Understanding Drug Resistance and Antibody Neutralization Escape in Antivirals: A Dissertation

UNDERSTANDING DRUG RESISTANCE AND ANTIBODY NEUTRALIZATION ESCAPE IN ANTIVIRALS

A Dissertation Presented

By

KRISTINA PRACHANRONARONG

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

APRIL 6, 2016

MD/PHD PROGRAM

UNDERSTANDING ANTIVIRAL DRUG RESISTANCE AND MOLECULAR RECOGNITION IN INFLUENZA BROADLY NEUTRALIZING ANTIBODIES

A Dissertation Presented By

KRISTINA PRACHANRONARONG

This work was undertaken in the Graduate School of Biomedical Sciences
MD/PHD PROGRAM

The signature of the Thesis Advisor signifies
validation of Dissertation content

CELIA A. SCHIFFER, PH.D., Thesis Advisor

The signatures of the Dissertation Defense Committee signify
Completion and approval as to style and content of the Dissertation

DAN BOLON, PH.D., Member of Committee

FRANCESCA MASSI, PH.D., Member of Committee

JENNIFER WANG, M.D., Member of Committee

SCOTT GARMAN, PH.D., Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee

LAWRENCE STERN, PH.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduation requirements of the School.

ANTHONY CARRUTERS, PH.D.
Dean of the Graduate School of Biomedical Sciences
APRIL 6, 2016

TABLE OF CONTENTS

Page

ABSTRACT vii
LIST OF TABLES ix
LIST OF FIGURES x
LIST OF THIRD PARTY COPYRIGHTED MATERIAL xiv
PREFACE xvii
1 Introduction 1
1.1 Therapeutic Strategies Against Viruses 1
1.2 Structure Based Drug Design in Antivirals 7
1.3 Drug Resistance and Antibody Neutralization Escape in Antivirals 8
1.3.1 Resistance to Competitive Small Molecule Antivirals and the Substrate Envelope 10
1.3.2 Resistance to Antibodies as Therapeutics: Antibody Neutralization Escape 14
1.4 Protein Dynamics in Molecular Recognition 17
1.4.1 Role of Protein Dynamics in Antiviral Resistance 18
1.4.2 Role of Protein Dynamics in Antibody Neutralization Escape 19
1.5 Scope of Thesis 19
2 THE SUBSTRATE ENVELOPE HYPOTHESIS DESCRIBES DIFFERENTIAL PATTERNS OF DRUG RESISTANCE IN N1 AND N2 NEURAMINIDASE 22
2.1 Abstract 22
2.2 Introduction 23
2.3 Results 29
2.3.1 Description of Molecular Dynamics Simulations 29
2.3.2 Static and Dynamic Substrate and Inhibitor Envelopes 35
2.3.3 Differences in Van der Waals Contacts Correlate with Differential Patterns of Drug Resistance in N1 and N2 45
2.3.4 Hydrogen Bond Interactions 57
2.4 Discussion 58
2.5 Methods 61
2.5.1 Influenza Neuraminidase Substrate and Inhibitor Complex Structures 61
2.5.2 Structure Preparation 62
2.5.3 Molecular Dynamics Simulation Protocol 63
2.5.4 Root Mean Squared Deviation (RMSD) 63
2.5.5 Root Mean Squared Fluctuation (RMSF) and Simulation-Derived Temperature Factors 64
2.5.6 Dynamic Substrate Envelope 64
2.5.7 Van der Waals Contact Potential 65

2.5.8 Hydrogen Bond Calculations 65
2.5.9 Plots and Figures 66
3 COMPARATIVE ANALYSIS OF HEPATITIS C NS3/4A INHIBITOR STRUCTURES 67
3.1 Abstract 67
3.2 Introduction 68
3.2.1 HCV NS3/4A Protease and Current Inhibitors 70
3.2.2 Drug Resistance to HCV NS3/4A Protease Inhibitors 79
3.3 Results 83
3.3.1 Substrate Envelope and V_{IN} and V_{OUT} 83
3.3.2 Inhibitor-Protease Contacts at the Active Site 91
3.3.3 Hydrogen Bond Interactions 102
3.4 Discussion 108
3.5 Methods 109
3.5.1 Substrate Envelope and V_{IN} and V_{OUT} Calculations 109
3.5.2 Van der Waals Contact Potential Energy 111
3.5.3 Hydrogen Bond Interactions 112
3.5.4 Plots and Figures 112
4 IDENTIFICATION OF INFLUENZA A VIRUS CANDIDATE RESISTANCE MUTATIONS TO BROADLY NEUTRALIZING ANTIBODY 113
4.1 Abstract 113
4.2 Introduction 114
4.3 Results 118
4.3.1 Serial passage of influenza in the presence of F10 monoclonal antibody 118
4.3.2 Sequence analysis reveals candidate F10 escape mutations 122
4.3.3 Structural mapping of candidate F10 hemagglutinin escape mutants 122
4.3.4 Potential impact of specific mutations on hemagglutinin function 123
4.3.5 Neuraminidase mutant 137
4.3.6 PA, PB1, and synonymous mutants 137
4.4 Discussion 138
4.5 Conclusion 143
4.6 Methods 144
4.6.1 Cells, virus stocks, and chemicals. 144
4.6.2 Viral titer determination by plaque assay 144
4.6.3 Determination of the ED_{50} for F10 antibody 145
4.6.4 Viral culture 145
4.6.5 High-throughput sequencing 146
4.6.6 Bioinformatics analysis 146
4.6.7 Population genetic analysis 147
4.6.8 Structural analysis methods 148

5 STRUCTURAL BASIS OF AN INFLUENZA HEMAGGLUTININ STEM-DIRECTED ANTIBODY RETAINING THE G6 IDIOTYPE 150
5.1 Abstract 150
5.2 Introduction 151
5.2.1 An example of a stem-directed broadly neutralizing antibody that neutralizes influenza and binds G6 154
5.3 Results 155
5.3.1 Elucidation of the G6 Idiotype: Preferential binding to germline-like V-segments 155
5.3.2 BCR Cross-Linking: F10 antibody maturation abrogates the G6 idiotype 159
5.3.3 Influenza stem-directed broadly neutralizing antibody binding to G6 versus HA 165
5.3.4 The crystal structures of the D80 Fab fragment and the D80-HuG6.3 Fab fragment complex 166
5.3.5 Intermolecular binding interface 173
5.4 Discussion 181
5.5 Methods 183
5.5.1 Panning the naïve human scFv-phagemid library against beads coupled with G6 183
5.5.2 B cell cross-linking assay and binding kinetics assay 183
5.5.3 Protein expression, purification, and crystallization 184
5.5.4 Diffraction data collection and structure solution 184
5.5.5 Structural data analysis 185
6 DISCUSSION 186
6.1 Impact of uncertainty in coordinates on molecular dynamics simulation analysis 186
6.2 Assessing equilibration of molecular dynamics simulations 189
6.3 Dynamic substrate envelopes compared to static substrate envelopes 197
6.4 Generalization of N1 and N2 NA sequences examined in molecular dynamics simulations compared to all N1 and N2 NA sequences. 198
6.5 Role of electrostatics in analysis of drug resistance residues in influenza neuraminidase 199
6.6 Incorporating the substrate envelope hypothesis and protein dynamics into drug design 200
6.7 Understanding additional drug resistance mutations in influenza neuraminidase N1 and N2 subtypes 204
6.8 Understanding molecular recognition and antibody neutralization escape in influenza broadly neutralizing antibodies 206
6.9 Impact of crystal contacts on D80 Fab fragment antigen binding site. 209

6.10 Understanding binding of D80 to HA 210
6.11 Using broadly neutralizing antibodies in influenza therapy 217
6.12 Concluding remarks 219
7 APPENDIX CHAPTER I 220
7.1 Molecular Dynamics Protocol 220
7.1.1 Protein Preparation Wizard Protocol 220
7.1.2 Molecular Dynamics Minimization Protocol 220
7.1.3 Molecular Dynamics Protocol 221
8 APPENDIX CHAPTER II 223
8.1 Determination of D80 Fab Fragment Resolution Cutoff 223
9 APPENDIX CHAPTER III 231
9.1 Hepatitis C NS3/4A Protease Inhibitor Ki Determination Assay: Data and Modeling 231
10 APPENDIX CHAPTER IV 238
10.1 Influenza Neuraminidase Baculovirus Insect Cell Expression and Purification 238
11 APPENDIX CHAPTER V 244
11.1 Influenza Neuraminidase 1D NMR Cleavage Assay 244
12 BIBLIOGRAPHY 253

ABSTRACT

Antiviral drug resistance is a major problem in the treatment of viral infections, including influenza and hepatitis C virus (HCV). Influenza neuraminidase (NA) is a viral sialidase on the surface of the influenza virion and a primary antiviral target in influenza. Two subtypes of NA predominate in humans, N1 and N2, but different patterns of drug resistance have emerged in each subtype. To provide a framework for understanding the structural basis of subtype specific drug resistance mutations in NA, we used molecular dynamics simulations to define dynamic substrate envelopes for NA to determine how different patterns of drug resistance have emerged in N1 and N2 NA. Furthermore, we used the substrate envelope to analyze HCV NS3/4A protease inhibitors in clinical development. In addition, influenza hemagglutinin (HA) is a primary target of neutralizing antibodies against influenza. Novel broadly neutralizing antibodies (BnAbs) against the stem region of HA have been described and inhibit several influenza viral subtypes, but antibody neutralization escape mutations have emerged. We identified potential escape mutations in broadly neutralizing antibody F10 that may impact protein dynamics in HA that are critical for function. We also solved crystal structures of antibody fragments that are important for understanding the structural basis of antibody binding for influenza BnAbs. These studies can inform the design of improved therapeutic strategies against viruses by incorporating an understanding of structural elements that are critical for function, such as substrate processing and protein

dynamics, into the development of novel therapeutics that are robust against resistance.

