states $S_1 - S_6$ can occur within the closed and open conformational ensembles identified in this work. The initiation of RNA synthesis is likely to begin with S_1 in the closed

ensemble as defined by inter-domain angles $< 69^\circ$. It would be necessary for the enzyme to then adopt the intermediate state ($69^\circ < \text{inter-domain angle} < 74^\circ$) before commencing processive elongation. After visiting the intermediate state, the enzyme must adopt an open conformation (inter-domain angle $> 74^\circ$) in order to allow elongation of the daughter RNA strand. This transition must necessarily encompass states $S_2 - S_6$. Once this is accomplished the enzyme must repeatedly convert between $S_1 - S_6$ in the open state in order for the cycle of nucleotide addition to proceed during processive elongation. Consequently, the local active site rearrangements theorized by Gong and Peersen are likely to occur for multiple global enzyme conformations.

Gong and Peersen suggest that the $S_1 - S_6$ sequence is likely to be common to viral RNA polymerases that employ (+)-sense templates, such as the poliovirus and HCV polymerases. This proposal is further supported by more recent work by Shu and Gong (48). These authors used seven different crystal structures of the RNA polymerase from a human enterovirus bound to RNA template to characterize different stages of the nucleotide addition cycle. They identified structures corresponding to each of local active site conformations $S_1 - S_6$ identified by Gong and Peersen. Given the similar observations for two different viral species, we believe it is highly likely that Gong and Peersen are correct in their suggestions regarding commonalities in the RNA replication cycle mediated by (+)-sense viral RNA polymerases. This would suggest that the conformational ensembles we have identified for the HCV polymerase are functionally relevant to other viruses as well.

2.4 Summary

Our present studies demonstrate that the HCV polymerase can explore diverse conformations that appear relevant to the RNA replication cycle. This phenomenon is observed even in the absence of other components of the replication complex such as nucleotides or the RNA template. This observation suggests the enzyme adopts a conformational selection mechanism, in which functionally relevant conformations are accessible to the free enzyme. Distinct conformations may be preferentially stabilized by specific binding partners, such as components of the replication complex or small molecule inhibitors, but are not necessarily required for the conformation to be observed. The conformational distribution we observe is consistent with previous studies that identify four conformational states with unique kinetic signatures (34). While that work was informative in demonstrating the presence of distinct conformational states, our present investigation reveals molecular details regarding intrinsic fluctuations of the polymerase as well as the structural ensembles that allow these states to be adopted.

Each state is characterized by specific patterns of flexibility, large amplitude fluctuations and correlated motions. Of particular interest are the unique hydrogen bond networks that appear to play a key role in stabilizing each state and in facilitating transitions between them. Given that the RNA replication cycle is thought to be highly similar among viral polymerases (particularly those that employ (+)-sense templates) our observations are likely to be relevant to these enzymes in other viruses as well.

Our work suggests concrete ways these observations could be experimentally verified. The role of specific residues in adopting distinct conformations can be evaluated by mutating these residues to remove H-bonding groups. One could then determine whether the kinetic

signatures measured by Reich et al. are perturbed. For example, if one removes H-bonds important for stabilizing the intermediate state, we would expect that the waiting time for product release should increase, as it should become more difficult for the enzyme to convert from initiation to the elongation stage of RNA replication. One could also label hydrogen-bonding partners with FRET donor-acceptor pairs to determine whether measured distance distributions are consistent with the identified conformational states.

The observed conformational distribution has important consequences for understanding how therapeutics may work to target the HCV polymerase. Distinct conformational states may modulate the affinity or efficacy of small molecule inhibitors. In previous studies, we have shown that inhibitors bound to the different allosteric sites on the enzyme can alter its conformational preferences (7, 8, 10). Conversely, it is not unreasonable to expect that the specific conformation adopted by the enzyme could modulate its interactions with small molecule ligands. Thus, drugs that target the initiation stage of RNA replication may be ineffective during elongation and vice versa. Such a model is consistent with studies demonstrating that certain allosteric inhibitors disrupt the transition from elongation to inhibition, while others inhibit initiation (3, 33, 49-51). Thus, it may be necessary to employ combinations of inhibitors if one desires to effectively target all the functionally relevant conformational states adopted by the enzyme. It is possible that a collection of distinct conformational states defined by large-scale structural changes is also present in other viral polymerases given the extensive functional and structural similarities between these enzymes. Thus, the potential impact of this conformational distribution on modulating inhibitor efficacy should be an important consideration for developing therapeutics that can target viral RNA polymerases.

Finally, our findings are consistent with previous structural studies of similar viral RNA polymerases that identified conformational states of the active site (46, 48). The work of these authors is complementary to our own in that those studies focused on local structural rearrangements while our present work examines global conformational changes and structural fluctuations. Thus, it should be possible to combine both the detailed active site description afforded by these previous studies in concert with the global structural changes identified in this work to better understand the molecular processes that govern RNA replication in viral polymerases.

2.5 Acknowledgments

The hardware used in the computational studies is part of the University of Maryland, Baltimore County (UMBC) High Performance Computing Facility (HPCF). This facility is funded by the U.S. National Science foundation through the MRI program (Grants CNS-0821258 and CMS-1228778) and the SCREMS program (Grant DMS-0821311), with significant support from the university. This investigation also used clusters of the Extreme Science and Engineering Discovery Environment (XSEDE). XSEDE is supported by the National Science Foundation (Grant OCI-1053575).

We thank Dr. Brittny Davis for carrying out some of the simulations that were analyzed in this study. We also thank Dr. Patrick O'Neill and Christopher Porter for technical assistance with data analysis.

For additional information see appendix A.

2.6 References

1. Boyce SE, *et al.* (2014) Structural and regulatory elements of HCV NS5B polymerase--beta-loop and C-terminal tail--are required for activity of allosteric thumb site II inhibitors. *PLoS One* 9(1):e84808.
2. Messina JP, *et al.* (2015) Global Distribution and Prevalence of Hepatitis C Virus Genotypes. *Hepatology* 61(1):77-87.
3. Ontoria JM, *et al.* (2009) Identification and Biological Evaluation of a Series of 1H-Benzo[de]isoquinoline-1,3(2H)-diones as Hepatitis C Virus NS5B Polymerase Inhibitors. *J. Med. Chem.* 52(16):5217-5227.
4. Shaw AN, *et al.* (2009) Substituted benzothiadiazine inhibitors of Hepatitis C virus polymerase. *Bioorg. Med. Chem. Lett.* 19(15):4350-4353.
5. Ruebsam F, *et al.* (2008) Pyrrolo[1,2-b]pyridazin-2-ones as potent inhibitors of HCVNS5B polymerase. *Bioorg. Med. Chem. Lett.* 18(12):3616-3621.
6. Di Marco S, *et al.* (2005) Interdomain Communication in Hepatitis C Virus Polymerase Abolished by Small Molecule Inhibitors Bound to a Novel Allosteric Site. *J. Biol. Chem.* 280(33):29765-29770.
7. Davis B & Thorpe IF (2013) Thumb Inhibitor Binding Eliminates Functionally Important Dynamics in the Hepatitis C Virus RNA Polymerase. *Proteins* 81(1):40-52.
8. Davis BC, Brown JA, & Thorpe IF (2015) Allosteric inhibitors have distinct effects, but also common modes of action, in the HCV polymerase. *Biophys. J.* 108(7):1785-1795.
9. Ago H, *et al.* (1999) Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Structure* 7(11):1417-1426.
10. Brown JA & Thorpe IF (2015) Dual Allosteric Inhibitors Jointly Modulate Protein Structure and Dynamics in the Hepatitis C Virus Polymerase. *Biochemistry* 54(26):4131-4141.
11. Adachi T, *et al.* (2002) The essential role of C-terminal residues in regulating the activity of hepatitis C virus RNA-dependent RNA polymerase. *Bba-Proteins Proteom.* 1601(1):38-48.
12. Brooks BR, *et al.* (2009) CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* 30(10):1545-1614.
13. Phillips JC, *et al.* (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781-1802.
14. Lesburg CA, *et al.* (1999) Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nat Struct. Biol.* 6(10):937-943.
15. Ferrenberg AM & Swendsen RH (1988) New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.* 61(23):2635-2638.
16. Ferrenberg AM & Swendsen RH (1989) Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* 63(12):1195-1198.
17. Swendsen RH, Wang JS, & Ferrenberg AM (1992) New Monte-Carlo Methods for Improved Efficiency of Computer-Simulations in Statistical-Mechanics. *Top. Appl. Phys.* 71:75-91.
18. Shirts MR & Chodera JD (2008) Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* 129(12):124105.
19. Kim I & Allen TW (2012) Bennett's acceptance ratio and histogram analysis methods enhanced by umbrella sampling along a reaction coordinate in configurational space. *J. Chem. Phys.* 136(16):164103.

20. Roux B (1995) The Calculation of the Potential of Mean Force Using Computer-Simulations. *Comput. Phys. Commun.* 91(1-3):275-282.
21. Souaille M & Roux B (2001) Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.* 135(1):40-57.
22. Chodera JD, Swope WC, Pitera JW, Seok C, & Dill KA (2007) Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. *J. Chem. Theory Comput.* 3(1):26-41.
23. Swendsen RH (1993) Modern Methods of Analyzing Monte-Carlo Computer-Simulations. *Physica A* 194(1-4):53-62.
24. Kumar S, Bouzida D, Swendsen RH, Kollman PA, & Rosenberg JM (1992) The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. *J. Comput. Chem.* 13(8):1011-1021.
25. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, & Kollman PA (1995) Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. *J. Comput. Chem.* 16(11):1339-1350.
26. Van Dijk AA, Makeyev, E.V, & Bamford, D.H (2004) Initiation of viral RNA-dependent RNA polymerization. *J. Gen. Virol.* 85:1077-1093.
27. Ng KK, Arnold JJ, & Cameron CE (2008) Structure-function relationships among RNA-dependent RNA polymerases. *Curr. Top. Microbiol. Immunol.* 320:137-156.
28. Harrus D, *et al.* (2010) Further Insights into the Roles of GTP and the C Terminus of the Hepatitis C Virus Polymerase in the Initiation of RNA Synthesis. *J. Biol. Chem.* 285(43):32906-32918.
29. Mosley RT, *et al.* (2012) Structure of Hepatitis C Virus Polymerase in Complex with Primer-Template RNA. *J. Virol.* 86(12):6503-6511.
30. Lohmann V (2013) *Hepatitis C Virus: From Molecular Virology to Antiviral Therapy* (Springer-Verlag Berlin Heidelberg, Berlin Heidelberg).
31. Caillet-Saguy C, Lim SP, Shi PY, Lescar J, & Bressanelli S (2014) Polymerases of hepatitis C viruses and flaviviruses: structural and mechanistic insights and drug development. *Antiviral Res.* 105:8-16.
32. Chinnaswamy S, *et al.* (2008) A locking mechanism regulates RNA synthesis and host protein interaction by the hepatitis C virus polymerase. *J. Biol. Chem.* 283(29):20535-20546.
33. Tomei L, *et al.* (2003) Mechanism of Action and Antiviral Activity of Benzimidazole-Based Allosteric Inhibitors of the Hepatitis C Virus RNA-Dependent RNA Polymerase. *J. Virol.* 77(24):13225-13231.
34. Reich S, Golbik RP, Geissler R, Lilie H, & Behrens SE (2010) Mechanisms of activity and inhibition of the hepatitis C virus RNA-dependent RNA polymerase. *J. Biol. Chem.* 285(18):13685-13693.
35. Ferrer-Orta C, Arias A, Escarmis C, & Verdaguer N (2006) A comparison of viral RNA-dependent RNA polymerases. *Curr. Opin. Struct. Biol.* 16(1):27-34.
36. Choi KH (2012) Viral polymerases. *Adv. Exp. Med. Biol.* 726:267-304.
37. McDonald SM (2013) RNA synthetic mechanisms employed by diverse families of RNA viruses. *WIREs RNA* 4(4):351-367.

38. Cameron CE, Moustafa IM, & Arnold JJ (2009) Dynamics: the missing link between structure and function of the viral RNA-dependent RNA polymerase? *Curr. Opin. Struct. Biol.* 19(6):768-774.
39. Shatskaya GS & Dmitrieva TM (2013) Structural organization of viral RNA-dependent RNA polymerases. *Biochemistry (Mosc)* 78(3):231-235.
40. Sesmero E & Thorpe IF (2015) Using the Hepatitis C Virus RNA-Dependent RNA Polymerase as a Model to Understand Viral Polymerase Structure, Function and Dynamics. *Viruses* 7(7):3974-3994.
41. Durk RC, *et al.* (2010) Inhibitors of Foot and Mouth Disease Virus Targeting a Novel Pocket of the RNA-Dependent RNA Polymerase. *PLoS ONE* 5(12).
42. Malet H, *et al.* (2008) The flavivirus polymerase as a target for drug discovery. *Antivir. Res.* 80(1):23-35.
43. Manvar D, *et al.* (2016) Discovery of conjugated thiazolidinone-thiadiazole scaffold as anti-dengue virus polymerase inhibitors. *Biochem. Biophys. Res. Commun.* 469(3):743-747.
44. Yokokawa F, *et al.* (2016) Discovery of Potent Non-Nucleoside Inhibitors of Dengue Viral RNA-Dependent RNA Polymerase from a Fragment Hit Using Structure Based Drug Design. *J. Med. Chem.* 59(8):3935-3952.
45. Brown JA, Espiritu MV, Abraham J, & Thorpe IF (2016) Computational predictions suggest that structural similarity in viral polymerases may lead to comparable allosteric binding sites. *Virus Res.* 222:80-93.
46. Gong P & Peersen OB (2010) Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* 107(52):22505-22510.
47. Zamyatkin DF, *et al.* (2008) Structural insights into mechanisms of catalysis and inhibition in Norwalk virus polymerase. *J. Biol. Chem.* 283(12):7705-7712.
48. Shu B & Gong P (2016) Structural basis of viral RNA-dependent RNA polymerase catalysis and translocation. *Proc. Natl. Acad. Sci. U S A* 113(28):E4005-4014.
49. Nyanguile O, *et al.* (2010) 1a/1b Subtype Profiling of Nonnucleoside Polymerase Inhibitors of Hepatitis C Virus. *J. Virol.* 84(6):2923-2934.
50. Deredge D, Li JW, Johnson KA, & Wintrobe PL (2016) Hydrogen/Deuterium Exchange Kinetics Demonstrate Long Range Allosteric Effects of Thumb Site 2 Inhibitors of Hepatitis C Viral RNA-dependent RNA Polymerase. *J. Biol. Chem.* 291(19):10078-10088.
51. Li JW & Johnson KA (2016) Thumb Site 2 Inhibitors of Hepatitis C Viral RNA-dependent RNA Polymerase Allosterically Block the Transition from Initiation to Elongation. *J. Biol. Chem.* 291(19):10067-10077.