LIST OF TABLES

Page

Table 2.1 Drug resistance mutations in NA subtypes N1 and N2 that have been observed in viruses clinically or during influenza surveillance. 25

Table 2.2 Drug resistance mutations in NA subtypes N1 and N2 that have been observed experimentally. 26

Table 2.3 Additional residue positions in NA associated with decreased neuraminidase inhibitor susceptibility (41). 27

Table 2.4 Table of NA ligand complexes that were analyzed. 30

Table 3.1 Sites of drug resistance mutations to HCV NS3/4A protease inhibitors 82

Table 4.1 Top-ranking polymorphic sites (posterior probability s>0>99%) 124

Table 5.1 Crystallographic statistics for the D80-G6 complex crystal structure and D80 unbound crystal structure. 168

Table 6.1 Starting crystal structures for molecular dynamics simulations 188

Table 8.1 Overall R_{work}/R_{free} for D80 Refinement 226

Table 8.2 CC*, $CC^{1/2}$, CC_{work}, and CC_{free} for D80 Refinement 227

Table 8.3 High resolution limit for refinement set at 2.7 Å 228

Table 8.4 High resolution limit for refinement set at 2.59 Å 229

Table 8.5 High resolution limit for refinement set at 2.52 Å 230

LIST OF FIGURES

Page

Figure 1.1 Influenza virus life cycle and targets for therapeutic strategies against influenza (9). 4

Figure 1.2 Hepatitis C virus life cycle and points of intervention for therapeutic strategies (14). 5

Figure 1.3 Substrate and inhibitor envelopes in HIV protease (33, 49). 12

Figure 2.1 Overall root mean squared deviation (RMSD) calculations for each MD simulation were performed over 100 ns. 31

Figure 2.2 Experimental average B factors and RMSF values for simulations. 33

Figure 2.3 Influenza NA structure and sites of drug resistance. 37

Figure 2.4 Two-dimensional structures of substrates and inhibitors of NA. 39

Figure 2.5 Dynamic substrate and inhibitor envelopes for NA. 41

Figure 2.6 Average ligand root mean squared fluctuation (RMSF) in NA. 43

Figure 2.7 Van der Waals interactions in NA. 47

Figure 2.8 Van der Waals interactions for drug resistance residues in NA. 49

Figure 2.9 Intermolecular hydrogen bonding interactions during MD simulations. 51

Figure 2.10 Increased van der Waals contact potential energies of inhibitors compared to substrates. 53

Figure 3.1 Cartoon representation of the hepatitis C NS3/4A helicase-protease structure. 71

Figure 3.2 Hepatitis C NS3/4A protease inhibitors in clinical use or development. 73

Figure 3.3 Viral substrates of HCV NS3/4A protease share a conserved binding mode in the active site. 85

Figure 3.4 Inhibitors protruding outside of the substrate envelope in HCV NS3/4A protease. 87

Figure 3.5 Volume of substrates and inhibitors inside (V_{IN}) and outside (V_{OUT}) of the substrate envelope by substrate moiety. 89

Figure 3.6 Substrate van der Waals contacts mapped on the surface of the NS3/4A protease active site for each substrate. 92

Figure 3.7 Inhibitor van der Waals contacts mapped on the surface of the NS3/4A protease active site. 94

Figure 3.8 Substrate envelope and van der Waals surface representations for modeled HCV NS3/4A protease inhibitors 96

Figure 3.9 Differential van der Waals interactions on the binding surface of HCV NS3/4A protease. 98

Figure 3.10 Hydrogen bond interactions between ligands and backbone atoms in the HCV NS3/4A protease active site 104

Figure 3.11 Hydrogen bond interactions between ligands and side chain atoms in the HCV NS3/4A protease active site. 106

Figure 4.1 Experimental design and viral titers for F10 trajectories. 120

Figure 4.2 Significant mutations arising in influenza A virus under selection with the broadly neutralizing antibody F10. 125

Figure 4.3 Candidate escape mutations identified in the F10 trajectories are mapped onto the structure of HA. 127

Figure 4.4 The N203V mutation in HA is located in the HA receptor binding site. 129

Figure 4.5 The N460S mutation in HA is located adjacent to the fusion peptide in HA. 133

Figure 4.6 S123G is located in a hinge region of conformational change in an early fusion intermediate of HA1. 135

Figure 4.7 The F10 epitope on the stem region of HA is highly conserved. 141

Figure 5.1 The G6 idiotype is defined by the CDR2 loop of the antibody heavy chain. 157

Figure 5.2 Murine G6 and humanized G6.3 cross-link B cell receptors that bear the G6 idiotype. 161

Figure 5.3 The G6 idiotype appears on select influenza stem-directed broadly neutralizing antibodies. 163

Figure 5.4 Crystal structures of a stem-directed broadly neutralizing antibody that bears the G6 idiotype: D80 in complex with G6. .. 169

Figure 5.5 The bound and unbound structures of the D80 Fab fragment do not show significant differences. 171

Figure 5.6 Overview of the binding epitopes and intermolecular interactions between D80 and G6. 175

Figure 5.7 Characterizing the pocket in the binding site of G6. 177

Figure 5.8 Hydrogen bond and pi-stacking interactions between D80 and G6. 179

Figure 6.1 Potential Energy Over the Trajectory for N1 NA Molecular Dynamics Simulations 191

Figure 6.2 Potential Energy Over the Trajectory for N2 NA Molecular Dynamics Simulations 193

Figure 6.3 Per Chain Root Mean Squared Deviation for Molecular Dynamics Simulations 195

Figure 6.4 Competition ELISA assay between phage-bound broadly neutralizing antibodies and soluble broadly neutralizing antibodies. 212

Figure 6.5 Analysis of crystal structure of H5 HA and F10 single chain Fv fragment and HA binding epitope. 214

Figure 9.1 Ki values of seven novel HCV NS3/4A protease inhibitors with wild-type genotype 1a HCV NS3/4A protease. 232

Figure 9.2 Molecular dynamics simulation results for models of HCV NS3/4A protease bound to novel inhibitors. 234

Figure 9.3 Chemical structures of novel HCV NS3/4A protease inhibitors..... 236

Figure 10.1 Expression and purification of N1 NA using a nickel-NTA column analyzed using SDS-PAGE..... 240

Figure 10.2 Expression and purification of N1 NA using a nickel-NTA column analyzed by Western blot using an anti-His-tag antibody. 242

Figure 11.1 A previously published example of using proton NMR to detect neuraminidase substrate cleavage using sialyllactose, a trisaccharide (301). 247

Figure 11.2 1D proton NMR neuraminidase cleavage assay performed with commercially available neuraminidase (NA) from New England Biolabs. 249

Figure 11.3 1D proton NMR neuraminidase cleavage assay performed with the substrate α-1 acid glycoprotein and two different neuraminidases (NA). 251

LIST OF THIRD PARTY COPYRIGHTED MATERIAL

Figure 1.1 Adapted from Nature Reviews Drug Discovery, volume 6, issue 12, pages 967-974. Mark von Itzstein. The war against influenza: discovery and development of sialidase inhibitors. Copyright 2007, with permission from Nature Publishing Group (License Number: 3831240125345).

Figure 1.2 Adapted from Nature Medicine, volume 19, issue 7, pages 837-849. Troels K H Scheel, Charles M Rice. Understanding the hepatitis C virus life cycle paves the way for highly effective therapies. Copyright 2013, with permission from Nature Publishing Group (License Number: 3831250010380).

Figure 1.3 Adapted from Elsevier Chemistry & Biology, volume 11, issue 10, pages 1333-1338. Nancy M King, Moses Prabu-Jeyabalan, Ellen A Nalivaika, Celia A Schiffer. Combating susceptibility to drug resistance: Lessons from HIV-1 protease. Copyright 2004, with permission from Elsevier (License Number: 3831260448924). Also adapted from Viruses, volume 2, issue 11, pages 2509-2535. Akbar Ali, Rajintha M. Bandaranayake, Yufeng Cai, Nancy M. King, Madhavi Kolli, Seema Mittal, Jennifer F. Murzycki, Madhavi N.L. Nalam, Ellen A. Nalivaika, Aysegul Ozen, Moses M. Prabu-Jeyabalan, Kelly Thayer, and Celia A. Schiffer. Molecular Basis for Drug Resistance in HIV-1 Protease. Open Access

Figure 6.4 Adapted from Nature Structural and Molecular Biology, volume 16, issue 3, pages 265-273. Jianhua Sui, William C. Hwang, Sandra Perez, Ge Wei, Daniel Aird, Li-mei Chen, Eugenio Santelli, Boguslaw Stec, Greg Cadwell, Maryam Ali, Hongquan Wan, Akikazu Murakami, Anuradha Yammanuru, Thomas Han, Nancy J. Cox, Laurie A. Bankston, Ruben O. Donis, Robert C. Liddington, Wayne A. Marasco. Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. Copyright 2009, with permission from Nature Publishing Group (License Number: 3871491183988).

Figure 6.5 Adapted from Nature Structural and Molecular Biology, volume 16, issue 3, pages 265-273. Jianhua Sui, William C. Hwang, Sandra Perez, Ge Wei, Daniel Aird, Li-mei Chen, Eugenio Santelli, Boguslaw Stec, Greg Cadwell, Maryam Ali, Hongquan Wan, Akikazu Murakami, Anuradha Yammanuru, Thomas Han, Nancy J. Cox, Laurie A. Bankston, Ruben O. Donis, Robert C. Liddington, Wayne A. Marasco. Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. Copyright 2009, with permission from Nature Publishing Group (License Number: 3871491183988).

Figure 11.1 Adapted from Nature Communications, volume 4, article number 1491. Cristopher J. Vavricka, Yue Liu, Hiromasa Kiyota, Nongluk Sriwilaijaroen, Jianxun Qi, Kosuke Tanaka, Yan Wu, Qing Li, Yan Li, Jinghua Yan, Yasuo Suzuki, George F. Gao. Influenza neuraminidase operates via a nucleophilic mechanism and can be targeted by covalent inhibitors. Copyright 2013, with permission from Nature Publishing Group (License Number: 3879461354092).

PREFACE

Chapter II is a collaborative study that is in preparation for publication as:

Kristina Prachanronarong, L. Safak Yilmaz, Aysegul Ozen, Kelly Thayer, Konstantin B. Zeldovich, Daniel N. Bolon, Timothy F. Kowalik, Jeffrey D. Jensen, Robert W. Finberg, Jennifer P. Wang, Nese Kurt-Yilmaz, Celia A. Schiffer. (2016) The substrate envelope hypothesis describes differential patterns of drug resistance in N1 and N2 neuraminidase. [*in preparation*]

Author contributions: KP and CAS conceived and designed the experiments. *KP* performed the research. *KP*, NKY, and CAS interpreted the data and wrote the manuscript.

Contributions from Kristina Prachanronarong: I designed the experiments for this study with the guidance of my mentor Celia Schiffer and my thesis research committee. I performed all the research required for the study, including modeling, molecular dynamics simulations, computational script writing/editing, and all calculations, and I created all figures and tables. Nese Kurt-Yilmaz and Celia Schiffer guided interpretation of the data. I wrote the manuscript, and Nese Kurt-Yilmaz and Celia Schiffer provided editorial assistance. All remaining authors provided valuable advice.

Acknowledgements: We acknowledge the contributions of all members of the ALiVE (Algorithms to Limit Viral Epidemics) and AiRE (Anticipating Influenza Resistance Evolution) working group.

Chapter III is a collaborative study that is in preparation for publication as:

Aysegul Ozen*, *Kristina Prachanronarong*,* Nese Kurt-Yilmaz, Ashley Matthew, Djade Soumana, Celia A Schiffer. (2016) Comparative analysis of hepatitis C NS3/4A inhibitor structures. [*in preparation*]

Contributions from Kristina Prachanronarong: Aysegul Ozen, Celia Schiffer, and I conceived this study and planned the calculations and research that would be included in the study. I performed the substrate envelope Vin/Vout and van der Waals calculations, and Aysegul performed the hydrogen bond calculations. I completed all the modeling and calculations for Figure 3.8. Aysegul and I made the figures. Nese Kurt-Yilmaz and Celia Schiffer guided interpretation of the data, and I wrote the manuscript. Nese Kurt-Yilmaz and Celia Schiffer provided editorial assistance. All remaining authors provided valuable advice.

Author contributions: AO* and *KP** contributed equally to this work. AO, *KP*, and CAS conceived and designed the research. AO and *KP* performed the research. *KP*, NKY, and CAS interpreted the data and wrote the manuscript.

Chapter IV is a collaborative study that is in preparation for publication as:

Kristina Prachanronarong, Ping Liu, Konstantin B. Zeldovich, Yu-Ping Poh, Nicholas Renzette, Timothy F. Kowalik, Daniel N. Bolon, Jeffrey D. Jensen, Thomas Han, Wayne A. Marasco, Robert W. Finberg, Celia A. Schiffer, and Jennifer P. Wang. (2016) Identification of influenza A virus candidate resistance mutations to broadly neutralizing antibody. [*in preparation*]

Author contributions: KBZ, NR, TFK, DNB, JDJ, WAM, RWF, CAS, and JPW designed the research. *KP*, PL, KBZ, YP, and NR performed the research. TH and WAM contributed new reagents/analytic tools; *KP* and JPW wrote the paper.

Contributions from Kristina Prachanronarong: I mapped the locations of candidate resistance mutations on the structures of hemagglutinin and neuraminidase and provided hypotheses about the potential structural mechanisms of these mutations. I also wrote the portions of the manuscript in the methods, results, and discussion sections that describe the structural analyses related to this study.

Acknowledgements: We thank Melanie Trombly and Nese Kurt-Yilmaz for assistance with manuscript preparation. We acknowledge the contributions of all

members of the ALiVE (Algorithms to Limit Viral Epidemics) and AiRE (Anticipating Influenza Resistance Evolution) working group.

Chapter V is a collaborative study that is in preparation for publication as:

Yuval Avnir*, Zhen Zhang*, *Kristina Prachanronarong**, Kyle Gellatly, Brendan Hilbert, Djade Soumana, Markus Bohn, Daniel R. Caffrey, Konstantin B. Zeldovich, Timothy F. Kowalik, Jeffrey D. Jensen, Jennifer P. Wang, Robert W. Finberg, Celia A Schiffer[#], and Wayne A Marasco[#]. (2016) Structural basis of an influenza hemagglutinin stem-directed antibody retaining the G6 idiotype. [*in preparation*]

Author contributions: YA*, ZZ*, and *KP** contributed equally to this work. YA, ZZ, *KP*, CAS, and WAM conceived and designed the experiments. YA, ZZ, and *KP* performed the research. KG improved the refinement of the D80 crystal structure. BH, DS, and MB contributed crystallography-related assistance. YA, *KP*, CAS, and WAM interpreted the data and wrote the manuscript. CAS[#] and WAM[#] contributed equally to this work.

Contributions from Kristina Prachanronarong: Zhen Zhang expressed and purified the antibody fragments used in this study. Together, we performed crystallization trials and began optimizing the crystallization and cryo-protection conditions for the D80 and D80-G6 crystals. I reproduced and prepared the crystals for a trip to the synchrotron, shot the crystals, and solved and refined the crystal structures. I performed the structural analysis related to the crystal

structures. Kyle Gellatly worked on additional refinement of the D80 crystal structure to improve the model, and Brendan Hilbert, Djade Soumana, Kuan-Lin Hung, and Markus Bohn provided crystallography-related assistance. I wrote the portions of the manuscript about the structural work, and Yuval Avnir wrote the portions of the manuscript about the phage display and mutagenesis experiments.

Acknowledgements: We acknowledge the contributions of all members of the ALiVe (Algorithms to Limit Viral Epidemics) and AiRE (Anticipating Influenza Resistance Evolution) working group. We thank Nese Kurt-Yilmaz for editorial assistance with preparation of the manuscript and Bill Royer for technical assistance.

CHAPTER I

1 INTRODUCTION

1.1 Therapeutic Strategies Against Viruses

Vaccines and antiviral drugs are two major classes of agents for fighting viruses. Vaccines have eradicated smallpox and have also significantly decreased rates of measles, mumps, rubella, and polio infections (1-3). However, vaccines still need to be developed for many disease-causing viruses, such as human immunodeficiency virus (HIV) and hepatitis C virus (HCV). In influenza, seasonal vaccines are available, but a universal vaccine has not yet been developed and antiviral drugs are still needed for combating infection (4, 5). For example, oseltamivir, a direct acting influenza neuraminidase inhibitor, was the most prescribed antiviral drug and 34th most prescribed drug overall in 2014 (6). However, many antiviral drugs have flaws, such as low efficacy, toxicity, side effect profiles, and high susceptibility to resistance. Viruses use host cell mechanisms to replicate and many viruses mutate rapidly and have natural variation, so designing effective antiviral drugs and vaccines is challenging. Nevertheless, there are still many opportunities to develop and improve therapeutic strategies against viruses.

In contrast to antibiotics, there are not as many broad-spectrum antiviral drugs to fight infection, but there are several specific drugs that target individual viruses, including influenza, herpesviruses, HBV, HCV, and HIV. In general,

antiviral drugs combat viral infection by targeting a specific stage of the viral life cycle to inhibit viral replication as is the case for direct acting antivirals (DAAs) or by improving the immune response to infection (7). For example, in influenza, antiviral drugs target several stages of the viral life cycle (**Figure 1.1**). Amantadine inhibits fusion by blocking the M2 ion channel and preventing the acidification of the endosome, which is required for the conformational changes in hemagglutinin that are necessary for viral and endosomal membrane fusion (8). Ribavirin interferes with replication of the viral genome, and oseltamivir and zanamivir inhibit influenza neuraminidase, which is a sialidase on the surface of the influenza virion that facilitates viral budding (9). Ribavirin is also used to treat viral hemorrhagic fevers, chronic hepatitis C virus infection, and severe acute respiratory syndrome (SARS) (10). In addition, seasonal vaccines prevent influenza binding to a cell by eliciting the development of antibodies that target the head region of influenza hemagglutinin. Recently, broadly neutralizing antibodies (BnAbs) against influenza hemagglutinin and neuraminidase have been discovered that also interfere with the viral life cycle by neutralizing these proteins in locations that are important for function, such as sialic acid binding, membrane fusion, and viral release from an infected cell (11-13). Many BnAbs target the conserved sialic acid binding site on the head of hemagglutinin or the conserved stem region that is important for fusion. In the HCV viral life cycle, there are also multiple points of intervention by antiviral drugs: viral entry, viral translation, polyprotein processing, HCV RNA replication, and viral assembly

(**Figure 1.2**). In addition, some peptides with anti-virucidal activity have been reported for HCV, but these are still in early development (14).

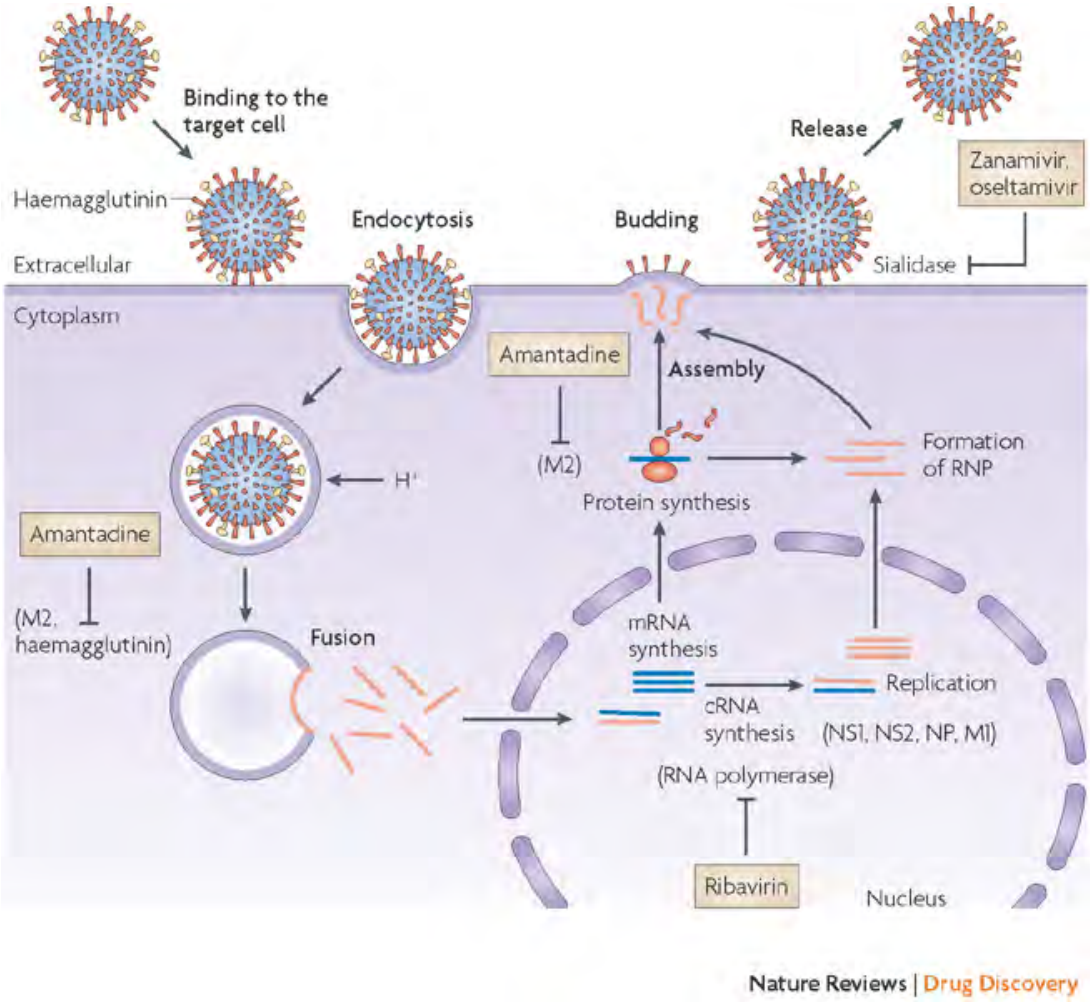
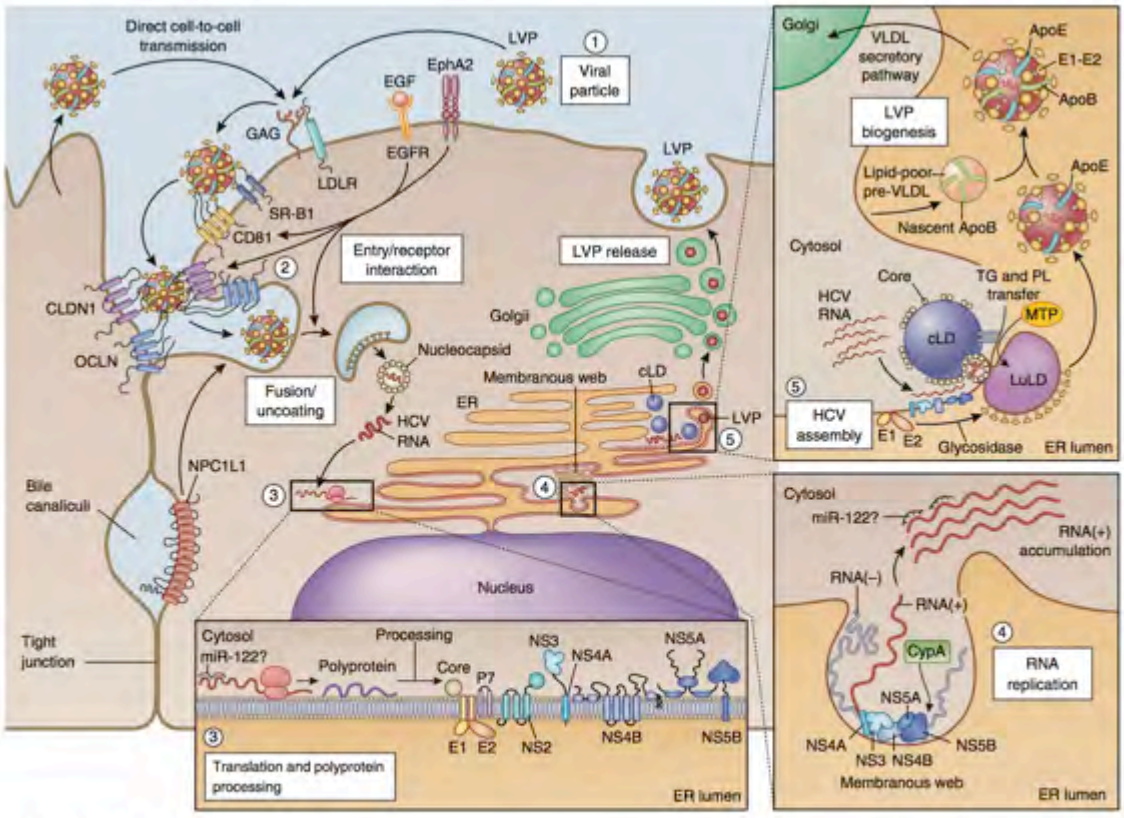


Figure 1.1 Influenza virus life cycle and targets for therapeutic strategies against influenza (9).