Chapter 3. Effects of structural perturbations on conformational sampling in the HCV polymerase

3.1 Introduction

In this chapter we describe how the structural and dynamic properties of the NS5B conformations, reported in chapter 2, are perturbed by mutations and the binding of inhibitors.

One of our goals has been to further understand the relationship of allosteric inhibition in NS5B with its conformations. In NS5B, as we pointed out in chapter 1, four allosteric (NNI) sites have been identified: two in the thumb (NNI-1 and NNI-2) and two in the palm (NNI-3 and NNI-4) (see Figure 1.5). Brown *et al.* (1) showed that NNI-3 and NNI-4 are likely to be distinct regions within a single large pocket rather than two individual pockets. For this reason, we use the nomenclature NNI-3/4 to denote both of these partially overlapping sites. The inhibition mechanisms for NNIs reported to-date are:

- 1) Thumb NNIs (NNI-1 & NNI-2) inhibit an early phase of replication that occurs after initiation but before elongation starts (2-5). Inhibitors targeting the NNI-1 pocket seem to prevent enzyme function by reducing overall stability of the enzyme and preventing it from stably adopting functional conformations. In contrast, NNI-2 inhibitors seem to reduce conformational sampling, preventing the transitions between the open and closed conformational states that are required for NS5B to function (6).
- 2) Most of the NNIs that bind to the palm domain (NNI-3/4) have been found to stabilize the β -flap via critical interactions with Tyr448 (7), fixing it in the closed,

initiation-appropriate conformation and preventing these residues from moving out to allow the RNA double helix to egress (8). Thus, it has been suggested that palm NNIs inhibit initiation. NNI-3/4 inhibitors were also observed to restrict conformational sampling, though the dominant mode of action of these molecules was predicted to result from blocking access of the RNA template (6).

In the present study we complement these previous findings by showing how some of these observations, especially in the cases of NN-1 and NNI-2, have their root in the modification of the enzyme FEL. H-bond analysis of the inhibitor-bound systems suggest two dominant mechanisms by which inhibitors mediate their effects: i) disrupting the communication between domains, ii) preferentially stabilizing one of the available conformations. These strategies could be adopted moving forward in drug design of new inhibitors for NS5B and related polymerases for which fewer treatments are available.

Introducing enzyme mutations allowed us to understand to what degree the inhibitor effects depend on the enzyme sequence and thus to what extent we expect our findings to be transferable to other viruses. We focused on the impact of these mutations on the conformations sampled by the enzyme. Our results showed that the mutations have a much smaller impact on the enzyme FEL than the inhibitors. In addition, they also showed that the genotype 2a systems sampled analogous closed and intermediate conformations as the genotype 1b. Thus, we expect that the insights we garnered from this study will be applicable to other related viral polymerases.

3.2 Methods

3.2.1 Structure Preparation and Simulation Conditions

The structures employed in this study are specified in Table 3.1. In the naming of the inhibitor-bound systems the first set of letters before the dash is the PDB ID of the inhibitor(s) followed by the PDB ID of the protein structure. In the naming of the mutated systems, the first set of letters before the underscore is the PDB ID of the protein structure. The next set of letters indicates a mutation made by our group. Structures where no mutation is specified were taken directly from the PDB database.

For the inhibitor-bound systems, simulations of 2WHO with the VGI inhibitor were previously performed by Davis and Thorpe (9); simulations of 2WHO, 3HHK, 3CO9 and 2BRL with the PFI, 77Z, 3MS and POO inhibitors respectively by Davis, Brown and Thorpe (6) and simulations of 1QUV with 3MS and VGI inhibitors by Brown and Thorpe (1). This data was subjected to further analysis in this study. Structure 2BRL does not contain residues 22-35 or 148-152. Consequently, coordinates for these residues were taken from 2WHO, placed into the 2BRL structure and energy minimized before simulations were initiated as described by Davis, Brown and Thorpe (6). Inhibitor parameters were generated using the methods of Vanommeslaeghe et al. (10-12) as described by Davis and Thorpe (6) and ligand binding was confirmed by visual comparison to the crystal structure. All structures were solvated with TIP3P (13) water and placed in a truncated octahedron that was larger than the protein by at least 10 Å in each dimension. This generated unit cells with an edge length of at least 93 Å for each structure. Nineteen chloride ions were added to each system to neutralize the system. Any water molecules overlapping with protein, ligand, or chloride ions were removed. Each system contained approximately 60,000 atoms.

For the mutated polymerase systems there are three different categories: 1) mutations that we made, 2) structures that we obtained already mutated from the PDB data base and 3) structures from HCV genotype 2a. Structures in the third category contain a collection of amino acid changes relative to genotype 1b. Genotype 2a has 75.53% identity and 85.18% similarity respect to genotype 1b.

Category 1.

For the first category, bound inhibitors were removed from 2WHO, (9) 3CO9 (14) and 2BRL (6) so that all structures were simulated as free enzymes. One mutation was generated for each of the three enzymes. Criteria for a residue to be mutated were as follows: i) it was identified as a critical residue using NetworkView,(15) ii) it was not critical in both free and ligand-bound systems, iii) it did not directly interact with the ligand, and iv) it was in a conserved region. Satisfaction of criterion (iv) was determined via sequence alignment of NS5B sequences from several genotypes available at NCBI (<http://www.ncbi.nlm.nih.gov/protein>). The genotypes and sequence accession numbers employed are as follows: 1a (AAB67036.1), 1b (BAA83719.1), 2a (AAF01178.1), 2b (AAF59945.1), 2c (BAA88057.1) and 3a (AAC03058). The identification of criterion (i) was done as follows: trajectories containing the centers of mass of each residue for our mutated structures were generated. These were analyzed to identify allosteric communication networks as described below. In this analysis each residue center of mass is considered a node. Two nodes are connected by an edge if they are within 8.5 Å of each other for at least 75% of the trajectory. Residues adjacent in sequence are excluded and cannot be connected by an edge. Edges are weighted according to the correlation between nodes as computed via Eq. 3. The Girvan–Newman algorithm (16) was used to calculate communities in

the network: these are collections of nodes that have more connections to other nodes in the community than to nodes outside the community. Allosteric communication is thought to occur preferentially over nodes between two communities with the greatest betweenness, where betweenness equals the number of shortest paths that pass through a node. These nodes/residues may be part of important routes of communication across the enzyme. This identification was done by Dr. Brittny Davis and the mutated systems in this category were originally generated for a different study carried out by Dr. Brittny Davis and Dr. Ian Thorpe. However, these simulations were employed to perform new data analysis described in our present work.

Mutations D55A, L469P, and Y452H were originally selected as mutations that would counteract the effect of the NNI-1, NNI-2 and NNI-3 inhibitors, respectively. Here we use the simulations of Davis and Thorpe to study the effect of mutations on the FEL of the free enzyme. In addition, the L30S mutation was selected to disrupt functional motions and conformational sampling in the free enzyme. Finally, A73L was generated to serve as a control. Note that mutations L30S and Y452H were also chosen because of available experimental data regarding the impact of these mutations on the properties of the enzyme (17, 18). L30S was shown to reduce enzyme activity while Y452H was shown to confer resistance to an inhibitor bound to the palm site. See figure 3.1 for location of the mutations on the enzyme.

Mutations were made using the MMTSB toolset (19). The protein coordinates employed were taken from the last snapshot of simulations of free structures of 2BRL (6), 2WHO (9) and 3CO9 (14). To make the mutation MMTSB preserves the positions of the α -Carbon and side-chain center of mass for the mutated residue. It then rebuilds an all-atom structure with the new sequence while choosing a side chain rotamer that minimizes unfavorable interactions with residues that are not being altered.

Category 2.

For the second category, the structures of 2HAI (20) and 3FRZ (21) were obtained from the PDB database with previously existing mutations. Both structures were crystallized with inhibitors bound so the inhibitors were removed for the simulations. The 2HAI simulations were previously performed by Davis and Thorpe (14). 2HAI has mutations L47Q, F101Y and K114R. 3FRZ contains all the mutations present in 2HAI as well as the additional mutation V59D.

Category 3.

For the third category we employed PDB coordinates 1YUY and 1YV2 (22) that are from genotype 2a. These structures are of particular interest because they were the first ones to be reported as “open” and “closed” (see chapter 1). In addition, these structures allow us to determine whether the observations made for genotype 1b are also transferable to genotype 2a.

The C-terminal residues of the enzyme are not necessary for its function in vitro (23). Thus, the C-terminal residues of all systems were truncated so that each contained a total of 531 residues.

Simulations were performed using the CHARMM 27 (24) protein force field and NAMD version 2.9 (25). Minimization of each solvated enzyme was carried out using NAMD for 1000 steps using the conjugate gradient method and applying periodic boundary conditions. All bonds to hydrogen atoms were constrained using the SHAKE algorithm, and electrostatic interactions were calculated by the Particle Mesh Ewald method. A distance of 9 Å was used for the coulomb cutoff and 11.4 Å for the non-bonded pair list. Minimization was followed by 2,500,000 steps of simulation in the NVT ensemble at 300 K using a 2 fs integration time step. Temperature was maintained via velocity reassignment every 100 steps, during which the

positions of alpha carbons was restrained using force constants of 10 kcal(mol⁻¹ Å⁻²). Restraints were then removed and simulations performed in the NPT ensemble at 300 K for 2,500,000 steps using Berendsen pressure coupling at 1.01 bar with a relaxation time of 100 fs. At this stage the temperature was still maintained via velocity reassignment every 100 steps. Finally, production

Table 3.1 Simulated systems.			
System's modification	PDB ID [a, b]	Ligand location in PDB structure [c]	Simulation time (ns)
inhibitor-bound	VGI-1QUV	NNI-2	1000
	3MS-1QUV	NNI-3	1000
	3MS-VGI-1QUV	NNI-2 & 3	1000
	3MS-3CO9	NNI-3	710
	VGI-2WHO	NNI-2	400
	PFI-2WHO	NNI-2	500
	77Z-3HHK	NNI-3	500
	POO-2BRL	NNI-1	520
mutated	3FRZ	-	700
	2HAI	-	700
	1YUY	-	500
	1YV2	-	600
	2BRL_D55A	-	300
	2BRL_A73L	-	300
	2WHO_L469P	-	300
	2WHO_A73L	-	300
	3CO9_L30S	-	400
	3CO9_Y452H	-	400
	3CO9_A73L	-	300
[a] In the naming of the inhibitor-bound systems, the first set of letters before the dash is the PDB ID of the ligand(s) followed by the PDB ID of the protein structure.			
[b] In the naming of the mutated systems, the first set of letters before the underscore is the PDB ID of protein structure. The next set of letters indicates a mutation made by our group. Structures where no mutation is specified were taken directly from the PDB database.			
[c] NNI: Non-nucleoside inhibitor site			

NVT simulations were performed using the same conditions as the original NVT simulations, except that temperature was maintained using a Langevin thermostat with a damping coefficient of 1 ps^{-1} applied to heavy atoms and restraints were not employed. The pressure was checked every 10 ns to confirm that it remained within reasonable range of 1.01 bar. The final 200 ns of each trajectory was analyzed.

For explanation of the metrics for conformational sampling (i.e. inter-domain angle, PCA) and FEL and WHAM see Methods section in chapter 2.

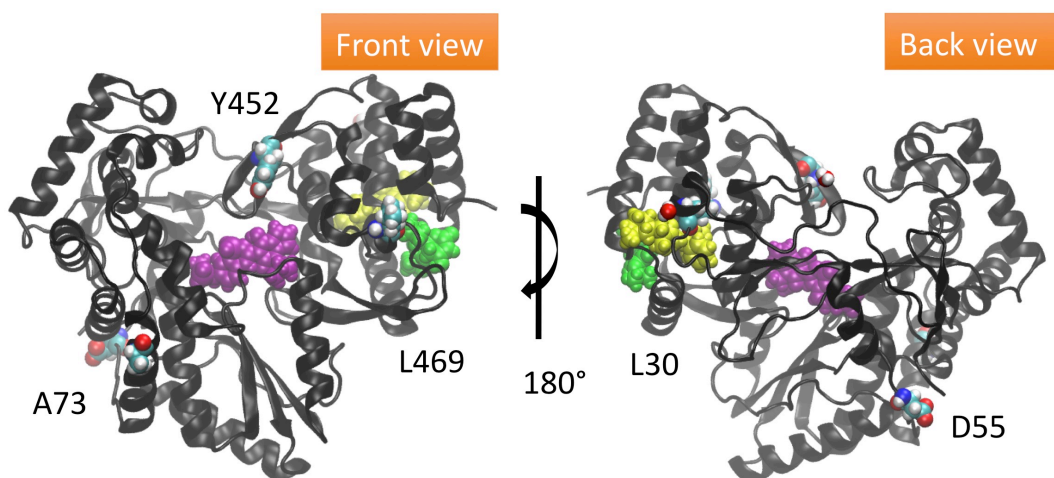


Figure 3.1. Location of mutations in the NS5B polymerase. In the figure we show inhibitors in the allosteric sites so that the locations of mutations with respect to the inhibitors are easily recognizable. NNI-1 is in yellow, NNI-2 in green and NNI-3 in purple.