Points of intervention in the HCV life cycle

① The viral particle (neutralizing antibodies, virocidal peptides)

② Entry and receptor interaction (antibodies and small molecules targeting receptors, kinase inhibitors)

③ Translation and polyprotein processing (NS3-NS4A protease inhibitors)

④ HCV RNA replication (NS5B polymerase and NS5A inhibitors, miR-122 antagonists, cyclophilin inhibitors, statins, PI4KIIIα inhibitors)

⑤ Assembly and virion morphogenesis (NS5A inhibitors, DGAT1 inhibitors, glycosidase inhibitors, MTP inhibitors)

Figure 1.2 Hepatitis C virus life cycle and points of intervention for therapeutic strategies (14).

Antiviral drugs can also be broad spectrum and non-specific. One well-known class of these types of antiviral drugs are interferons, which were a main treatment for HCV until recently and are also a treatment for chronic active HBV infection (15-17). Antiviral drugs have host or viral targets, and drugs that target host mechanisms may be broad spectrum, but they also have a greater likelihood of toxicity and side effects. For instance, interferons are cytokines with broad antiviral activity that activate the immune system through various pathways and also activate other immune cells, including natural killer cells and macrophages (18). As a treatment, however, interferon has many undesirable side effects and is administered intravenously, increasing patient burden (19). Antiviral DAAs that inhibit specific viral targets, in contrast, are in general less toxic and more selective. For instance, influenza neuraminidase inhibitors have a reasonable side-effect profile, and long-term administration for prophylaxis is well tolerated in adults and children (20).

Antiviral drugs can be designed as small molecules or larger biological medical products (biologics) that are active against viruses through a variety of mechanisms. Examples of biologics include interferon, immunoglobulin therapy, vaccines, and monoclonal antibody therapy. The mechanisms of action for antiviral drugs include competitive and non-competitive antagonism and agonism. For example, Toll-like receptor agonists are being studied for their potential antiviral applications (21). Non-nucleoside reverse-transcriptase inhibitors (NNRTI) are an example of a class of non-competitive allosteric

antagonist antiviral drugs (22). There are also several examples of direct-acting antivirals that act as competitive antagonists, such as influenza neuraminidase inhibitors and HIV and HCV protease inhibitors.

1.2 Structure Based Drug Design in Antivirals

Structure based drug design is an important tool for improving the activity, selectivity, and pharmacokinetic profile of antiviral drugs. For instance, to increase activity and selectivity, HIV and HCV protease inhibitors were designed through extensive structure activity relationship (SAR) studies as substrate peptidomimetics with modifications to reduce peptide character and improve bioavailability while increasing favorable active site interactions (23-28). Influenza neuraminidase inhibitors were also designed to mimic substrates while optimizing interactions in the active site (9, 29-31). In addition, various strategies are used to improve the pharmacokinetics of drugs by altering properties such as absorption, distribution, metabolism, and excretion (ADME).

Although these strategies have increased specificity and potency of drugs, losses in efficacy still occur as a result of antiviral resistance, which is a major challenge in developing antiviral drugs (8, 32). Often, antiviral drug targets are not evolutionarily conserved, and these targets can acquire mutations that result in a loss of antiviral drug efficacy while still maintaining function. Understanding the structural basis of interactions between antiviral drugs and their targets is

critical for improving potency and specificity while also avoiding a loss of efficacy due to resistance.

1.3 Drug Resistance and Antibody Neutralization Escape in Antivirals

Drug resistance and antibody neutralization escape occur when drugs and antibodies lose the ability to treat or prevent infection through their usual mechanism of action. Drug resistance is a major health problem that impacts not only antivirals but also antimicrobial and antineoplastic drugs. Loss of efficacy from resistance can result through a variety of direct and indirect mechanisms. In the main type of direct drug resistance mechanism, primary drug resistance mutations or antibody neutralization escape mutations prevent binding by directly altering the binding site of the drug or antibody. Various indirect mechanisms of resistance have also been observed. Permissive compensatory mutations, also known as secondary resistance mutations, reduce the fitness cost of primary resistance mutations and can occur in various locations throughout a disease system. For instance, in HIV protease, primary drug resistance mutations have occurred in the protease active site and secondary mutations have developed in different locations outside of the active site (33, 34). Substrate co-evolution has also been observed where substrates mutate to become better substrates in the presence of HIV protease resistance mutations (35, 36). In influenza, hemagglutinin can also evolve to weaken binding affinity for the substrate sialic acid to compensate for reduced cleavage activity in neuraminidase (37). In

influenza neuraminidase and HIV and HCV proteases, there are also examples of drug resistant targets with multiple resistance mutations, and changes in protein dynamics can also confer resistance (34, 38-42).

Resistance mutations against antivirals have emerged as a result of selective pressures in the environment. Many viruses have a high rate of error-prone replication, creating heterogeneous populations of viruses with drug resistant variants at low frequency in the absence of selective pressure. In fact, many viruses exist as quasispecies, which is a population of viruses with mutant genomes. When these viruses are exposed to selective pressure from antiviral drugs, the fitness of drug resistant variants increases, allowing them to dominate in the population, while the growth of drug susceptible variants is inhibited. This process also occurs when antiviral drug doses are sufficiently low (suboptimal) and administered over long periods of time, encouraging the development of drug resistant variants. In addition, reservoirs of virus also exist in other species besides humans for some viruses, such as in birds, pigs, horses, and sea mammals for influenza, and these reservoirs allow new viral variants in humans to emerge periodically (12). If antiviral drug targets are not evolutionarily constrained and can accommodate resistance mutations while still maintaining function, resistance mutations can develop.

1.3.1 Resistance to Competitive Small Molecule Antivirals and the Substrate Envelope

Drug resistance is a change in molecular recognition such that drug resistant variants no longer bind inhibitors while maintaining biological function. Understanding the salient molecular features of a drug target necessary for maintaining function is critical for designing antiviral drugs that are less susceptible to drug resistance. Through our work studying HIV and HCV proteases, we established the *substrate envelope* hypothesis, which predicts that substrates fill a conserved enzyme-specific volume when bound to the active site even though substrates vary in amino acid sequence, and primary drug resistance mutations occur where inhibitors protrude beyond the substrate envelope (**Figure 1.3**) (43, 44). The substrate envelope has been established for HIV and HCV proteases, but should also be applicable to understanding primary drug resistance mutations in other enzymes that process substrates. In fact, the substrate envelope hypothesis has been applied prospectively to a set of drug targets including Abl kinase, thymidylate synthase, chitinase, dihydrofolate reductase, and influenza neuraminidase, and the protrusion of inhibitors outside of the substrate envelope correlates with different patterns of drug resistance (45). In addition, the substrate envelope has been used in the development of a reverse transcriptase inhibitor, tenofovir (46, 47). Recently, crystal structures of inactive influenza NA in complex with uncleaved substrates were reported, and the substrate envelope hypothesis should be applicable to understanding drug resistance in NA in different subtypes using these crystal structures (48). In HCV,

there are also several protease inhibitors in clinical development, and comparing how these inhibitors fit within the substrate envelope may also help elucidate how different patterns of drug resistance have emerged for these inhibitors.

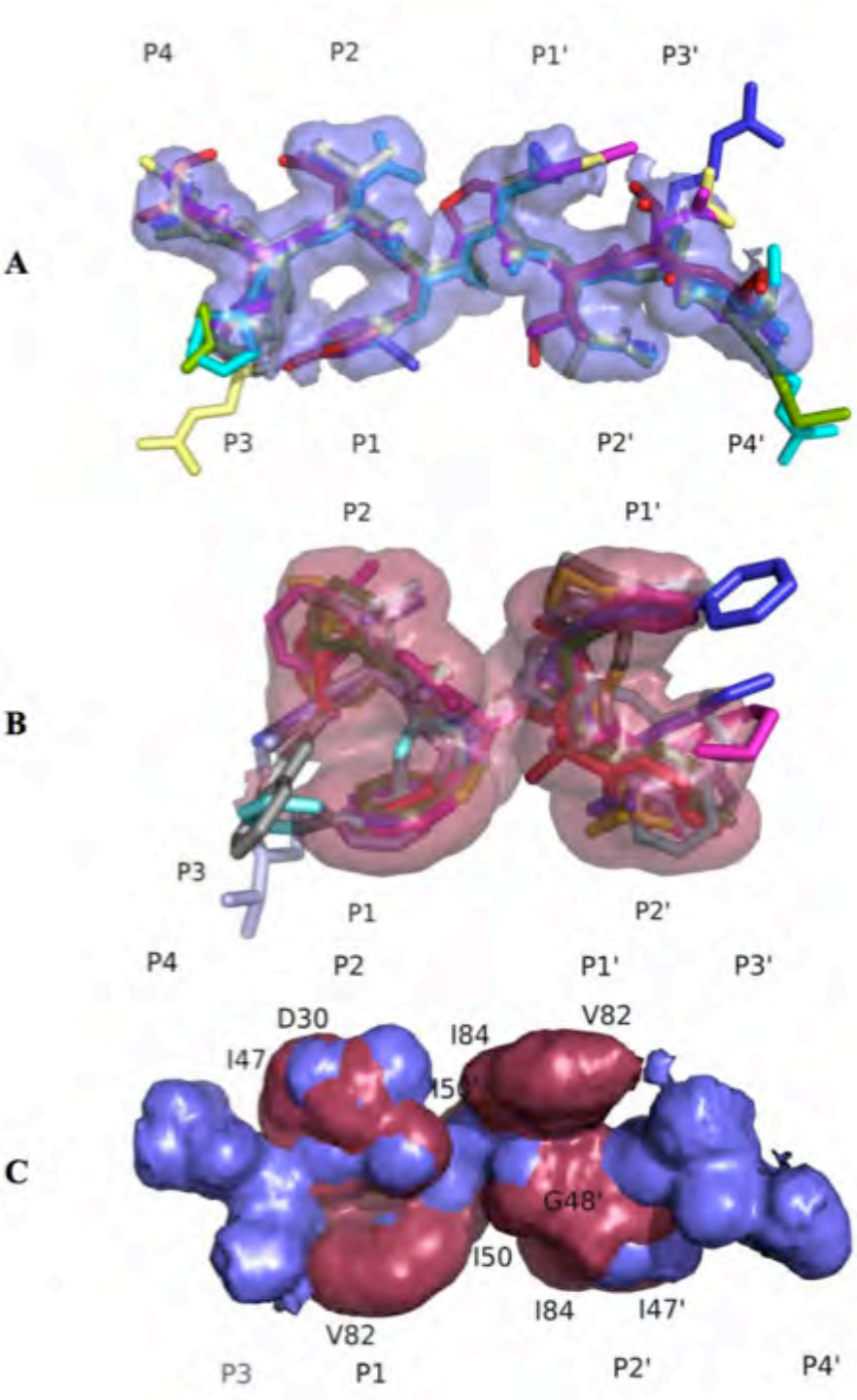


Figure 1.3 Substrate and inhibitor envelopes in HIV protease (33, 49).

Figure 1.3 Substrate and inhibitor envelopes in HIV protease (33, 49). A) The overlapping van der Waals volumes of substrates fill a conserved shape known as the substrate envelope, shown in blue. Each substrate moiety is labeled. B) The overlapping van der Waals volumes of inhibitors fill a conserved shape known as the inhibitor envelope, shown in red. C) Superposition of the inhibitor and substrate envelopes shows that where inhibitors protrude outside of the substrate envelope is where primary drug resistance mutations occur.

1.3.2 Resistance to Antibodies as Therapeutics: Antibody Neutralization Escape

Neutralizing antibodies are also an important element of an immune response to viruses, and antibody neutralization escape resistance mutations occur when antibodies that were once capable of neutralizing antigens are no longer effective. Antibodies can develop over the course of an infection naturally without treatment, and vaccines can also actively stimulate the immune system to develop antibodies that protect against viral infection. Antibodies can also be administered as passive immunotherapy before or during infection.

There are three main types of viral vaccines: live attenuated, inactivated, and recombinant vaccines (50). In general, these antiviral vaccines are effective for long periods of time but influenza vaccines are a major exception. Seasonal influenza vaccines are very effective against specific strains, but they lose efficacy regularly against new dominating circulating strains as a result of viral evolution and variation, and new vaccines must continually be developed (12). Therefore influenza vaccines should be improved so that they provide broader and longer acting protection against more influenza strains.

Seasonal influenza vaccines become less effective over time due to several factors. Current influenza vaccines mount an immune response against the highly variable immunodominant globular head region of influenza hemagglutinin (HA), and since this region is highly variable, vaccines become a mismatch to circulating strains as this region evolves. The influenza viral RNA-dependent RNA polymerase is highly error prone and does not contain a

proofreading mechanism (51, 52). Consequently, new mutations accumulate rapidly over time in evolutionarily unconstrained regions of the genome, such as the globular head region of HA, in a process known as antigenic shift (53). In addition, the genome of the influenza virus is segmented, so that if an organism or a cell is infected by at least two different influenza viruses simultaneously, the segmented strands of the genome can reassort and create new viruses with strands of RNA from each parental virus, in a process known as antigenic shift. These changes also contribute to influenza virus diversity.

Reservoirs of influenza viruses exist in different species, and viruses from these sources periodically emerge in humans. These viruses can mutate to sustain human-to-human transmission, which occurred with the 2009 swine flu (54). When influenza strains emerge to which a population lacks immunity, epidemics and pandemics can occur. However, vaccines that can be designed to mount an immune response to antigens that are highly conserved between multiple strains of influenza may provide broader and longer acting protection.

In addition to vaccines, there are nearly a dozen specific monoclonal antibody immunotherapies against different viruses (55, 56). However, antibody neutralization escape resistance mutations have also been reported with these therapies. For instance, *in vivo* and *in vitro* resistance mutations have developed in respiratory syncytial virus (RSV) in response to palivizumab, an antiviral monoclonal antibody therapy and the first commercially available antibody approved by the FDA for use against an infectious disease (57).

Recently, cross-reactive broadly neutralizing antibodies (BnAbs) in influenza have been discovered that neutralize a wide spectrum of influenza strains, and these antibodies are currently in development as passive or active immunotherapy (12, 58-63). These antibodies target highly conserved regions on hemagglutinin, such as the stem region, which is important for the conformational changes that occur during fusion, and the sialic acid binding region on the globular head domain of HA (12). BnAbs have reactivity across subtypes, groups, and even types of influenza, and there is widespread interest in developing them into therapies (12). For instance, Diridavumab, or CR6261, is a monoclonal antibody therapy against influenza A HA based on a BnAb that is being developed by Crucell, a Johnson & Johnson subsidiary (60). However, resistance mutations have emerged with some of these BnAbs. Two antibody neutralization escape mutations, D19N and G33E, were discovered against the stem-directed BnAb CR8020 after four viral passages in the presence of CR8020 as a selective pressure (61, 64). These mutations are both located at the binding interface of CR8020 with HA. Two escape mutations were also discovered against stem-directed BnAb C179 at positions 318 in HA_1 and 52 in HA_2 (58, 65). CR6261, which also binds the stem, was initially refractory to developing resistance, but still developed the resistance mutation H111L after 10 passages (12, 66, 67). Therefore, revealing the effect of such mutations on antibody–target interaction is needed to design BnAbs that are less susceptible to resistance.

1.4 Protein Dynamics in Molecular Recognition

Proteins undergo conformational changes on various time scales, which are essential in molecular recognition and biological function. Therefore, understanding protein dynamics is critical for understanding biological functions of proteins, including target binding and substrate turnover. Several experimental and computational methods are available for studying protein dynamics, including nuclear magnetic resonance spectroscopy (NMR), hydrogen-deuterium exchange mass spectrometry, small angle X-ray scattering (SAXS), spin label electron paramagnetic resonance (EPR) spectroscopy, fluorescence spectroscopy, and time-resolved x-ray crystallography. In addition, molecular dynamics (MD) simulations are a computational method for studying protein dynamics with detailed atomic resolution. Some important considerations for deciding which experimental and computational methods to use for studying protein dynamics include the time scale of the motion, the size and properties of the biological system being studied, and the specific limitations of each method. However, recent technological advances have greatly improved many of these methods. For instance, advances in x-ray sources, data collection, and data analysis methods have also increased the accessibility of time-resolved x-ray crystallography (68). The speed with which MD simulations can be calculated today has greatly increased based on the availability of supercomputing clusters and increased parallelization of calculations using powerful graphics processing unit (GPU) technology. Thus, MD simulations can be run on much longer time

scales, and they are very valuable as a complementary method to experiments in revealing the underlying molecular details of experimentally observed data on protein dynamics.

1.4.1 Role of Protein Dynamics in Antiviral Resistance

Protein dynamics can contribute to antiviral drug resistance (42, 69-71). For example, secondary permissive mutations can improve enzyme fitness in the presence of primary drug resistance mutations by altering protein dynamics to enhance biological function. Differences in the conformation of loops and other structures in an enzyme active site can affect catalytic activity. For example, in certain subtypes of influenza neuraminidase, the 150-loop in the active site closes down on ligands after binding (72, 73). This loop contains a catalytic residue D151, and therefore, the motion of this loop may impact substrate processing. In HIV protease, the flaps close down on the substrate, which requires an extensive rearrangement of hydrophobic residues in the core of the enzyme, known as hydrophobic sliding, and changes in these protein dynamics can alter the balance between inhibitor binding and substrate processing (69). Mutations outside of the active site in HIV protease can change the dynamic ensemble of the protease, and the effect of such distal mutations are propagated to critical active site residues, suggesting that a network of interactions is involved in how distal drug resistance mutations impact the active site (34). Furthermore, in the HIV gp41/gp120 envelope fusion protein, gp120 variants that

have faster fusion kinetics are more resistant to the fusion inhibitor enfuvirtide (74, 75). Overall, changes in protein dynamics can alter the balance between biological function versus inhibitor binding, and this balance can be tipped in favor of biological function while limiting inhibitor binding and conferring drug resistance.

1.4.2 Role of Protein Dynamics in Antibody Neutralization Escape

Alterations in protein dynamics can also be involved in antibody neutralization escape. With the HIV gp41/gp120 envelope protein, antibody potency is related to the ability of the antibody to bind to the pre-fusion closed form of the HIV envelope protein, and therefore changes in dynamics can impact the binding affinity of antibodies. In addition, conformational diversity can inhibit antibody-mediated neutralization (76, 77). In human T cell leukemia transmembrane protein, which is a class I fusion protein like HIV gp41/gp120 envelope and influenza HA, binding of antibodies to the coiled-coil pre-hairpin fusion intermediate is highly conformation dependent, and this mechanism appears to be applicable to other retroviruses, providing additional examples of how protein dynamics can impact antibody binding to its target (78).

1.5 Scope of Thesis

In this thesis, I attempt to fill gaps in our understanding of antiviral drug resistance in influenza NA and HCV NS3/4A protease, and broadly neutralizing antibodies against influenza HA. First, I use molecular dynamics simulations to

define the *dynamic substrate envelope* in influenza neuraminidase to understand how different patterns of drug resistance have emerged in N1 and N2 NA subtypes. I demonstrate that differences in van der Waals contact potential energies and hydrogen bonding with the enzyme between substrates and inhibitors in the two subtypes can explain the different patterns of drug resistance for mutations that are in the active site. Although highly homologous, the differences between N1 and N2 NA are sufficient to define distinct substrate envelopes in these subtypes, leading to selection of different drug resistance mutations. Then, I use a similar approach for HCV NS3/4A protease, and use the substrate envelope hypothesis to understand different patterns of drug resistance that have emerged with various inhibitors in clinical development. Next, I examine potential antibody neutralization escape mutations that emerged in influenza HA during viral passaging in the presence of the broadly neutralizing antibody F10 as a selective pressure. The antibody neutralization escape mutations in HA occur outside of the F10 binding epitope and may impact the conformational changes in HA that are important for fusion. Finally, I solved novel crystal structures of an influenza broadly neutralizing antibody fragment D80 alone and in complex with an anti-idiotypic antibody G6. I show that the most important interactions for binding to D80 occur in the CDR2 and CDR3 loops of the heavy chain of D80, and an extensive pi stacking interaction is conserved between the CDR loops in the heavy chain of the bound and unbound structures of D80, suggesting a pre-arranged binding interface.

CHAPTER II

2 THE SUBSTRATE ENVELOPE HYPOTHESIS DESCRIBES DIFFERENTIAL PATTERNS OF DRUG RESISTANCE IN N1 AND N2 NEURAMINIDASE

2.1 Abstract

Currently, the main antiviral drugs against influenza are neuraminidase (NA) inhibitors, but drug resistance is widespread and highly prevalent, such as during the 2007-2008 influenza season. Two subtypes of NA predominate in humans, N1 and N2, but different patterns of drug resistance have emerged in each subtype despite highly homologous active sites. To understand the molecular basis for the alternative patterns of drug resistance, structural and dynamic analysis on N1 and N2 in complex with substrates and inhibitors were performed. Comparing dynamic substrate and inhibitor envelopes and interactions at the active site during MD simulations revealed how different patterns of drug resistance have emerged for specific drug resistance mutations, including residues I222, N246, and H274 in N1 and E119 in N2. This work provides guidelines for predicting and understanding mutations that decrease susceptibility to NA inhibitors in different subtypes, insights toward development of novel inhibitors to avoid drug resistance by better mimicking the dynamic binding features and molecular interactions of substrates in the active site.