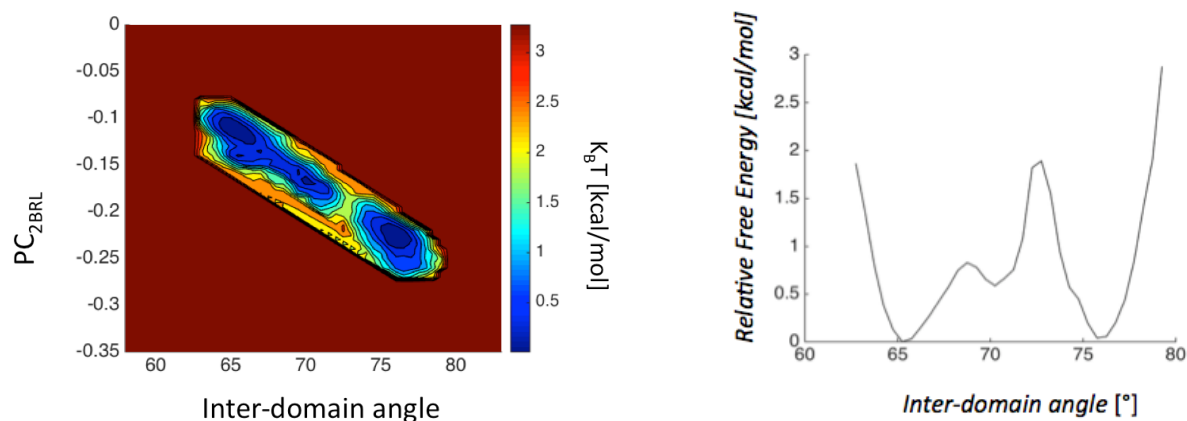
3.3 Results and Discussion

3.3.1 The FEL of the enzyme maintains its essential features despite the presence of perturbations

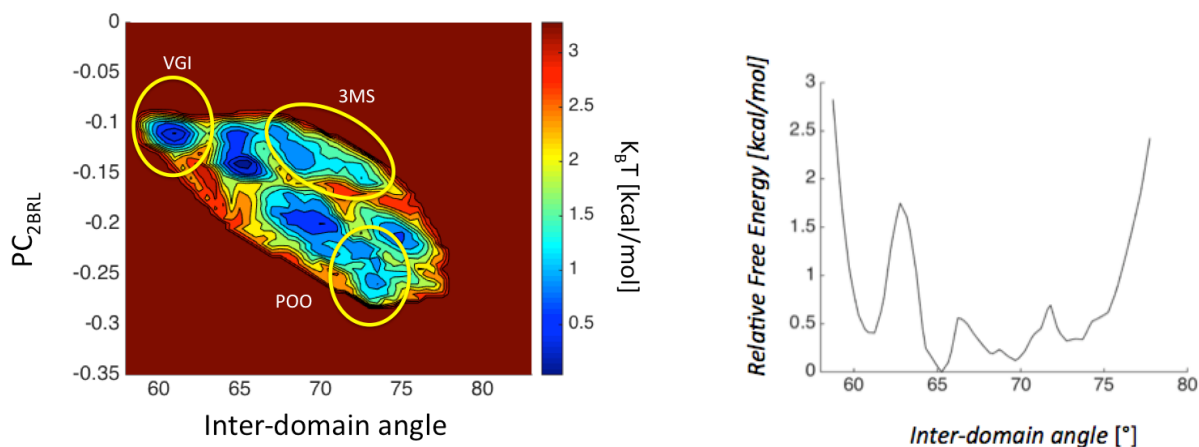
Comparing the 1-D FELs, we observe that in the inhibitor-bound systems the barriers are reduced between closed/intermediate and intermediate/open relative to the free systems.

Following the classification we did in chapter 2: closed (inter-domain angle $< 69^\circ$), intermediate ($69^\circ < \text{inter-domain angle} < 74^\circ$) and open (inter-domain angle $> 74^\circ$).

a) Free energy landscapes of the free enzyme (26)



b) Free energy landscapes from the inhibitor-bound systems.



c) Free energy landscapes of the mutated systems.

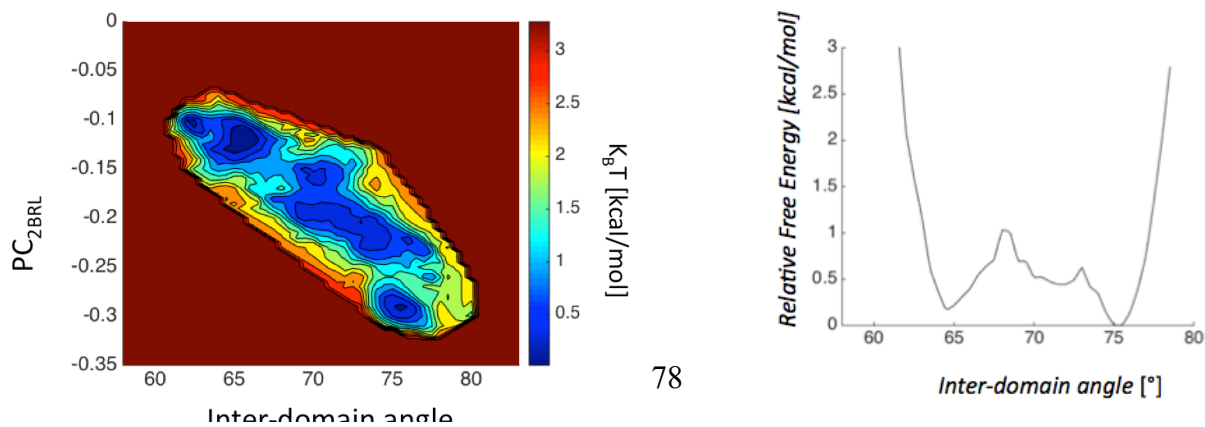


Figure 3.2. 1-D (right) and 2-D (left) Free energy landscapes of free systems and ligand bound systems.

However, the stability of the open state is reduced as well so that the closed state is now most stable rather than both closed and open states displaying the same stability. This suggests that one strategy for inhibition may be preferentially stabilizing one of the enzyme conformations so that the enzyme gets “locked” in that conformation and cannot cycle between conformations as required for the replication cycle. In the inhibitor-bound systems we also clearly see that the hyper-closed conformation (~62 degrees) reported by Davis et al. (9) is a trap that is off the pathway of functional conformations transitions as described in their paper. Looking at the mutated systems, we also see a reduction in the barriers between conformations. In addition, we observe that the minimum corresponding to the closed conformation now is higher in energy than in the free enzyme and the minimum corresponding to the open conformation now is lower in energy with respect to the free enzyme. Thus, the closed conformation becomes slightly less stable than the open. We also observe that the closed/intermediate barrier is now larger and the intermediate/open barrier is now smaller. If the barrier from intermediate to open is lower it may be easier for the enzyme to transition to the elongation stage. This would make the enzyme more processive.

The 1-D and 2-D free energy landscapes (FELs) obtained using the combined data for the inhibitor-bound systems and the mutated systems are shown in figure 3.2. Individual FELs of each system are in appendix B. The FEL for the free systems, from chapter 2, is also shown here for completeness and for easier comparison.

Comparing the 2-D FELs, we readily notice that the inhibitors cause a much larger effect in perturbing the FEL than the mutations but that the range of conformations sampled by the NS5B as a free enzyme is maintained. In all three scenarios (free, inhibitor-bound and mutated

systems) we see minima analogous to the minima of the FEL of the free systems that sample the open (inter-domain angle > 74), intermediate ($69 < \text{inter-domain angle} < 74$) and closed (inter-domain angle < 69) conformations. Thus, this suggests that the model that we reported in chapter 2 for the free enzyme is also applicable to other HCV genotypes. It is not unreasonable to expect that this model will be relevant for other viral polymerases as well.

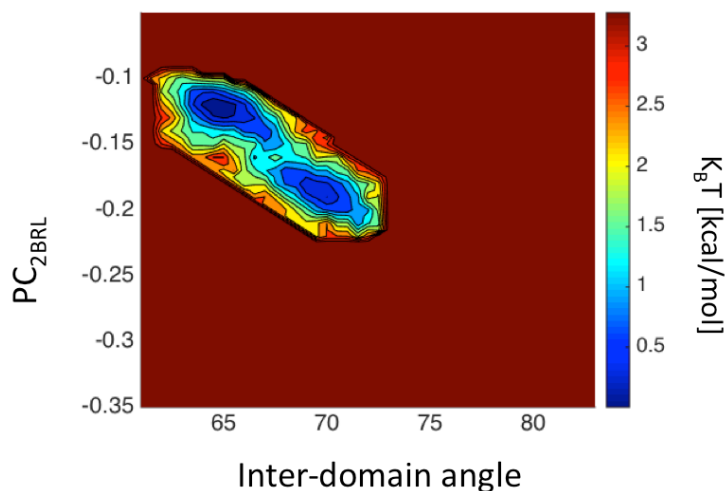


Figure 3.3. FEL of genotype 2a systems.

Looking at the FEL (see figure 3.3) that combines the results that we obtained for the genotype 2a systems (i.e. 1YUY and 1YV2), we observe closed and intermediate states similar to those sampled by genotype 1b. The open state is not sampled by those structures but it is worth noting that to generate this FEL we only used

the simulations started from 2 structures, while to generate the FEL for genotype 1b we used 5 structures. Only one of simulation of genotype 1b sampled the open state, so this may suggest that more simulation data would be needed to see the full conformational space with genotype 2a.

From the comparison of the 2-D FELs we see that, like the free enzyme, the inhibitor-bound and mutated systems sample a variety of conformations. Thus, the conformational selection mechanism that we proposed for the free enzyme appears to still apply for perturbed systems. In addition, we observe that the closed minimum remains unchanged but that the positions and shapes of the intermediate and open states are drastically modified. The fact that

there are additional minima in the intermediate state compared to the free enzyme may slow the transition from closed to open by trapping the enzyme as it converts between them. The closed state is the least impacted by binding of inhibitors or mutations, what may be due to the fact that in this state the domains are closer together and can be more easily stabilized via hydrogen bonding.

We picked snapshots that corresponded to the FEL minima based on the metrics shown in Figure 3.2. and generated PDB files representative of minimum. Aligning the palm domain we performed a pair-wise comparison of the root mean square deviation (RMSD) for each structure (see appendix B). This analysis confirmed that inhibitors perturb the free structure more than mutations in all three (open, intermediate and closed) minima. Thus, inhibitors likely have a larger effect than differences due to sequence that might be observed for other genotypes or even other viral polymerases. This observation further suggests that the general feature of the FEL may be transferable to other HCV genotypes and related viral polymerases. Comparing the inhibitor-bound systems and the free systems we observe that the inhibitors cause the largest perturbation to the intermediate state. Thus, inhibitor binding has a dominant impact on the transition between open and closed states.

As we see in the 2-D FEL, the minimum for the closed state of inhibitor-bound systems also has a different shape and position compared to the free enzyme. Disruption of the closed state may make the initiation stage more difficult since initiation is thought to occur in this state. Inhibitors likely do not impact elongation significantly, given that the RMSD differences between the open state in the free and the inhibitor-bound minima is small (2.4 Å). This agrees with what has been reported for NNI-2 inhibitors (5). Looking at the RMSD values for the mutated systems, we see that the open state is most perturbed, consistent with the FEL.

Perturbing the open state should have an impact on transition to the elongation stage of replication and the elongation itself.

We superimposed the structures representative of each minima by aligning those corresponding to a given conformation (see appendix B) and noticed that the closed state is the least perturbed among different systems, in agreement with our observations for the 2-D FELs. The main feature that differentiates the closed states seems to be the position of the beta-flap, what can affect the positioning of the template, since the beta-flap has been reported to be involved in stabilizing it (27).

3.3.2 Inhibitors modify the Free energy landscape

In previous publications (1, 6) it was hinted that perturbing the FEL would be one mechanism used by inhibitors to prevent the enzyme's function but it was never fully developed or tested. The present results confirm this hypothesis and show that there are two different ways used by the inhibitors to alter the FEL: i) sampling a novel conformation (such as in the case of VGI, 3MS and POO) ii) changing the probability with which the conformations are sampled (as observed for the remaining inhibitors studied).

3.3.3 Inhibitors disrupt communication between domains and preferentially stabilize specific conformations

We analyzed the Hydrogen bonds (H-bonds) of the inhibitor-bound systems (see appendix B) for the open, intermediate and closed conformations and identified the H-bonds that

were different between them. We observed that H-bond networks involving the delta 1 loop, delta 2 loop and fingers were eliminated in the open and closed conformations. H-bonds between the fingers, thumb and palm domains were also removed in the inhibitor bound systems. All these observations lead us to conclude that one of the strategies for inhibition is breaking the H-bonds between domains, thus disrupting inter-domain communication.

The comparison of the number of H-bond networks that differ between conformations when looking at the free enzyme and the inhibitor-bound enzyme revealed a much lower number for inhibitor-bound systems (see appendix B), this shows how the inhibitor-bound systems are less coupled than the free enzyme as a consequence of the disruption of the H-bonds between domains as it was noted in the previous paragraph.

We also observed that in the inhibitor-bound systems there are fewer H-bonds between the thumb and fingers in the open conformation and fewer H-bonds between the thumb and palm in the closed conformation. Our conclusion from these observations is that another mechanism for inhibition is preferentially stabilizing specific conformations, thus preventing transitions between conformations as is also suggested by the 1-D FEL.

3.3.4 Mutations destabilize enzyme conformations.

For the H-bond analysis of the mutated systems we used one system that primarily sampled a given minimum to represent that minimum. The systems we used were: 2HAI for the open conformation (sampled this conformation in 1914 out of 2000 frames), 3FRZ for the intermediate conformation (sampled this conformation in 1675 frames out of 2000) and 1YUY for the closed conformation (sampled this conformation in 1988 frames out of 2000). From this

analysis (see appendix B) we observed that there are more H-bonds forming and breaking during the transitions in both the fingers and the thumb domain compared to the free enzyme. In addition, we observed that some H-bonds that were only broken in the intermediate state of the free enzyme were not present in the mutated systems. These suggest that the mutations destabilize the enzyme conformations, making it more difficult for the enzyme to establish the interactions that buttress each conformation.

Both 3FRZ and 2HAI sample the open conformation and have the mutation L47Q. In the H-bond analysis of the free enzyme it is apparent that L47 is involved in the H-bond network that we reported (see chapter 2) between the delta 1, delta 2 and the rest of the fingers domain, stabilizing the closed conformation. The fact that L47 is mutated to a glutamine could result in this network not to be formed and favor the sampling of the open conformation versus the closed.

Highlighted in cyan are hydrogen bonds between residues at the same position but which have different sequence identity in genotype 1b and genotype 2a. For example, in the delta 1 loop in the closed conformation the hydrogen bonded pair is CYS 39 – GLU 18 and in the open conformation the pair is ALA 39 – GLU 18. The fact that the H-bond is still formed even if the amino acid sequence is different agrees with our previous observations that support the possible transferability of our results to other genotypes and related viruses.

3.4 Conclusions

The study in this chapter showed that the free energy landscape we described for NS5B as a free enzyme (in chapter 2) may be widely applicable to other HCV genotypes. Both

inhibitor-bound and mutated systems in this study have shown how the NS5B polymerase is a very tightly coupled allosterically regulated system. However, the structural aspect has been shown to have a more decisive impact in the function of the enzyme than the sequence. For that reason, we expect for the insights garnered from this study to be transferable to other genotypes of HCV and also other viral polymerases. This knowledge may be of crucial importance in drug development in particular for other viruses for which there is not much information and treatments are scarce, such as West Nile virus or foot-and-mouth disease virus.

The results from this study also indicate that previously reported inhibitors have a large impact on modifying the Free energy landscape. We showed that these inhibitors can encourage novel conformations to be adopted or change the relative probability of sampled conformations. Analyzing the H-bonds of the inhibitor-bound systems, we found that the inhibitors accomplish these effects by disrupting communication between domains and preferentially stabilizing specific conformations. These strategies are of remarkable importance because they may be applied in a new stage of drug design. The H-bonds that we have reported could be used as targets for inhibition of the enzyme function.