2.2 Introduction

Seasonal influenza infects over 24 million people annually in the United States, causing over 200,000 hospitalizations and 40,000 deaths (79). Vaccines prevent influenza infection, but are not always effective because of antigenic drift, a high mutation rate, and mismatches between vaccine and circulating strains (51, 53). Therefore, direct acting antiviral medications are needed as another strategy for combating influenza infection.

Currently, the main antiviral drugs against influenza are neuraminidase (NA) inhibitors (5). Influenza NA is a viral sialidase necessary for viral maturation, and cleaves terminal sialic acid residues from glycoproteins to release the budding virus from the surface of infected cells (80-82). NA may also increase viral motility in mucus (83). In 1999, the U.S. Food and Drug Administration (FDA) approved two competitive active site NA inhibitors, oral oseltamivir (OST) and inhaled zanamivir (ZMR), and in 2014 peramivir for intravenous administration (5, 29-31).

Subtype specific patterns of drug resistance have developed against NA inhibitors even though sialic acid binding sites between subtypes are highly homologous (**Table 2.1**) (41, 84-93). Type A influenza is most prevalent and is divided into subtypes based on two surface proteins, hemagglutinin (HA) and NA, and two subtypes predominate in human infection, N1 and N2. In addition to drug resistance mutations observed clinically in N1 and N2, additional resistance mutations have been identified experimentally through *in vitro* and *in vivo*

experiments, and more residues have been associated with decreased NA inhibitor susceptibility (**Tables 2.2 and 2.3**)(41). Double mutations and secondary permissive mutations have also been observed(40). Between N1 and N2, the residues in the active site are 94% identical, but overall, N1 and N2 share approximately 45% amino acid sequence identity and 60% similarity. The particular viral genetic background and differences in protein dynamics between N1 and N2 may contribute to different patterns of drug resistance, but the mechanisms underlying these differences are still not well understood (87).

Table 2.1 Drug resistance mutations in NA subtypes N1 and N2 that have been observed in viruses clinically or during influenza surveillance.

Subtype N1	Subtype N2
N70S[Z]	
V116A[OZ]	
	E119V[O]/I[OZP]
Q136R[ZP]/K[ZP]	Q136K[Z]
	T148K[Z]
D151E[ZP]	
Y155H[OZP]	
D198E[O]	
I222V[O]/L[O]/R[OZP]/T[O]/K[O]	
	V215I[O]
S246N[OZ]/G[O]	
H252Y[O]	
H274Y[OP]	
	R292K[OZP]
N294S[OZP]	**N294S[O]**
	N329K[OZ]
	S331R[O]

Specific inhibitors that are susceptible to each mutation are noted in the superscript for each mutation by the letters O, Z, and P for oseltamivir, zanamivir, and peramivir, respectively. Residues in **bold** make direct van der Waals contacts with the ligands analyzed in this study.

Table 2.2 Drug resistance mutations in NA subtypes N1 and N2 that have been observed experimentally.

Subtype N1	Subtype N2
E119V^{OZP}/A^{OZ}/D^{OZP}/G^{OZP}	**E119A^{OZP}/D^{OZP}/G^{ZP}/I^{OP}**
Q136L^{OZ}	
	D151E^{OZ}
D198G^{OZ}	
I222M^{O}	**I222V^{O}**
	R224K^{OZ}
	E276D^{OZ}
	R371K^{OZ}

Specific inhibitors that are susceptible to each mutation are noted in the superscript for each mutation by the letters O, Z, and P for oseltamivir, zanamivir, and peramivir, respectively. Residues in **bold** make direct van der Waals contacts with the ligands analyzed in this study.

Table 2.3 Additional residue positions in NA associated with decreased neuraminidase inhibitor susceptibility (41).

Additional Residue Positions Associated With Decreased NAI Susceptibility
R118
L134
R152
R156
W178
S179
E227
D243
E277
D293
D330
T365
Y406
E425
I117
G209
N221
V233
D347
S404

Residues in **bold** make direct van der Waals contacts with the ligands in this study.
Residues in i*talics* are considered outside of the active site.

Drug resistance in enzyme targets is a change in molecular recognition such that drug resistant variants no longer bind inhibitors while processing substrates effectively. Understanding substrate recognition is critical because biological function constrains NA under the selective pressure of inhibitors. Interestingly, NA inhibitors were rationally designed based on the molecular features of substrates and to optimize binding interactions in the active site, however resistant variants that avoid inhibition but still process substrates have emerged (29-31, 41).

Through our work studying HIV and Hepatitis C virus (HCV) NS3/4A proteases, we established the *substrate envelope* hypothesis, which predicts that substrates fill a conserved enzyme-specific volume when bound to the active site, and primary drug resistance mutations occur where inhibitors protrude beyond the *substrate envelope* (43, 44). We also incorporated protein dynamics into a *dynamic substrate envelope* using molecular dynamics (MD) simulations to determine how conformational fluctuations of inhibitors compared to substrates contribute to drug resistance (94, 95). Differences in dynamic interaction profiles between substrates and inhibitors at the active site illustrate how primary drug resistance mutations occur at residues that interact more extensively with inhibitors than substrates.

How influenza N1 and N2 interact with substrates versus inhibitors at the molecular level and how specific mutations alter the balance between inhibitor binding versus substrate processing to lead to different patterns of drug

resistance in the two subtypes is not clear. In this study, we performed molecular dynamics simulations on representative N1 and N2 NA structures in complex with substrates and inhibitors, and calculated *dynamic substrate envelopes* for these NAs. These *dynamic substrate envelopes* reveal how different patterns of drug resistance have emerged in N1 and N2 NA for specific drug resistance mutations, including those at residues I222, N246, and H274 in N1 and E119 in N2. These data are predictive for mutations that decrease susceptibility to NA inhibitors in different subtypes, and insights from this work can guide the development of novel inhibitors to avoid drug resistance by better mimicking the dynamic binding features and molecular interactions of substrates in the active site.

2.3 Results

2.3.1 Description of Molecular Dynamics Simulations

Eight 100 ns molecular dynamics (MD) simulations on prototypic N1 and N2 tetrameric NAs in complex with substrates and inhibitors were performed (**Table 2.4**). All MD simulations were performed on the full tetramer to simulate the biological unit with the added benefit of multiple copy simulation sampling (72). The root mean squared deviation (RMSD) calculations converge and indicate that the systems are stable (**Figure 2.1**). In addition, experimental and simulation derived B-factors and alpha carbon root mean squared fluctuation (RMSF) values agree well (**Figure 2.2**).

Table 2.4 Table of NA ligand complexes that were analyzed.

Influenza NA Subtype	Ligand	Protein Starting Structure (PDB ID)	Ligand Starting Structure (PDB ID)
N1	Natural Avian Substrate	3B7E	4GZW
N1	Natural Human Substrate	3B7E	4GZX
N1	Oseltamivir	3B7E	4GZP
N1	Zanamivir	3B7E	3B7E
N2	Natural Avian Substrate	4GZW	4GZW
N2	Natural Human Substrate	4GZX	4GZX
N2	Oseltamivir	4GZP	4GZP
N2	Zanamivir	4GZW	3B7E

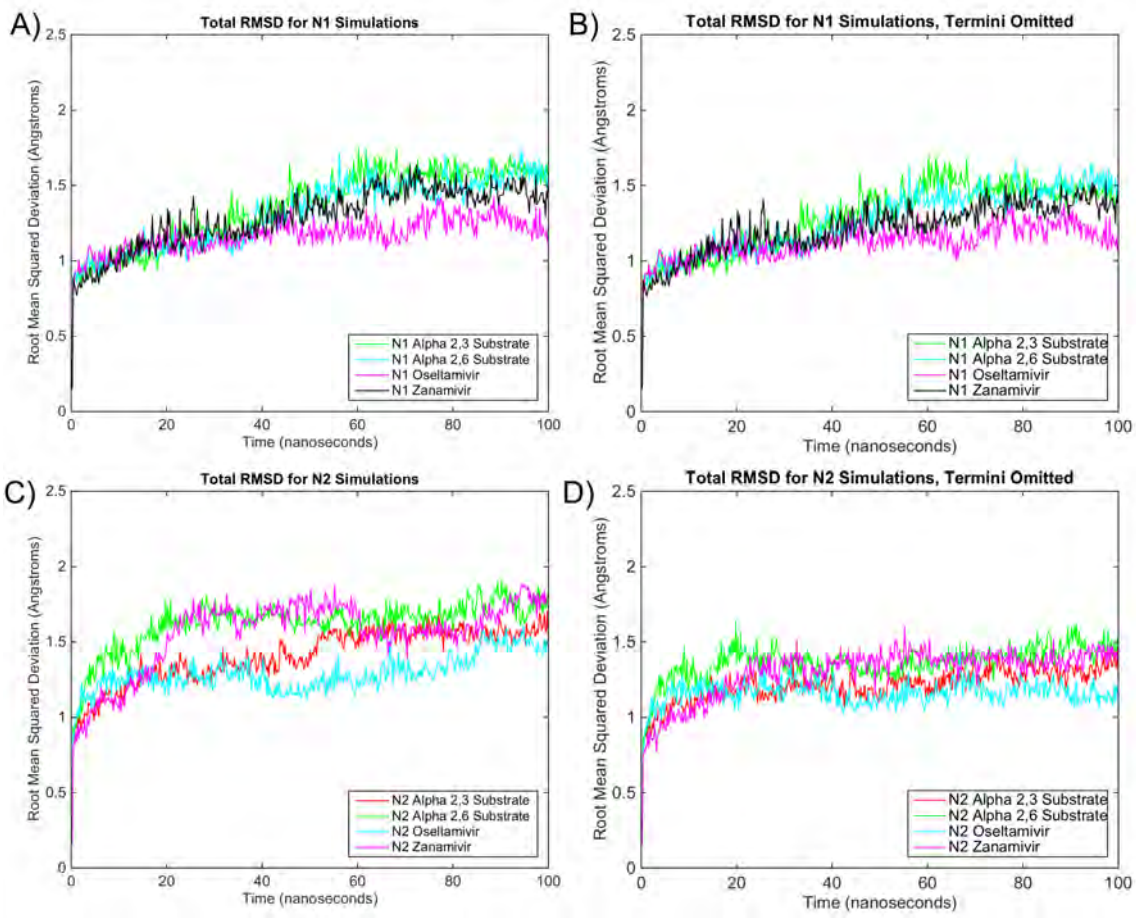


Figure 2.1 Overall root mean squared deviation (RMSD) calculations for each MD simulation were performed over 100 ns.

Figure 2.1 Overall root mean squared (RMSD) calculations for each MD simulation were performed over 100 ns. A) and C) RMSD calculations for N1 and N2 NA, respectively, for all simulations over all alpha carbons and B) and D) over all alpha carbons except five flexible N-terminal and C-terminal residues that were omitted from the calculation.

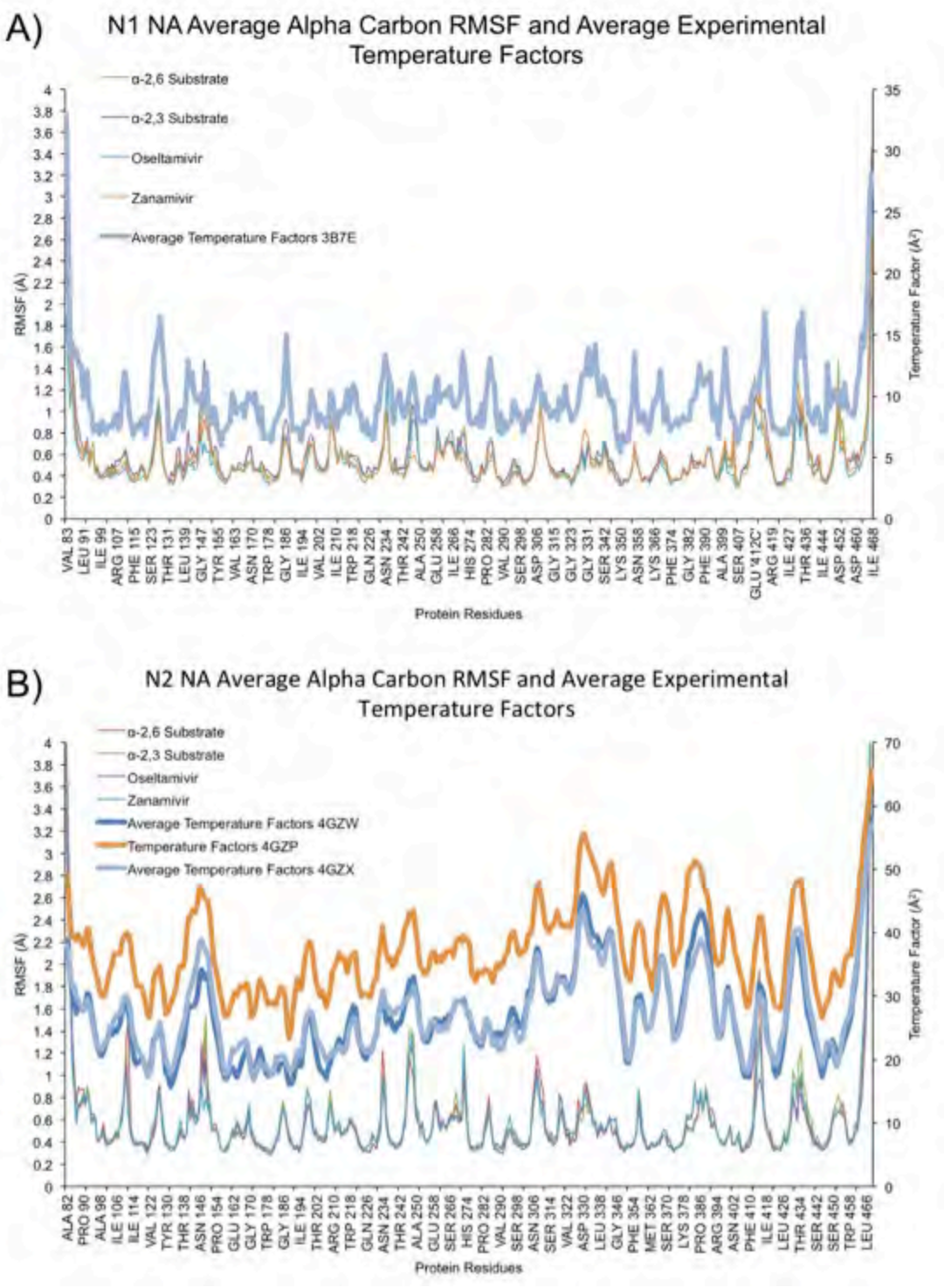


Figure 2.2 Experimental average B factors and RMSF values for simulations.

Figure 2.2 Experimental average temperature factors (B factors) and alpha carbon root mean squared fluctuation (RMSF) values for simulations. Values were compared for all simulations in N1 (A) and N2 (B) NA.

Influenza neuraminidase is a homotetrameric transmembrane protein (**Figure 2.3A**), and sites of drug resistance are located both in and outside of the active site (**Figure 2.3B).** Influenza NA also cleaves two types of substrates (**Figure 2.4A**) (48, 96). One substrate has alpha-2,3 glycosidic linkages between the terminal sialic acid and the neighboring galactose, and these substrates are present in avian gastrointestinal epithelium, human respiratory tract mucin, and human lower airway epithelium (96). The other type of substrate has alpha-2,6 glycosidic linkages, and these substrates are present in human upper airway epithelium (96). Neuraminidase inhibitors, such as oseltamivir and zanamivir are also very similar in structure to each other and the substrate cleavage product sialic acid (**Figure 2.4B**).

2.3.2 Static and Dynamic Substrate and Inhibitor Envelopes

We incorporated dynamics into a *dynamic substrate envelope* by mapping the van der Waals (vdW) volumes of substrates in the active site over time on a three-dimensional grid (**Figure 2.5**). This calculation produces a probabilistic distribution of conformers in the active site, providing more detail compared to static *substrate envelopes*. We also calculated *dynamic inhibitor envelopes* for N1 and N2 NA for both oseltamivir and zanamivir individually (**Figure 2.5**). Carbohydrate moieties galactose and N-acetyl-glucosamine that extend from sialic acid in the substrates contribute additional flexibility, which is also evident in the high root-mean-squared-fluctuation (RMSF) for these atoms (**Figure 2.6**).

The *dynamic substrate envelopes* also reflect this flexibility, where the more solvent exposed carbohydrate moieties have a broader probabilistic volume distribution (**Figure 2.5**). The movement in the galactose and N-acetylglucosamine moieties may also propagate additional flexibility to sialic acid. In contrast, the inhibitors oseltamivir and zanamivir are smaller and more rigid than the substrate (**Figure 2.5)**. However, the glycerol moiety in zanamivir better mimics the flexibility of the glycerol moiety in sialic acid (**Figure 2.6**).

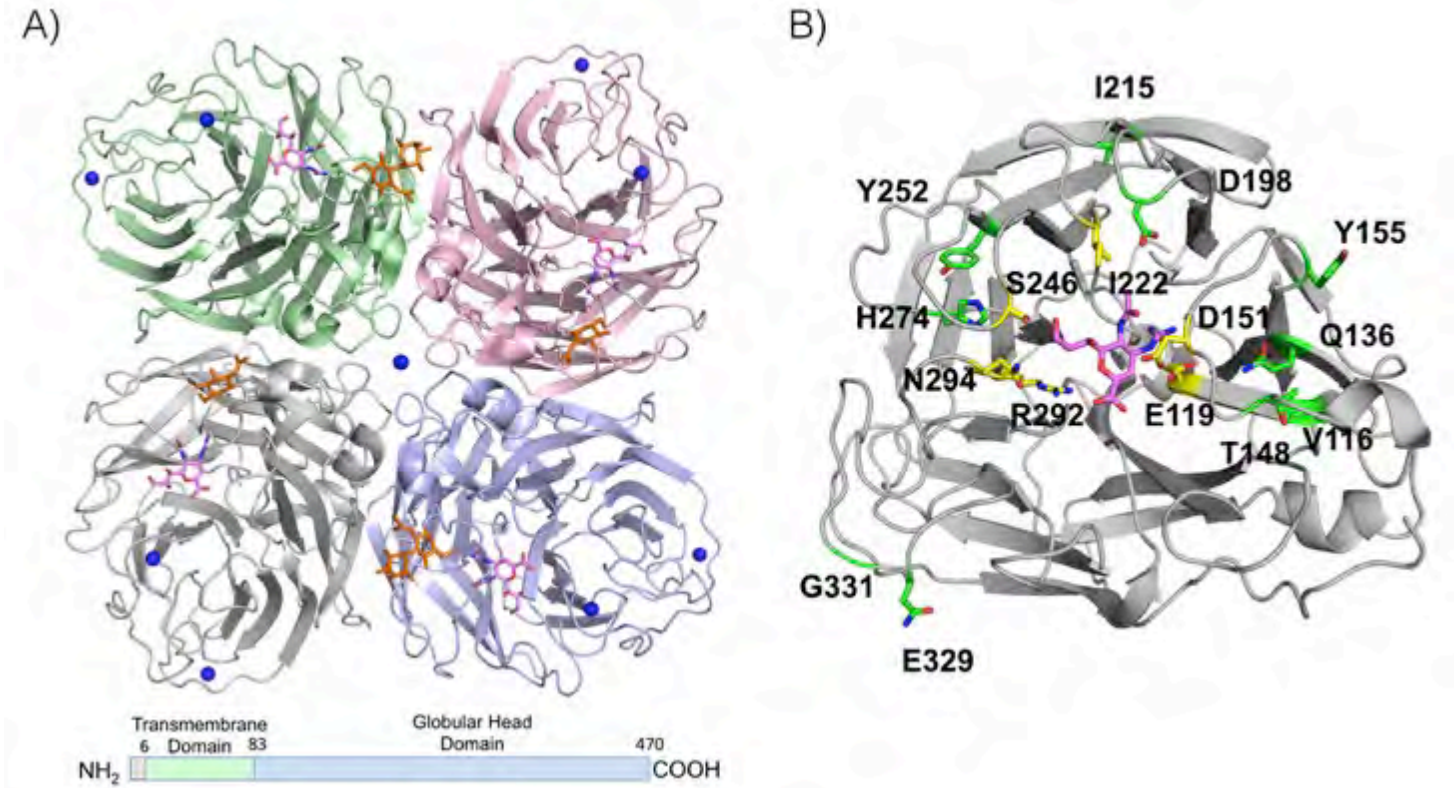


Figure 2.3 Influenza NA structure and sites of drug resistance.

Figure 2.3 Influenza NA structure and sites of drug resistance. (A) Ectodomain of an NA tetramer. Each chain is a different color, calcium ions are dark blue, and glycosylation is gold. Zanamivir is in violet sticks bound to each subunit. B) The active site in one monomer. Zanamivir is in violet sticks bound in the active site. Drug resistance residues from Table 1 that form direct van der Waals contacts with inhibitors are shown in yellow sticks, and additional drug resistance residues from Table 1 are shown in green sticks.

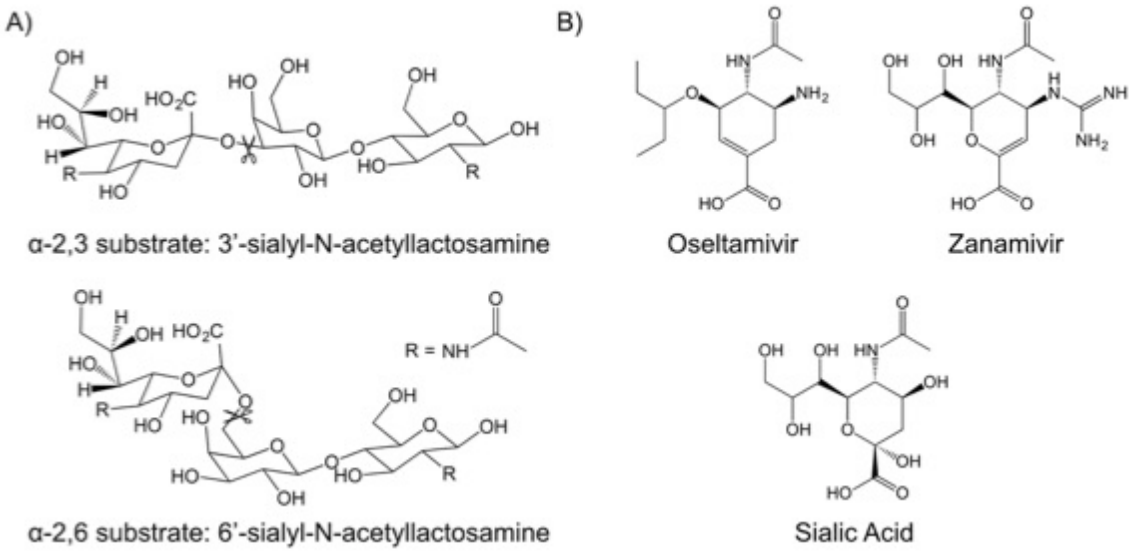


Figure 2.4 Two-dimensional structures of substrates and inhibitors of NA.