3.5 Acknowledgements

The hardware used in the computational studies is part of the University of Maryland, Baltimore County (UMBC) High Performance Computing Facility (HPCF). This facility is funded by the U.S. National Science foundation through the MRI program (Grants CNS-0821258 and CMS-1228778) and the SCREMS program (Grant DMS-0821311), with significant support from the university. This investigation also used clusters of the Extreme Science and Engineering

Discovery Environment (XSEDE). XSEDE is supported by the National Science Foundation (Grant OCI-1053575).

We thank Dr. Brittney Davis, Dr. Jodian Brown and Marie Espiritu for carrying out some of the simulations that were analyzed in this paper. We also thank Dr. Alexander Sodt, Dr. Patrick O'Neill and Christopher Porter for technical assistance with data analysis.

3.6 References

1. Brown JA & Thorpe IF (2015) Dual Allosteric Inhibitors Jointly Modulate Protein Structure and Dynamics in the Hepatitis C Virus Polymerase. *Biochemistry* 54(26):4131-4141.
2. Ontoria JM, *et al.* (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(16):5217-5227.
3. Nyanguile O, *et al.* (2008) 1,5-benzodiazepines, a novel class of hepatitis C virus polymerase nonnucleoside inhibitors. *Antimicrob Agents Chemother* 52(12):4420-4431.
4. Nyanguile O, Devogelaere B, & Fanning GC (2010) 1a/1bSubtype profiling of nonnucleoside polymerase inhibitors of hepatitis C virus. *J. Virol.* 84:2923-2934.
5. Tomei L, *et al.* (2003) Mechanism of action and antiviral activity of benzimidazole-based allosteric inhibitors of the hepatitis C virus RNA-dependent RNA polymerase. *J Virol* 77(24):13225-13231.
6. Davis BC, Brown JA, & Thorpe IF (2015) Allosteric inhibitors have distinct effects, but also common modes of action, in the HCV polymerase. *Biophys J* 108(7):1785-1795.
7. Pfefferkorn JA, *et al.* (2005) Inhibitors of HCV NS5B polymerase. Part 1: Evaluation of the southern region of (2Z)-2-(benzoylamino)-3-(5-phenyl-2-furyl)acrylic acid. *Bioorg Med Chem Lett* 15(10):2481-2486.
8. Mosley RT, *et al.* (2012) Structure of hepatitis C virus polymerase in complex with primer-template RNA. *J Virol* 86(12):6503-6511.
9. Davis BC & Thorpe IF (2013) Thumb inhibitor binding eliminates functionally important dynamics in the hepatitis C virus RNA polymerase. *Proteins* 81(1):40-52.
10. Vanommeslaeghe K, *et al.* (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* 31:671-690.
11. Vanommeslaeghe K & MacKerell AD, Jr. (2012) Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. *2012* 52:3144-3154.
12. Vanommeslaeghe KR, E. P.; MacKerell, A. D., Jr (2012) Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* 52:3155-3168.
13. Jorgensen WL, Chandrasekhar J, Madura JD, mpey RW, & Klein ML (1983) Comparison of Simple Potential Function for Simulating Liquid Water. *J. Phys. Chem* 79:926-935.

14. Davis BC & Thorpe IF (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(26):4541-4552.
15. Eargle J & Luthey-Schulten Z (2012) NetworkView: 3D display and analysis of protein.RNA interaction networks. *Bioinformatics* 18:3616-3621.
16. Girvan M & Newman ME (2002) Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U S A* 99:7821-7826.
17. Labonte P, *et al.* (2002) Modulation of Hepatitis C Virus RNA-dependent RNA Polymerase Activity by Structure-based Site-directed Mutagenesis. *J Biol Chem* 277(41):38838-38846.
18. Ando I, *et al.* (2012) Preclinical characterization of JTK-853, a novel nonnucleoside inhibitor of the hepatitis C virus RNA-dependent RNA polymerase. *Antimicrob Agents Chemother* 56(8):4250-4256.
19. Feig M, Karanicolas J, & Brooks CL (2004) MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *Mol. Graph. Model.* 22:377-395.
20. Li H, Tatlock J, Linton A, Gonzalez J, & Borchardt A (2006) Identification and structure-based optimization of novel dihydropyrones as potent HCV RNA polymerase inhibitors. *Bioorg Med Chem Lett.* 16(18):4834-4838.
21. Li H, Tatlock J, Linton A, Gonzalez J, & Jewell T (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. *J Med Chem* 52(5):1255-1258.
22. Biswal BK, Cherney MM, Wang M, & Chan L (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-nucleoside inhibitors. *J.Biol.Chem.* 280:18202-18210.
23. Adachi T, *et al.* (2002) The essential role of C-terminal residues in regulating the activity of hepatitis C virus RNA-dependent RNA polymerase. *Biochim Biophys Acta* 1601(1):38-48.
24. Brooks BR, Brooks CL, Mackerell ADJ, & Nilsson L (2009) CHARMM: the biomolecular simulation program. *J Comput Chem.* 30(10):1545-1614.
25. Phillips JC, Braun R, Wang W, Gumbart J, & Tajkhorshid E (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781-1802.
26. Sesmero E & Thorpe IF (2017) Using molecular simulations to delineate functional conformational transitions in the HCV polymerase. *J. Comp. Chem* (in press).
27. Ferrer-Orta C, Arias A, Escarmis C, & Verdaguer N (2006) A comparison of viral RNA-dependent RNA polymerases. *Curr Opin Struct Biol* 16(1):27-34.

Chapter 4. Summary, Significance and Future Directions

4.1 Dissertation Summary: Major Findings of this Dissertation

Our overall findings contribute to understanding the role that enzyme conformational changes play in replication of the viral genome. Specifically, we have examined, for the first time, the Free Energy Landscape (FEL) of the NS5B RNA polymerase from Hepatitis C Virus and the way in which it is altered by structural perturbations. The enzyme FEL revealed the existence of an intermediate conformation that was not hitherto reported. We characterized each conformation and the transitions between them, identifying the stabilizing H-bond interactions that are specific to each. From these findings we suggested new targets and strategies for developing therapeutics. The strategies we suggested in chapter 2 as possible avenues for inhibition were i) disrupting the communication between domains, and ii) preferentially stabilizing one of the conformations. Those strategies were confirmed in chapter 3 as mechanisms by which the previously reported inhibitors act. From this outcome, we determined principles underlying the mechanism of action of the inhibitors included in this study. Furthermore, we suggested that given that these strategies (shown in chapter 3) are adopted by the inhibitors in this study, they likely decrease enzyme function and thus may be suitable for drug development with the key H-bond interactions that we described as targets.

4.1.1 Characterization of the NS5B conformations and the transitions between them

(Chapter 2)

Our study demonstrated that the HCV polymerase without ligands present explores diverse conformations that appear relevant to the RNA replication cycle. This is observed even in the absence of other components needed for the RNA replication cycle such as nucleotides or the RNA template. This observation may not have been necessarily expected. It could have been the case that the enzyme stays in the closed conformation (needed for initiation) until the components needed for the replication complex are bound and converts to the open conformation at the elongation stage. These observations would be more consistent with an induced fit model of enzyme function. However, the fact that the free enzyme appears to sample many of the conformations needed for replication suggests that the enzyme adopts a conformational selection mechanism in which functionally relevant conformations are accessible to the free enzyme. Specific conformations may be preferentially stabilized by particular binding partners, such as components for the replication complex or small molecule inhibitors, as seen in chapter 3, but these molecules are not necessarily required for the conformation to be observed.

Each conformation is characterized by distinct patterns of flexibility, large amplitude structural changes and correlated motions. Noteworthy are the unique H-bonds of each conformation and the H-bond networks formed among them that play a key role in stabilizing each state and in facilitating transitions between them. Those H-bonds may be good targets for new drug development strategies. Given that the RNA replication cycle is thought to be highly similar among viral polymerases (particularly those that employ (+)-sense templates) our observations are likely to be relevant to these enzymes in other viruses as well. Thus, the

strategies suggested in this work of breaking inter-domain H-bonds and preferentially stabilizing specific conformations as a means to inhibit the function of NS5B may also be applicable to other viruses for which fewer treatments are available.

4.1.2 The effect of mutations and inhibitors show that insights from NS5B may be transferable to other viral polymerases (Chapter 3)

This study showed that the FEL of the enzyme maintains its essential elements despite the presence of perturbations such as mutations or allosteric inhibitors. This observation suggests that the insights obtained from NS5B will be transferable to other polymerases since the structural aspects of the enzyme seem to have a more dominant impact on its function than its sequence.

Another highly valuable conclusion from our results was how modification of the Free Energy landscape is the underlying principle behind mechanisms of inhibition observed previously. We observed how this modification could occur if an inhibitor encourages the enzyme to sample a novel non-functionally relevant conformation or if it changes the probability by which the conformations are sampled.

The H-bond analysis of the inhibitor-bound systems revealed that they mainly disrupt communication between domains and preferentially stabilize one or more conformations. This finding is of particular importance because it brings to light the mechanism of action of the previously reported inhibitors, and it also suggests strategies to be used for designing new drugs. Thus, the H-bonds identified in chapter 2 to be characteristic of each conformation may be targets for new therapies. Inhibitors could be designed that would break the inter-domain H-

bonds, preventing communication between domains and, as a result, the conformational changes needed for the replication cycle to take place.

4.2 Significance

Characterization of NS5B conformations and the deeper understanding of the replication process that we obtained from the results of the present study is a major milestone for viral polymerases. The details of the conformational distribution that the Free Energy landscape of the free enzyme (see chapter 2) revealed have important consequences for understanding how therapeutics may work to target the HCV polymerase. Specific conformational states may modulate the affinity or efficacy of small molecule inhibitors. We have shown in previous studies that inhibitors bound to the different allosteric sites on the enzyme can alter its conformational preferences (1-3). Conversely, it is not unreasonable to expect that the specific conformation adopted by the enzyme could modulate its interactions with small molecule ligands. As a consequence, drugs that target the initiation stage of RNA replication (when the enzyme is in the closed conformation) may be ineffective during elongation (when the enzyme is in the open conformation) and vice versa. Such a model is consistent with studies demonstrating that certain allosteric inhibitors disrupt the transition from elongation to inhibition, while others inhibit initiation (4-8). Thus, it may be advantageous to employ a combination therapy using a variety of inhibitors directed at different conformational states if one desires to effectively target the enzyme at all stages of RNA replication. It is likely that a collection of distinct conformational states defined by large-scale structural changes, as we see in NS5B, is also present in other viral polymerases, given the extensive functional and structural similarities

between these enzymes. As a result, the potential impact of this conformational distribution on modulating inhibitor efficacy should be an important factor to take into consideration for developing therapeutics that can target viral RNA polymerases.

4.3 Future Directions

The present study opens door for several research opportunities in the future. Three main studies would be recommended after this work:

- 1) MD simulations with mutations that are expected to reproduce the action of the inhibitors. To design these simulations, the residues that form the H-bonds that have been shown to be affected by the inhibitors would be mutated to residues that are unable to form those H-bonds. If the mutations reproduce the effect of the inhibitors, we would expect to see a shift in the relative probability with which the conformations are sampled by the enzyme or the sampling of a novel conformation such as we have seen to result from inhibitor action. Thus, those simulations will confirm our understanding of the mechanism of action of the given inhibitors and will contribute to further our knowledge for new drug discovery efforts.
- 2) Hydrogen-deuterium exchange experiments coupled to mass spectrometry may be employed to evaluate the impact of inhibitor binding on the H-bonds of the HCV polymerase. Results from such a study may provide additional evidence for the changes to H-bonds and dynamics for each conformation predicted in this dissertation and further validate the use of computational approaches in understanding protein conformations and allosteric inhibition in viral polymerases. Validating the computational studies

experimentally would be highly beneficial since the *in silico* approaches require a lower investment of funds and could be useful in guiding the experimental work. Moreover, such studies reveal the molecular details underlying the experimental observations that are not always readily accessible from the experiments.

- 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. Our work revealed that the free enzyme sampled the closed, intermediate and open conformations and that inhibitor binding perturbs the free energy landscape, changing the frequency with which the minima are sampled. The presence of RNA template may also favor specific conformations. Insights regarding changes to the free energy landscape 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 replication. An important piece of information resulting from such a study is whether the enzyme automatically opens up in preparation for elongation, or if the enzyme stays in the closed conformation until duplex RNA is formed to force the enzyme open. Each of these scenarios would facilitate the understanding of how effective specific inhibitors can be during different stages of the RNA replication cycle. If a specific inhibitor can only bind to the closed conformation, it would be most effective in the initiation stage of replication. However, if the enzyme readily transitions to the open conformation once bound to template RNA, then an inhibitor that can bind to the open conformation found during the elongation stage of replication might be more beneficial. Thus, this knowledge would be very helpful in designing effective combination therapies.

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). 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).

4.4 References

1. Davis BC & Thorpe IF (2013) Thumb inhibitor binding eliminates functionally important dynamics in the hepatitis C virus RNA polymerase. *Proteins* 81(1):40-52.
2. Davis BC, Brown JA, & Thorpe IF (2015) Allosteric inhibitors have distinct effects, but also common modes of action, in the HCV polymerase. *Biophys J* 108(7):1785-1795.
3. Brown JA & Thorpe IF (2015) Dual Allosteric Inhibitors Jointly Modulate Protein Structure and Dynamics in the Hepatitis C Virus Polymerase. *Biochemistry* 54(26):4131-4141.
4. Ontoria JM, *et al.* (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(16):5217-5227.
5. Tomei L, *et al.* (2003) Mechanism of action and antiviral activity of benzimidazole-based allosteric inhibitors of the hepatitis C virus RNA-dependent RNA polymerase. *J Virol* 77(24):13225-13231.
6. Nyanguile O, Devogelaere B, & Fanning GC (2010) 1a/1bSubtype profiling of nonnucleoside polymerase inhibitors of hepatitis C virus. *J. Virol.* 84:2923-2934.
7. Deredge D, Li J, Johnson KA, & Wintrode PL (2016) Hydrogen/Deuterium Exchange Kinetics Demonstrate Long Range Allosteric Effects of Thumb Site 2 Inhibitors of Hepatitis C Viral RNA-dependent RNA Polymerase. *J Biol Chem.* 291(19):10078-10088.
8. Li J & Johnson KA (2016) Thumb Site 2 Inhibitors of Hepatitis C Viral RNA-dependent RNA Polymerase Allosterically Block the Transition from Initiation to Elongation. *J Biol Chem.* 291(19):10067-10077.

Appendix A

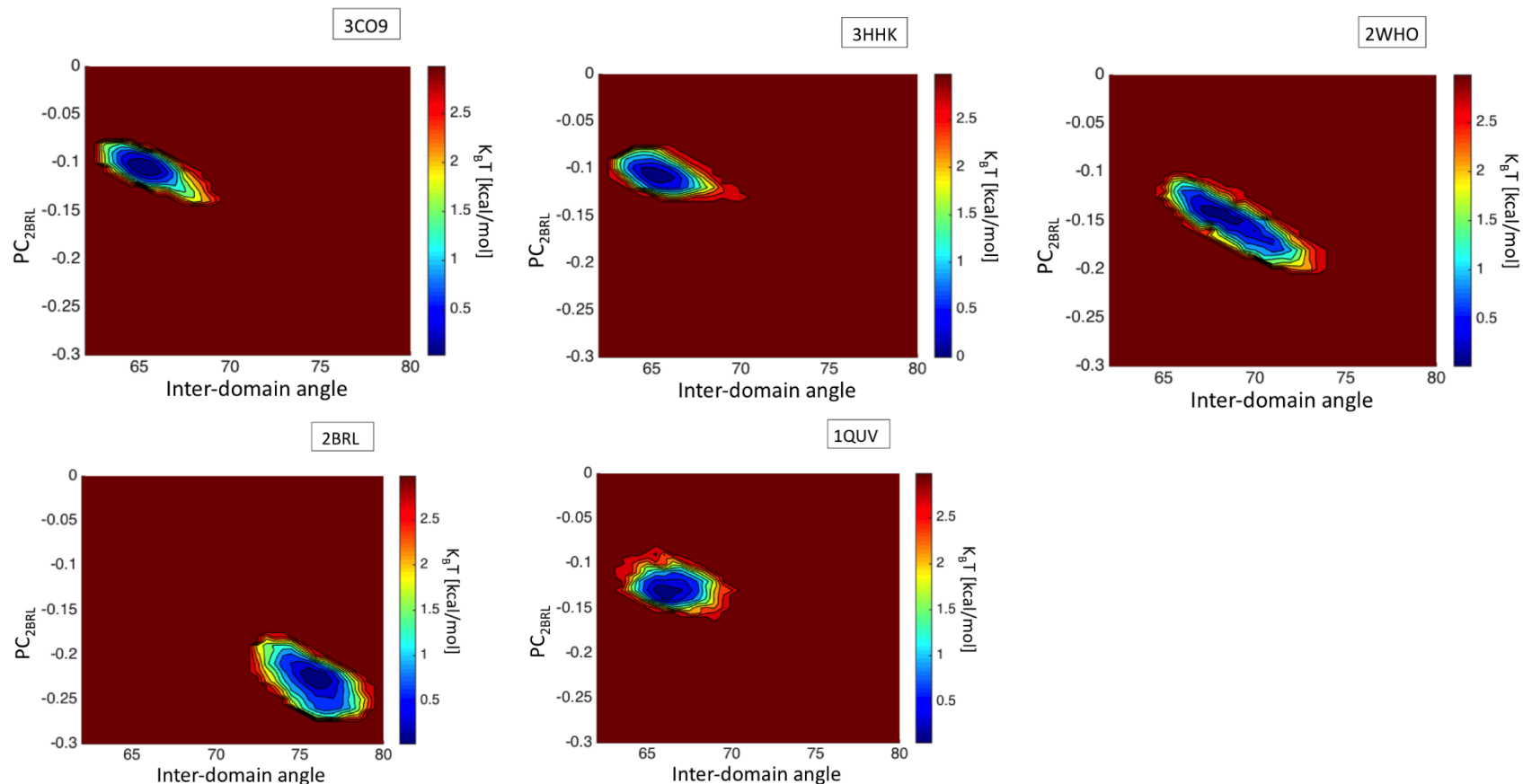


Figure A1. Free energy landscapes computed for individual trajectories.

In figure A1 are displayed the individual free energy landscapes generated from each trajectory. It is apparent that the trajectories explore distinct regions of conformational space despite representing the same system biological system and beginning from roughly the same starting point with regard to the structural coordinates shown in Table 1 of the main text. Thus, any given trajectory only provides a limited description of the overall landscape--it is only by combining all the data that

we are able to obtain a representative picture of the conformational distribution. When we examine the space explored by the distinct starting points, most systems are observed to sample only one minimum in the landscape and are thus limited in the information they can provide. This limited sampling is despite most of the simulations being performed for hundreds of ns using both conventional MD and TAMD and demonstrates the utility of employing diverse starting points for MD simulation studies.

Table A1: WHAM weights and number of snapshots in each substate for individual trajectories.

	Substate			
PDB ID	Closed	Intermediate	Open	f_i
2BRL	---	113	1887	0.062
2WHO	979	1021	---	0
3C09	1998	2	---	0.329
3HHK	1992	8	---	0.345
1QUV	1966	34	---	0.108

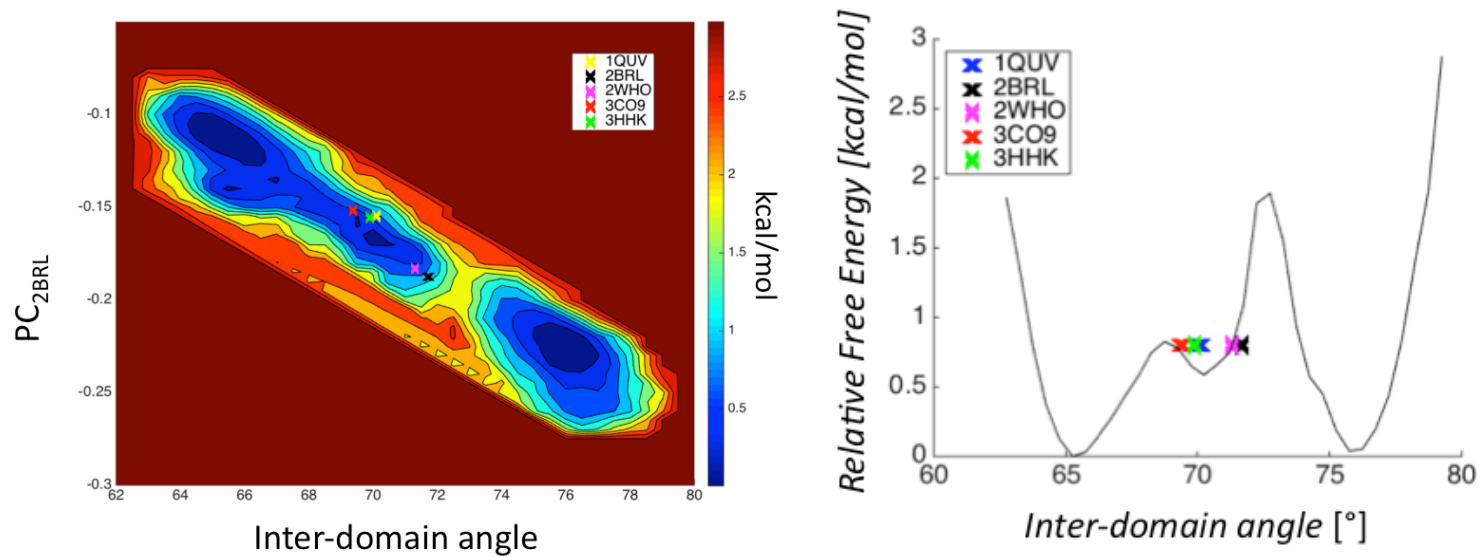


Figure A2. Location of the original PDB coordinates with respect to free energy landscapes presented in the main text. Symbols corresponding to PDB IDs for each structure are provided in the legend of each panel. Note that all of the coordinates display conformations in the vicinity of the free energy minimum in the intermediate region of the landscapes ($69^\circ < \text{inter-domain angle} < 74^\circ$). It is possible that conditions amenable to crystallization may preferentially stabilize structures with these domain angle values.

Table A2: Inter-domain angles computed for a random sample of HCV polymerase structures from the PDB

PDB ID	Inter-domain Angle (°)
3I5K	68.89
1YVF	70.51
1Z4U	70.71
3HKY	69.86
2FVC	69.26
1OS5	69.05
3GNV	69.76
2GC8	69.11
2AX0	69.99
2AWZ	70.03
2AX1	69.88
2XI2	69.35
1CSJ	71.92
1GX5	71.79
1GX6	71.46
1YVZ	71.17
1YVX	70.92
1NB4	69.55
1NB6	69.59
1NB7	68.84
2QE5	69.42
1YV2	71.15
1YUY	69.58

Table A3. Characteristic hydrogen bonds in each conformational state

		Closed – intermediate transition				Intermediate – open transition			
		Broken in the transition from closed to intermediate states		Formed in the transition from closed to intermediate states		Broken in the transition from intermediate to open states		Formed in the transition from intermediate to open - states	
		Residues	occ ^a	Residues	occ	Residues	occ	Residues	occ
Fingers domain	Rest of the fingers domain	ARG 250 – TYR 240	0.2	GLN184 – SER 180	0.1	ARG 65 – ASP 62	0.2	HSD 63 – ASP 61	0.2
		ARG 278 – GLU 230	0.4	ASP 232 – GLN 58	0.5	LEU 91 – ALA 88	0.2	GLU 171 – ARG 168	0.2
		CYS 279 – SER 3	0.2	ARG 250 – GLN 241	0.3	THR 92 – ALA 88	0.2	VAL 228 – ARG 56	0.8
		THR 286 – GLU 171	0.2	LYS 254 – GLU 237	0.2	ARG 109 – THR 92	0.6	CYS 243 – ALA 75	0.2
		ARG 56 – ASP 55	0.2			ARG 168 – ALA 97	0.7	ARG 277 – TYR 261	0.3
		VAL 228 – ARG 56	0.2			ARG 168 – SER 96	0.7	ARG 278 – GLU 230	0.7
						ARG 168 – PRO 163	0.7	THR 286 – GLU 171	0.5
						SER 226 – SER 3	0.4		
						ARG 250 – GLN 241	0.3		
						LYS 254 – GLU 237	0.2		
						SER 255 – ASP 129	0.3		
						ARG 277 – THR 136	0.7		
						SER 282 – LEU 159	0.3		

Between delta 1 and delta 2 loops and rest of fingers domain	SER 44 – THR 41	0.3	LYS 141 – THR 40	0.3	LYS 141 – THR 40	0.3	LYS 155 – THR 41	0.4
	ALA 45 – SER 42	0.2	ASN 142 – GLU 18	0.6	ASN 142 – GLU 18	0.6	PRO 156 – ARG 48	0.9
	LEU 47 – SER 44	0.3	ARG 154 – ASN 35	0.3	ARG 154 – ASN 35	0.3	ALA 157 – ARG 48	0.9
	ARG 48 – SER 44	0.2					ALA 157 – THR 41	0.6
	ARG 48 – ALA 45	0.2					ALA 157 – SER 44	0.3
	PRO 156 – ARG 48	0.3						
		0.3						
	ALA 157 – THR 41	0.6						
	ALA 157 – SER 44	0.2						
	ALA 157 – ARG 48	0.4						
	LYS 155 – THR 41	0.5						
	ALA 140 – THR 12	0.3						
	ALA 140 – CYS 14	0.4						
	ASN 142 – CYS 14							
	ASN 142 – SER 42							
Delta 2 loop	GLY 153 GLU 150	0.2	LYS 151 – GLN 148	0.3	LYS 151 – GLN 148	0.3	GLY 153 – GLU 150	0.3
Delta 1 loop	LEU 26 – ASN 24	0.3	ALA 9 – THR 7	0.3	VAL 37 – HSD 34	0.4	LEU 26 – ASN 24	0.3
	HSD 34 – LEU 21	0.4	VAL 37 – HSD 34	0.4	ALA 39 – SER 19	0.7	HSD 34 – LEU 21	0.4
	ALA 39 – GLU 18	0.4	ALA 39 – SER 19	0.7	SER 42 – GLU 17	0.2	ALA 39 – GLU 18	0.3
			SER 42 – GLU 17	0.2			SER 42 – ALA 15	0.4
							ALA 45 – SER 42	0.3
							ASN 142 – SER 42	0.8
							SER 44 – THR 41	0.2
							LEU 47 – SER 44	0.4
							ARG 48 – SER 44	0.3

Palm domain	GLN 194 – GLY 188	0.3	TYR 195 – SER 189	0.7	TYR 195 – SER 189	0.6	GLY 198 – SER 196	0.4
	GLN 194 – TYR 191	0.3	ASP 310 – LYS 211	0.2	TYR 191 – GLY 188	0.2	ARG 200 – GLY 192	0.7
	ARG 200 – GLY 192	0.2	THR 312 – LYS 211	0.2	LEU 308 – CYS 303	0.6	ARG 200-TYR 195	0.9
	ARG 200 – TYR 195	0.4	MET 313 – SER 190	0.4	ASP 310 – LYS 211	0.2	LYS 211 – THR 207	0.4
	GLY 317 – CYS 295	0.2	ASN 316 – ARG 200	0.8	THR 312 – LYS 211	0.2	GLY 317 – CYS 295	0.3
			GLU 325 – LYS 212	0.2	MET 313 – SER 190	0.4	ASP 332 – LYS 307	0.3
			PRO 350 – ARG 222	0.2	VAL 315 – SER 190	0.9	SER 368 – SER 367	0.2
			SER 365 – ARG 200	0.8	VAL 315 – GLY 192	0.8	ASN 369 – ILE 363	0.3
			CYS 366 – ASP 319	0.8	ASN 316 – ARG 200	0.9		
					SER 365 – ARG 200	0.8		
					CYS 366 – ASP 319	0.8		
						0.3		
					SER 368 – CYS 366	0.4		
					ASN 369 – THR 364	0.2		
					GLU 325 – LYS 212	0.3		
					GLU 325 – GLN 309			
Thumb domain	VAL 405 – THR 403	0.3	HSD 402 – GLU 398	0.2	THR 390 – ASP 387	0.3	VAL 405 – THR 403	0.4
	ARG 401 – GLU 398	0.3	SER 407 – GLU 398	0.4	SER 407 – GLU 398	0.4	SER 407 – VAL 405	0.2
	ASN 411 – TRP 408	0.3	HSD 467 – TYR 448	0.4	ILE 412 – LEU 409	0.2	GLY 410 – SER 407	0.3
	TYR 415 – SER 368	0.5	GLY 468 – HSD 467	0.2	HSD 467 – TYR 448	0.3	LEU 409 – ASN 406	0.4
		0.3	GLY 468 – ILE 463	0.5	GLY 468 – HSD 467	0.2	HSD 428 – THR 399	0.6
	ARG 422 – ALA 416	0.7	ARG 484 – ARG 386	0.8	ARG 484 – ARG 386	0.8	ASP 444 – ASN 406	0.9
	HSD 428 – THR 399	0.2	GLY 493 – ARG 490	0.3	ASN 527 – TYR 524	0.6	SER 473 – LEY 469	0.4
	ASP 444 – ASN 406	0.3	ASN 527 – PHE 472	0.7	ALA 529 – ARG 505	0.2	SER 473 – ALA 376	0.7
	LYS 531 – TRP 528		ASN 527 – TYR 524	0.6	TRP 528 – ARG 501	0.3	SER 473 – ASP 375	0.3
			TRP 528 – ARG 501	0.3	LYS 531 – ARG 501	0.3	GLU 481 – SER 371	0.2
			THR 390 – ASP 387	0.3			TRP 528 – ASN 527	0.3

Between palm and fingers	THR 287 – ARG 280	0.3	TYR 191 – MET 187	0.2	SER 347 – GLN 58	0.4	GLY 188 – GLN 184	0.5
	LYS 298 – GLN 58	0.2	SER 347 – GLN 58	0.4	SER 190 – MET 187	0.8	SER 189 – GLN 184	0.3
	GLY 188 – GLN 184	0.3	SER 288 – VAL 284	0.3	TYR 191 – MET 187	0.2	SER 190 – SER 189	0.2
			LYS 298 – TYR 64	0.2	TYR 191 – GLY 188	0.2	THR 287 – ARG 280	0.6
			THR 227 – ARG 222	0.4	THR 227 – ARG 222	0.5	THR 287 – SER 282	0.3
					LYS 298 – TYR 64	0.2	LYS 298 – GLN 58	0.2
					THR 287 – VAL 284	0.2	ARG 345 – ASP 61	0.4
					SER 288 – VAL 284	0.3	TYR 346 – HSD 63	0.2
Between thumb and fingers	ASP 444 – ARG 109	0.2	ARG 394 – VAL 144	0.3	ARG 394 – VAL 144	0.3	THR 390 – CYS 146	0.2
	CYS 451 – LEU 91	0.2	ARG 401 – GLU 18	0.3			ALA 400 – ASN 24	0.3
	ALA 396 – SER 27	0.3	LEU 492 – ARG 32	0.3			ASN 406 – HSD 95	0.4
	LYS 491 – HSD 33	0.4					ASP 444 – ARG 109	0.4
							CYS 451 – LEU 91	0.3
							LYS 491 – HSD 33	0.3
							ARG 503 – SER 29	0.2
Between thumb and palm	HSD 467 – GLY 198	0.3					ARG 386 – SER 368	0.2
	SER 473 – ALA 376	0.3						
	SER 473 – ASP 375	0.3						
	HSD 475 – SER 473	0.3						
	GLU 481 – SER 371	0.4						
	GLU 481 – TYR 383	0.2						

^a: occupancy (fraction of ensemble in which H-bond is present)

H-bonds that differ between the conformations can be divided into two main categories: 1) H-bonds that are unique to one conformation (colored green above) and 2) H-bonds that are specific to two conformations. In the latter group we observe the following subcategories: a) H-bonds that are common to the closed and open states: colored yellow above), b) H-bonds that are common to the intermediate and open states: purple), c) H-bonds common to the intermediate and the closed states: red).

Table A4. Characteristic hydrogen bond networks in each conformational state

Broken in the transition from closed to intermediate states	Formed in the transition from closed to intermediate states	Broken in the transition from intermediate to open states	Formed in the transition from intermediate to open - states
<p>#1 (see below)</p> <p>GLN 184 – GLY 188 – GLN 194 – TYR 191</p> <p>GLY 192 – ARG200 – TYR 195</p> <p>ARG 109 – ASP 444 – ASN 406</p> <p>ALA 376 – SER 473 – ASP 375 & HSD 475</p> <p>TYR 383 – GLU 481 – SER 371</p>	<p>ASP 310 – LYS 211 – THR 312</p> <p>ASN 316 – ARG 200 – SER 365</p> <p>HSD 402 – GLU 398 – SER 407</p> <p>TYR 448 – HSD 467 – GLY 468 – ILE 463</p> <p>PHE 472 – ASN 527 – TYR 524</p>	<p>LEU 91 – ALA 88 – THR 92 – ARG 109</p> <p>ARG 168 – SER 96, ALA 97 & PRO 163</p> <p>SER 190 – MET 187 – TYR 191 – GLY 188</p> <p>THR 287 – VAL 284 – SER 288</p> <p>ASP 310 – LYS 211 – THR 312</p> <p>MET 313 – SER 190 – VAL 315 – GLY 192</p> <p>ASN 316 – ARG 200 – SER365</p> <p>ASP 319 – CYS 366 – SER 368</p> <p>LYS 212 – GLU 325 – GLN 309</p> <p>TYR 448 – HSD 467 – GLY 468</p> <p>TRP 528 – ARG 501 – LYS 531</p>	<p>SER 42 – ALA 15, ALA 45 & ASN 142</p> <p>SER 44 – THR 41, LEU 47 & ARG 48</p> <p>#2 (see below)</p> <p>GLY 188 – GLN 184 – SER 189 – SER 190</p> <p>ARG 280 – THR 287 – SER 282</p> <p>GLY 192 – ARG 200 – TYR 195</p> <p>THR 403 – VAL 405 – SER 407 – GLY 410</p> <p>ARG 109 – ASP 444 – ASN 406</p> <p>SER 473 – LEU 469, ALA 376 & ASP 375</p>

Colors above indicate the same categories described in Table S4. The occurrence of these H-bond categories in each state is summarized by the schematic below.

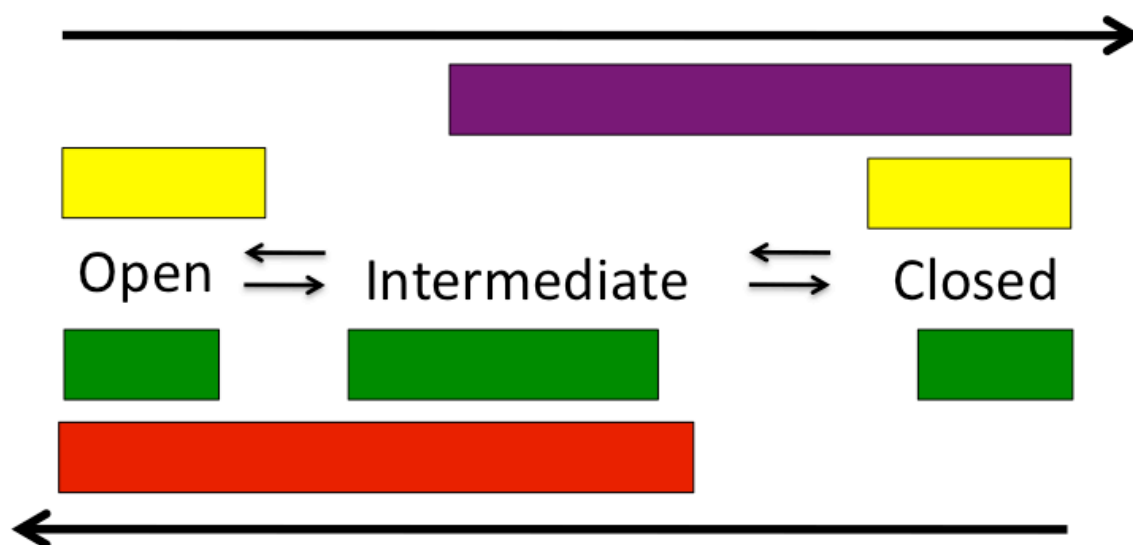
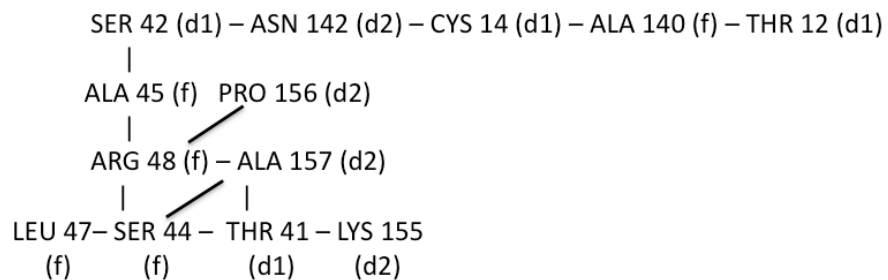


Figure A3. Schematic descriptions of H-bond categories.

Network #1 (closed state)



Network #2 (open state)

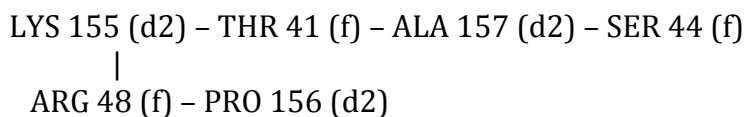


Figure A4. Schematic depictions of hydrogen bond networks in open and closed states. Text in parentheses refers to the enzyme domain or structural element in which a given residue is located as follows: fingers (f), delta 1 loop (d1), delta 2 loop (d2).

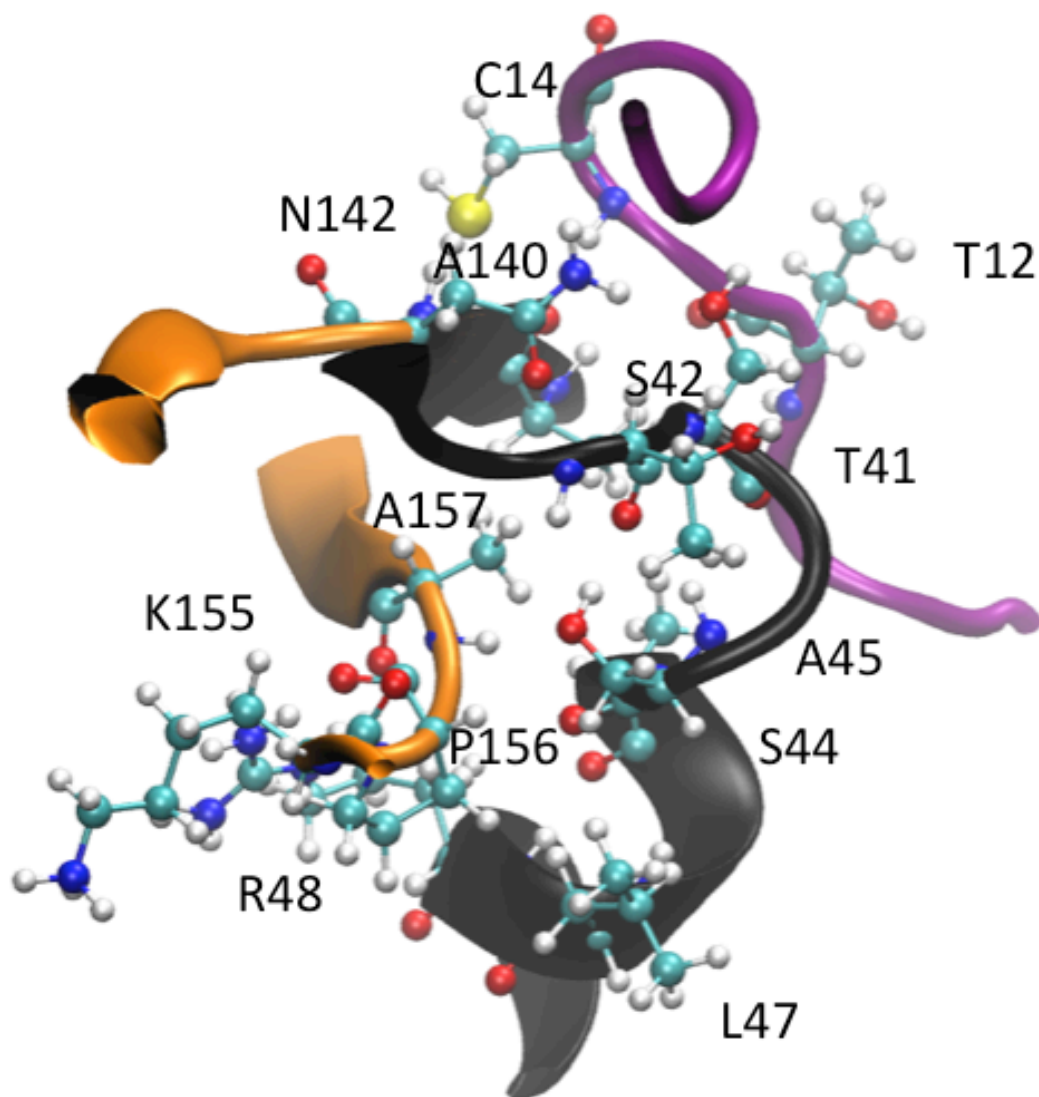


Figure A5. Molecular view of residues involved in H-bond network #1 that is schematically represented above. The polymerase is depicted in cartoon representation. The delta 1 and delta 2 loops are shown in purple and orange while residues in the fingers domain are shown in black. This extensive network occurs only in the closed conformation and involves twelve residues from the delta 1 loop, delta 2 loop and part of the adjoining fingers. The occurrence of this network emphasizes the key role of the delta 1 and 2 loops in stabilizing the closed conformation.

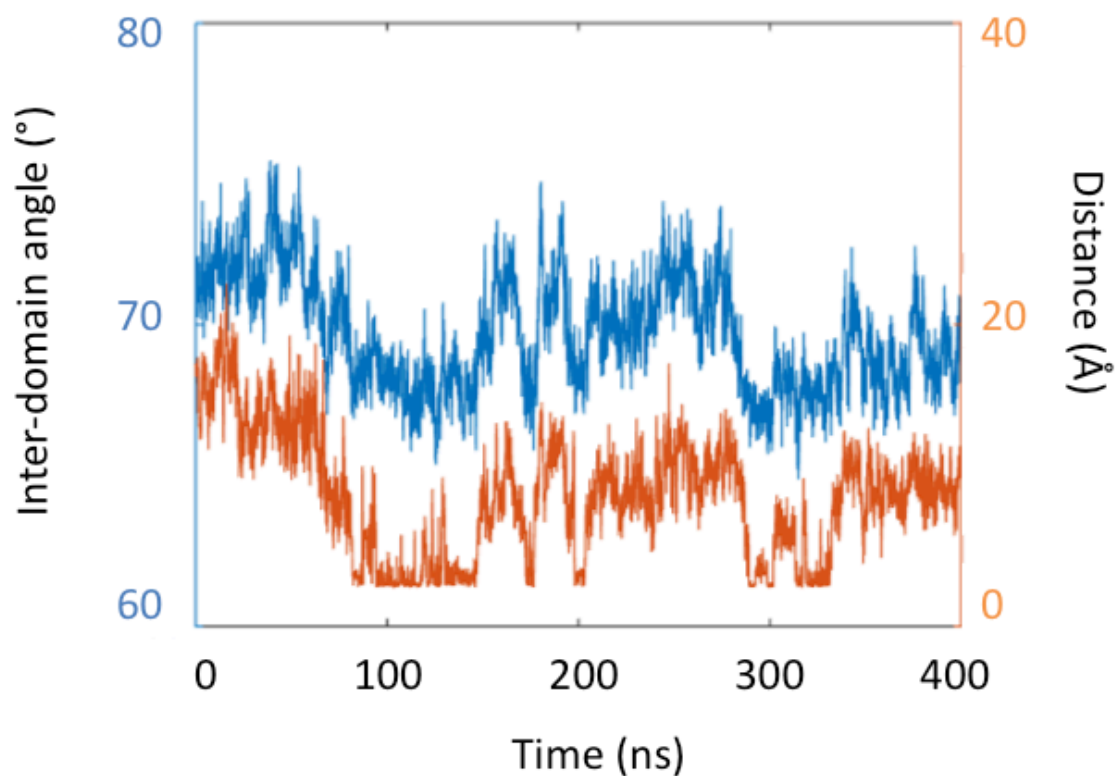


Figure A6. Length of hydrogen bond connecting ASP 444 and ARG 109 (blue) and inter-domain angle (orange) as a function of simulation time in the 2WHO trajectory.

Movies A11. Movies of largest amplitude PCs from 2BRL simulations and from ensembles for the closed, intermediate and open states are provided separately.

Appendix B

Table B1. Characteristic hydrogen bonds in each conformational state for the mutated systems. They are color coded following the same criteria that in appendix A.

		Closed – intermediate transition				Intermediate – open transition			
		Broken in the transition from closed to intermediate states		Formed in the transition from closed to intermediate states		Broken in the transition from intermediate to open states		Formed in the transition from intermediate to open states	
		residue	occ	residue	occ	residue	occ	residue	occ
Fingers domain	Rest of the fingers domain	MET 2 – SER 1	0.6	GLN 47 – SER 44	0.5	ALA 45 – SER 42	0.4	ASP 55 – SER 1	0.2
		ASP 70 – SER 66	0.2	ARG 48 – ALA 45	0.3	GLN 47 – SER 44	0.6	GLU 70 – ASP 66	0.2
		GLU 87 – ARG 81	0.6	ARG 56 – ASP 55	0.3	ARG 48 – SER 44	0.2	GLY 102 – SER 99	0.4
		SER 96 – PRO 93	0.7	HSD 63 – ASP 61	0.2	VAL 52 – SER 3	0.7	ARG 109 – THR 92	0.6
		GLY 102 – SER 99	0.4	ARG 65 – ASP 62	0.3	ARG 65 – ASP 62	0.3	LEU 111 – VAL 108	0.7
		GLY 102 – ARG 98	0.2	THR 92 – ALA 88	0.9	TYR 101 – SER 99	0.5	ASP 135 – LYS 100	0.2
		ARG 109 – THR 92	0.4	PRO 163 – LYS 98	0.2	ASP 107 – GLY 104	0.9	ASP 164 – LYS 98	0.2
		LEU 111 – VAL 108	0.8	LEU 165 – LYS 98	0.8	ASP 107 – ARG 114	0.9	VAL 167 – ASP 164	0.6
		LEU 165 – ASP 164	0.3	GLU 171 – ARG 168	1.0	PRO 163 – LYS 98	0.3	ARG 168 – LEU 165	0.3
		VAL 167 – ASP 164	0.6	LYS 172 – VAL 169	0.3	ASP 164 – LYS 100	0.2	ARG 168 – PRO 163	0.6
		ARG 168 – ALA 97	0.9	ASP 232 – GLN 58	0.2	ARG 168 – SER 96	0.8	LYS 172 – ARG 168	0.7
		ARG 168 – PRO 163	0.9	GLN 241 – SER 238	0.2	GLU 171 – ARG 168	1.0	SER 226 – ARG 56	0.4
		ARG 168 – LYS 172	0.7	ARG 250 – GLN 241	0.2	LYS 172 – VAL 169	0.3	ARG 234 – ASN 231	0.2
		LEU 91 – LYS 172	0.2	GLY 275 – ASN 273	0.2	ASP 232 – GLN 58	0.2	ARG 250 – LEU 245	0.3
		TYR 191 – MET 187	0.3	ARG 277 – TYR 261	0.4	ILE 233 – THR 229	0.6	LYS 254 – GLU 237	0.2
		SER 226 – CYS 223	0.4	ARG 278 – GLU 230	0.6	GLN 241 – SER 238	0.2	SER 269 – THR 12	0.5
		SER 226 – ARG 56	0.6	CYS 279 – SER 3	0.2	CYS 243 – ALA 75	0.2	SER 269 – PRO 13	0.4
		THR 227 – ARG 222	0.6	GLY 283 – ARG 168	0.8	GLU 248 – TRP 123	0.2	ARG 277 – THR 136	0.8
		VAL 228 – ARG 56	0.4			ARG 250 – GLN 241	0.2	GLY 283 – VAL 161	0.9
		THR 229 – GLN 58	0.4			ARG 277 – TYR 261	0.4		
		GLU 230 – SER 1	0.2			CYS 279 – SER 3	0.2		
		SER 238 – THR 235	0.2			GLY 283 – ARG 168	0.8		
		CYS 243 – ILE 239	0.3						

		ARG 250 - CYS 243 HSD 254 - GLU 237 GLY 263 - LEU 260 SER 269 - THR 12 SER 269 - PRO 13 ARG 277 - THR 136 GLY 283 - VAL 161 THR 286 - TYR 261	0.4 0.4 0.2 0.2 0.3 0.8 0.3 0.6						
	Between delta 1 and delta 2 loops and rest of fingers domain	SER 42 - GLU 19 ASN 142 - GLU 18 GLY 152 - ASP 148 ALA 157 - SER 44 TYR 162 - CYS 14	0.3 0.3 0.6 0.7 0.2	CYS 146 - MET 36 GLY 153 - GLU 150 LYS 155 - THR 41 ALA 157 - THR 41	0.2 0.5 0.9 0.8	SER 44 - THR 41 LYS 155 - THR 41	0.3 0.8	SER 42 - THR 41 SER 42 - ILE 11 ARG 43 - THR 41 ARG 43 - GLU 17 ASN 142 - ALA 15 PRO 156 - ARG 48 ALA 157 - ARG 48	0.2 0.4 0.4 0.5 0.3 0.5 0.5
	Delta 2 loop					GLY 153 - GLU 150	0.5	GLY 152 - GLN 148 GLY 153 - GLN 148	0.6 0.9
	Delta 1 loop	VAL 37 - HSD 34 CYS 39 - GLU 18	0.5 0.4	GLU 18 - GLU 17 VAL 37 - LEU 21 ALA 39 - SER 19	0.5 0.2 0.6	GLU 18 - GLU 17 VAL 37 - LEU 21 ALA 39 - SER 19 LEU 26 - ASN 24	0.6 0.2 0.6 0.4	ALA 39 - GLU 18 THR 41 - GLU 17	0.2 0.3

Palm domain	TYR 195 – GLY 188	0.2	TYR 191 – GLY 188	0.3	ARG 222 – ASP 220	0.8	PRO 349 – ARG 222	1.0
	LYS 211 – ALA 207	0.2	GLN 194 – TYR 191	0.5	ASN 291 – ASP 225	0.3	VAL 370 – ASN 369	0.3
	GLY 307 – LYS 304	0.2	ARG 222 – ASP 220	0.8	ASP 310 – LYS 211	0.2		
	GLU 325 – SER 324	0.4	THR 312 – LYS 211	0.2	THR 312 – LYS 211	0.2		
	ARG 334 – GLU 331	0.5	ASN 316 – ARG 200	0.9	PRO 349 – THR221	0.9		
	PRO 349 – ARG 222	0.8	GLY 317 – CYS 295	0.2	ASP 352 – ARG 222	0.2		
	GLU 357 – ARG 355	0.3	GLU 325 – LYS 212	0.2	SER 368 – CYS 366	0.3		
	CYS 366 – ASP 319	0.6	GLU 325 – GLN 309	0.3	ASN 369 – THR364	0.4		
			PRO 349 – THR 221	0.9				
			ASP 352 – ARG 222	0.2				
			SER 365 – ARG 200	0.7				
Thumb domain	ARG 401 – GLU 398	0.6	SER 377 – ASP 375	0.4	HSD 402 – GLU 398	0.5	ARG 401 – GLU 398	0.7
	SER 403 – GLU 398	0.3	GLY 378 – HSD 374	0.5	SER 407 – GLU 398	0.5	THR 403 – GLU 398	0.3
	ILE 412 – TRP 408	0.8	ARG 380 – HSD 374	0.2	TRP 408 – GLU 398	0.2	ILE 412 – TRP 408	0.6
	ILE 413 – LEU 409	0.7	ALA 400 – ALA 396	0.5	LEU 409 – ASN 406	0.6	ILE 413 – LEU 409	0.9
	ASN 444 – PRO 404	0.4	ARG 401 – TRP 397	0.8	GLY 410 – SER 407	0.3	ASP 444 – TRP 408	0.8
	TYR 448 – SER 288	0.2	HSD 402 – GLU 398	0.5	ASP 444 – ASN 406	0.7	GLN 446 – VAL 405	0.2
	LEU 457 – SER 455	0.3	HSD 402 – THR 399	0.2	CYS 445 – LEU 443	0.3	GLY 468 – GLU 464	0.4
	GLY 493 – ARG 490	0.3	VAL 405 – THR 403	0.3	SER 453 – ASP 444	0.2	SER 473 – LEU 469	0.7
	ARG 514 – ASP 437	0.3	SER 407 – GLU 398	0.5	ILE 462 – LEU 459	0.2	TYR 477 – ARG 422	0.9
	ARG 514 – ASP 440	0.2	TRP 408 – GLU 398	0.2	LEU 466 – TYR 448	0.2	GLU 481 – SER 371	0.2
	GLY 515 – ILE 512	0.2	GLY 410 – SER 407	0.3	HSD 467 – ILE 413	0.3	CYS 488 – VAL 485	0.4
	ALA 519 – GLY 516	0.2	CYS 445 – LEU 443	0.3	ALA 471 – ARG 422	0.2	GLY 515 – LEU 512	0.2
	ASN 527 – ARG 508	0.2	ILE 462 – LEU 459	0.2	HSD 475 – SER 473	0.2	ASN 527 – ARG 422	0.6
			LEU 466 – TYR 448	0.2	PHE 526 – GLY 522	0.3	ASN 527 – PHE 472	0.8
			HSD 475 – SER 473	0.2	LYS 531 – TRP 528	0.4	TRP 528 – ASN 527	0.6
			CYS 488 – THR 389	0.4	LYS 531 – ARG 501	0.3		
			ARG 517 – GLU 455	0.3				
			PHE 526 – GLY 522	0.3				
			LYS 531 – ARG 501	0.3				

Between palm and fingers	THR 287 - SER 282	0.2	THR 287 - ALA 281	0.2	THR 287 - ALA 281	0.2	SER 226 - CYS 223	0.5
	THR 287 - VAL 284	0.5	ASN 291 - ARG 280	0.2			THR 287 - SER 282	0.7
	TYR 346 - GLN 58	0.3	ARG 304 - ASP 66	0.3				
			ARG 304 - GLU 70	0.6				
Between thumb and fingers			ARG 345 - ASP 61	0.3				
	THR 390 - CYS 146	0.6	TRP 397 - GLU 17	0.2	TRP 397 - GLU 17	0.2	ALA 400 - ASN 24	0.4
	ARG 394 - GLU 143	0.2	ARG 401 - GLU 17	0.5	ARG 401 - GLU 18	0.2		
			LYS 491 - HSD 33	0.7	ARG 401 - GLU 17	0.5		
					LYS 491 - HSD 33	0.7		
					LEU 492 - ARG 32	0.3		
Between thumb and palm					LEU 492 - HSD 33	0.7		
	TYR 415 - SER 368	0.9					SER 478 - GLU 361	0.3
	HSD 467 - ALA 198	0.3					GLU 481 - TYR 383	0.2
							ARG 484 - ASN 369	0.2

Table B2. Characteristic hydrogen bonds in each conformational state for the inhibitor bound systems. They are color coded following the same criteria that in appendix A.

		Closed – intermediate transition				Intermediate – open transition			
		Broken in the transition from closed to intermediate states		Formed in the transition from closed to intermediate states		Broken in the transition from intermediate to open states		Formed in the transition from intermediate to open - states	
		residue	occ	residue	occ	residue	occ	residue	occ
Fingers domain	Rest of the fingers domain	TYR 64 – ASP 61	0.2	LEU 47 – SER 44	0.3	ARG 48 – SER 44	0.2	ASP 55 – SER 1	0.2
		GLU 70 – ASP 66	0.3	HSD 63–ASP 61	0.2	HSD 63 – ASP 61	0.2	ARG 56 – ASP 55	0.2
		ARG 109 – LYS 90	0.4	ARG 109 – THR 92	0.4	LEU 91 – ALA 88	0.2	TYR 64 – ASP 61	0.2
		GLU 230 – SER 1	0.4	SER 226 – SER 3	0.4	LYS 151–GLN148	0.3	ARG 65 – ASP 62	0.3
		ASP 232 – GLN 58	0.2	GLU 237 – ARG 234	0.2	PRO 156 –ARG48	0.2	GLU 70 – ASP 66	0.2
		GLU 248 – TRP 123	0.2	CYS 243 – ILE 239	0.2	VAL 228 –ARG56	0.2	ASP 135 – LYS 100	0.2
		ARG 250 – GLN 241	0.2	ARG 250 – TYR 240	0.2	CYS 243 - ILE239	0.2	GLY 166 – ASP 164	0.2
		ASN 268 – LEU 10	0.2	ARG 277 – THR 136	0.5	CYS 243 - ALA75	0.2	THR 229 – GLN 58	0.2
		CYS 279 – SER 3	0.2	GLY 283 – VAL 161	0.7	ARG 250-TYR240	0.2	GLU 230 – SER 1	0.3
				THR 286 – ARG 280	0.2	LYS 270 -ASN268	0.2	ARG 250 – GLN 241	0.2
				THR 286 – TYR 261	0.2	GLY 271-ASN268	0.5	THR 257 – GLU 237	0.4
						ASN 273 - THR7	0.2	SER 269 – THR 267	0.2
						GLY 275-ASN273	0.2	CYS 274 – ASN 268	0.2
						THR 286-GLU171	0.2	ARG 277 – TYR 261	0.2
								CYS 279 - SER 3	0.2
								THR 287 – SER 282	0.7

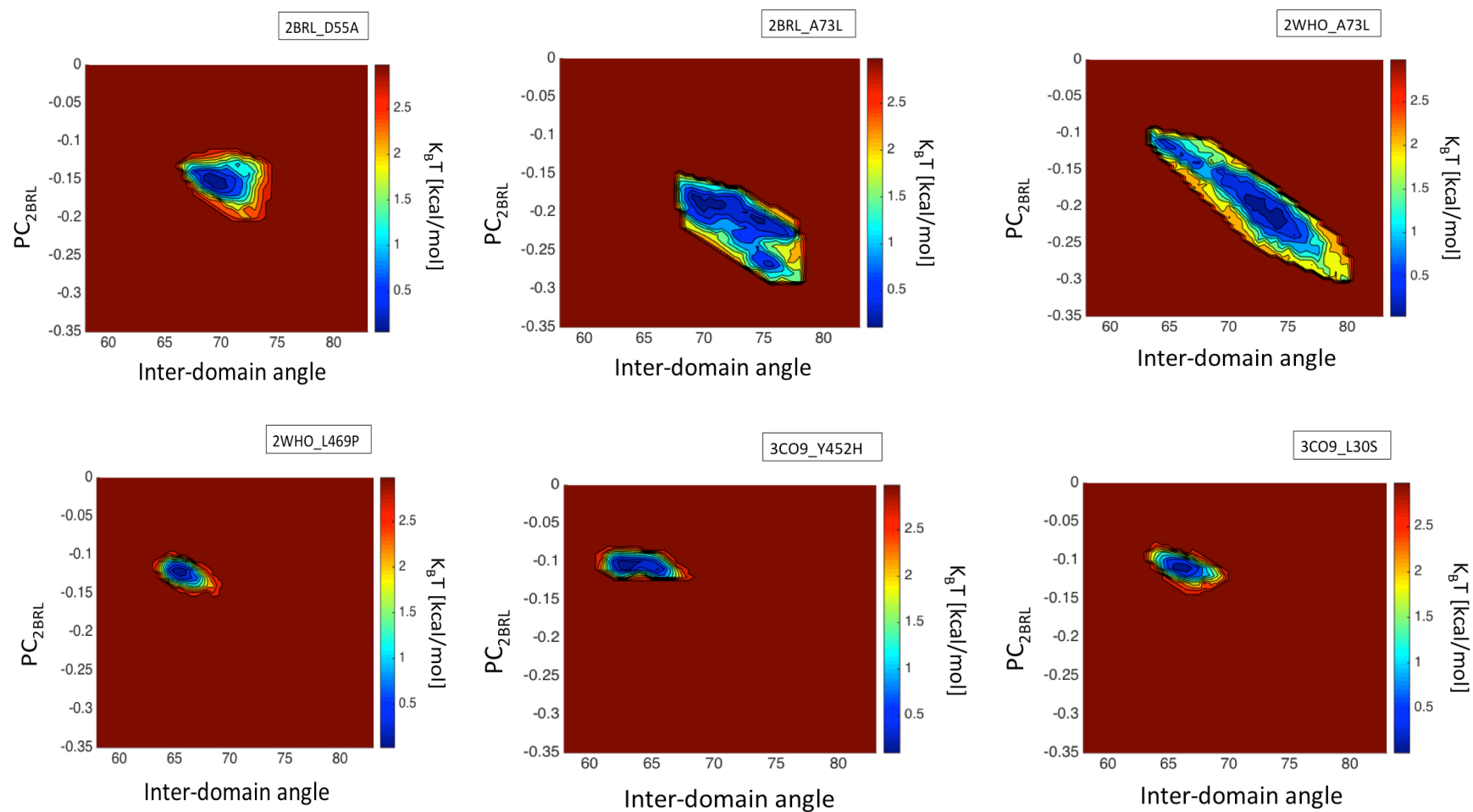
	Between delta 1 and delta 2 loops and rest of fingers domain	SER 44 – THR 41 GLN 49 – ILE 11 ASN 142 – ALA 15 ALA 157 – SER 44 ALA 157 – ARG 48	0.2 0.4 0.2 0.4 0.2	SER 42 – ALA 16 ARG 43 – THR 41 ARG 43 – ALA 15 ARG 154 – GLU 18 ASN 142 – SER 42 LYS 155 – THR 41	0.2 0.3 0.2 0.2 0.2 0.3	SER 42 – ALA 16 ARG 43 – ALA 15 ALA 140-THR 12 ALA 140-CYS 14 LYS 155 –THR 41	0.3 0.2 0.3 0.4 0.2		
	Delta 2 loop	GLY 152 – GLN 148 GLY 153 – GLU 150	0.2 0.2	LYS 151 – GLN 148	0.2				
	Delta 1 loop	SER 19 – GLU 18 LEU 30 – SER 27 HSD 34 – LEU 21 THR 41 – GLU 17	0.2 0.2 0.2 0.2	VAL 37 – HSD 34 VAL 37 – LEU 21	0.2 0.2			SER 19 – GLU 18 LYS 20 – GLU 18 LEU 30 – SER 27 HSD 33 – ARG 32 HSD 34 – ARG 32 HSD 34 – SER 27 HSD 34 – ASN 28 MET 36 – HSD 33 SER 44 – THR 41	0.4 0.4 0.3 0.2 0.3 0.3 0.2 0.2 0.2
Palm domain		ASN 369 – ILE 363	0.2	TYR 195 – GLY 192 TYR 195 – SER 189 LEU 308 – CYS 303 ASN 316 – ARG 200 GLY 317 – CYS 295 PRO 349 – ARG 222 SER 365 – ARG 200 CYS 366 – ASP 319 SER 368 – CYS 366 ASN 369 – THR 364 ASN 369 – GLU 361	0.3 0.7 0.3 0.5 0.4 0.6 0.4 0.2 0.3 0.3 0.2	TYR 191-GLY 188 ARG 200-TYR 195 PRO 349-THR 221 ASP 352-ARG 222 CYS 366 –ASP 319 ASN 369-THR 364 ASN 369-GLU 361	0.3 0.2 0.3 0.2 0.2 0.3 0.2	ASP 310 – LYS 211 THR 312 – LYS 211 ASP 318 – THR 221 VAL 370 – ASN 369	0.2 0.2 0.6 0.5

Thumb domain	VAL 372 – SER 371	0.3	GLY 378 – HSD 374	0.2	HSD 402 – GLU 398	0.2	SER 371 – GLU 361	0.2
	ARG 386 – SER 368	0.6	ARG 380 – HSD 374	0.2	VAL 405 – THR 403	0.3	VAL 372 – SER 371	0.3
	TRP 420 – ARG 386	0.5	ASP 387 – ARG 386	0.5	TRP 408 – ASN 406	0.2	ASN 406 – THR 403	0.2
	HSD 428 – THR 399	0.3	THR 389 – ASP 387	0.3	LEU 409 – ASN 406	0.2	SER 407 – THR 403	0.2
	HSD 467 – ILE 413	0.2	VAL 405 – THR 403	0.2	GLY 410 – ASN 406	0.3	SER 407 – GLU 398	0.2
	PHE 472 – LEU 469	0.2	TRP 408 – ASN 406	0.2	TYR 415 – SER 368	0.4	TRP 408 – THR 399	0.2
	ARG 484 – ASP 387	0.2	GLY 410 – ASN 406	0.3	ARG 422 – ALA 416	0.2	ILE 412 – LEU 409	0.3
	LYS 531 – TRP 528	0.3	ARG 422 – ALA 416	0.2	GLN 446 – VAL 405	0.4	HSD 428 – THR 399	0.6
			GLN 446 – VAL 405	0.4	PHE 472 – GLY 468	0.4	GLY 468 – HSD 467	0.4
			PHE 472 – GLY 468	0.4	SER 473 – LEU 469	0.3	CYS 488 – THR 389	0.4
			LEU 474 – SER 470	0.2	LEU 474 – SER 470	0.2	GLN 514 – LEU 511	0.2
			GLY 493 – ARG 490	0.2	HSD 475 – SER 473	0.3	VAL 530 – ASN 527	0.3
			LYS 531 – ARG 501	0.2	SER 476 – HSD 374	0.4	LYS 531 – TRP 528	0.2
					GLU 481 – SER 371	0.4		
					GLU 481 – TYR 383	0.3		
					TRP 528 – LEU 525	0.2		
					LYS 531 – ARG 501	0.2		
Between palm and fingers	SER 288 – VAL 284	0.5			SER 226 – CYS 223	0.5	THR 227 – ARG 222	0.3
	CYS 289 – LEU 285	0.8			THR 287 – VAL 284	0.3	ASN 291 – ASP 225	0.2
	ASN 291 – ARG 280	0.3			THR 287 – ALA 281	0.6	LYS 298 – TYR 64	0.2
	THR 294 – ARG 280	0.3						
	ASP 318 – THR 221	0.2						
Between thumb and fingers	ALA 396 – SER 27	0.2	LYS 491 – HSD 33	0.3			LEU 492 – ARG 32	0.2
	ASN 406 – HSD 95	0.2					ARG 503 – SER 29	0.6
	ASP 444 – ARG 109	0.3						
	TYR 448 – TYR 176	0.3						
	CYS 451 – LEU 91	0.5						
	ARG 503 – SER 29	0.2						
Between thumb and palm			TYR 415 – SER 368	0.3	LEU 466 – GLY 198	0.3	SER 478 – GLU 361	0.3
			LEU 466 – GLY 198	0.2			ARG 484 – ASN 369	0.8

Table B3. RMSD comparison of the PDBs representative of the different minima.

		free			ligands						mutations		
		closed	inter	open	closed	inter	open	D	E	F	closed	inter	open
free	Closed	0	3.6	4.1	3.1	4.3	4.7	3.3	3.2	4.9	2.8	3.8	6.3
	Inter	3.6	0	2.9	2.7	4.9	3.1	3.1	2.4	5.1	3.1	3.5	6.1
	open	4.1	2.9	0	4.0	4.0	2.4	4.3	3.1	4.0	3.8	2.2	4.1
ligands	Closed	3.1	2.7	4.0	0	4.4	3.8	2.4	2.6	4.8	3.1	3.8	6.4
	Inter	4.3	4.9	4.0	4.4	0	4.4	4.2	4.5	2.2	5.0	3.1	4.6
	open	4.7	3.1	2.4	3.8	4.4	0	4.5	3.0	4.1	4.5	3.2	4.7
	VGI	3.3	3.1	4.3	2.4	4.2	4.5	0	3.4	5.1	3.0	3.9	6.7
	3MS	3.2	2.4	3.1	2.6	4.5	3.0	3.4	0	4.7	3.2	3.5	5.9
	POO	4.9	5.1	4.0	4.8	2.2	4.1	5.1	4.7	0	5.7	3.5	3.9
mutations	Closed	2.8	3.1	3.8	3.1	5.0	4.5	3.0	3.2	5.7	0	3.8	6.4
	Inter	3.8	3.5	2.2	3.8	3.1	3.2	3.9	3.5	3.5	3.8	0	3.7
	open	6.3	6.1	4.1	6.4	4.6	4.7	6.7	5.9	3.9	6.4	3.7	0

Figure B1. Individual Free Energy landscapes of the inhibitor bound systems.



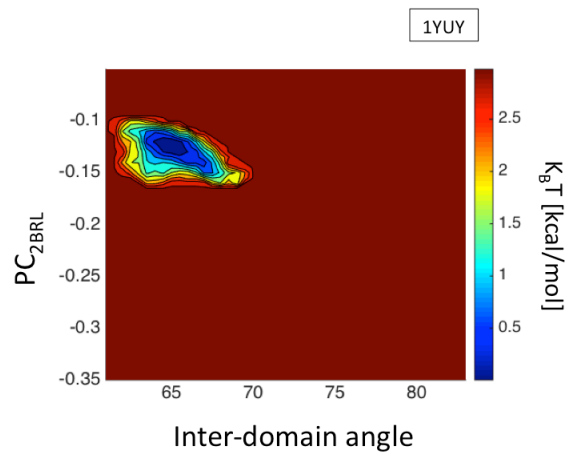
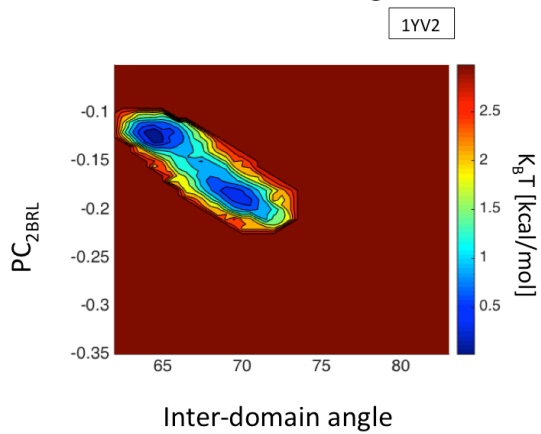
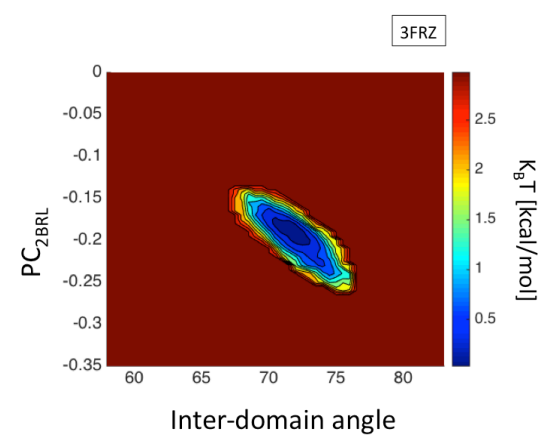
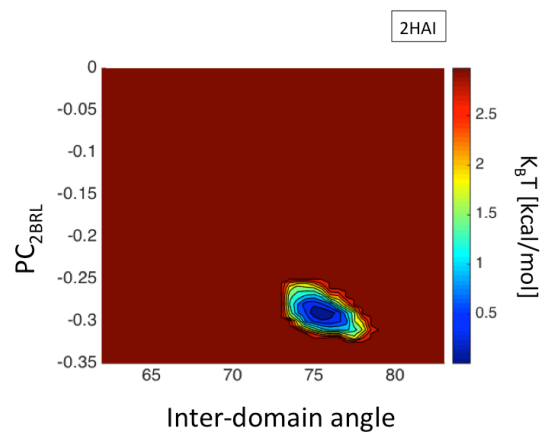
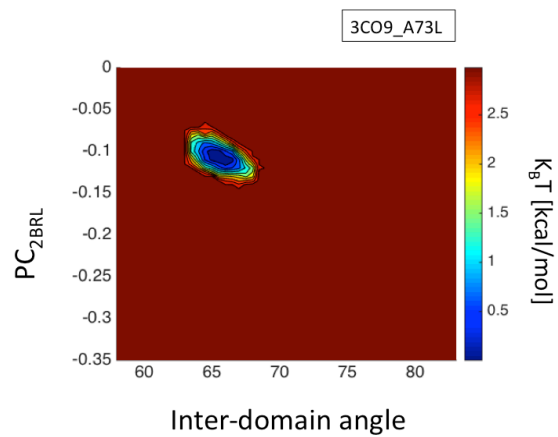


Figure B2. Individual Free Energy landscapes of the mutated systems.