Figure 2.4 Two-dimensional structures of substrates and inhibitors of NA. A) An α-2,3 substrate, 3'-sialyl-N-acetyllactosamine, and an α-2,6 substrate, 6'-sialyl-N-acetyllactosamine, are shown. Scissors indicate the location of the scissile bond. The R group is an N-acetyl group. B) Two dimensional structures of two NA inhibitors, zanamivir and oseltamivir, and the substrate cleavage product sialic acid.

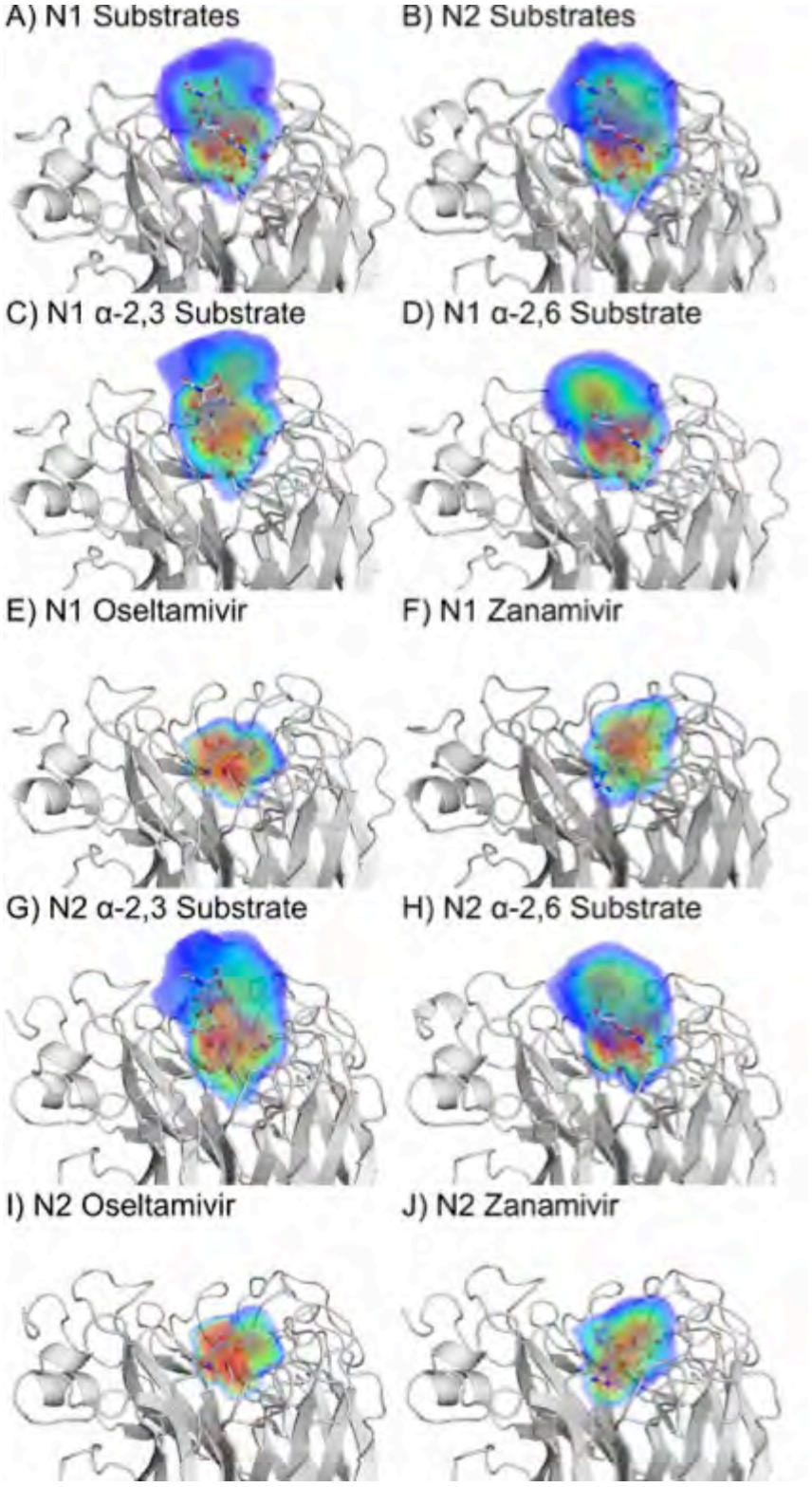


Figure 2.5 Dynamic substrate and inhibitor envelopes for NA.

Figure 2.5 Dynamic substrate and inhibitor envelopes for NA. Corresponding ligands are shown in the active site in gray sticks. (A) The combined dynamic substrate envelope of α-2,3 and α-2,6 substrates is shown in N1 NA, and in (B) N2 NA. (C) and (D) The individual substrate envelopes of α-2,3 and α-2,6 substrates are shown in N1 NA, respectively, and also in (G) and (H) for N2 NA. (E) and (F) The individual inhibitor envelopes of oseltamivir and zanamivir are shown in N1 NA, respectively, and also in (I) and (J) for N2 NA. The probabilistic volume distribution is represented using a rainbow color spectrum from red to blue to indicate more or less occupied regions, respectively.

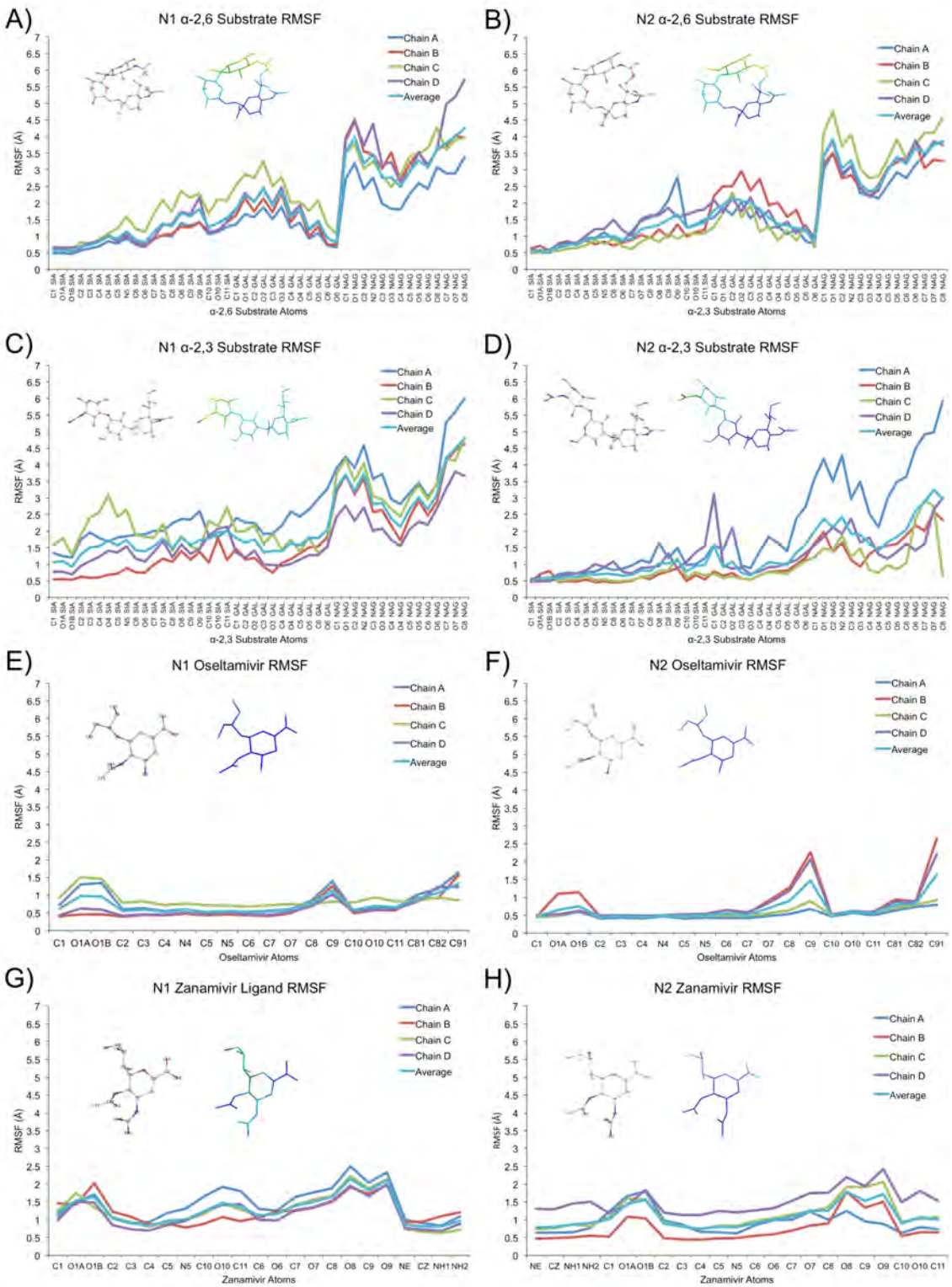


Figure 2.6 Average ligand root mean squared fluctuation (RMSF) in NA.

Figure 2.6 Average ligand root mean squared fluctuation (RMSF) in NA. The average RMSF over 400 ns is mapped onto the structures of substrate and inhibitor atoms for each system.

2.3.3 Differences in Van der Waals Contacts Correlate with Differential Patterns of Drug Resistance in N1 and N2

The average vdW contact energies were calculated over the course of MD simulations and mapped onto the active site for each system (**Figure 2.7**). Comparing the vdW energies between substrate versus inhibitor binding reveals what residues are more critical for inhibitor binding than substrate recognition. Specific drug resistance mutations occur at these sites as they are less determinant in substrate binding, such as E119 in N2 and I222 and S246 in N1.

Residue E119 is primarily a drug resistance residue in N2 NA, which can be explained by differences in van der Waals interactions and hydrogen bonds in N1 versus N2 NA (**Figure 2.8A and 2.9**). E119 interacts with both substrate and inhibitor to a similar extent in N1, and therefore, a mutation to decrease inhibitor's contacts at this residue would also impact substrate processing. Resistance mutations at this residue in N1 has not yet been observed clinically or though influenza surveillance (41), although mutations at 119 have been observed experimentally *in vitro* through reverse genetics and other *in vitro* mutagenesis studies. E119 is interacts with the C4-guanidinium group on zanamivir and peramivir and with the corresponding amino group on oseltamivir. Since E119 makes more extensive contact in substrate binding in N1, mutations at this site confer strong fitness penalties that cannot be overcome even in the presence of inhibitor. Experimental results support the *in silico* observations that E119 is critical for substrate binding and fitness in N1, and therefore, E119 is not a prevalent drug resistance site in N1 (97).

In contrast in N2, both oseltamivir and zanamivir have greater vdW contacts than substrates, especially the α-2,6 substrate, indicating that a drug resistance mutation would be well tolerated at this location in N2 NA. Drug resistance mutations in N2 E119V and E119I likely cause a loss of inhibitor contacts. Hydrogen bonds with inhibitors are also more prevalent at E119 compared to substrates (**Figure 2.9**) in N2 making E119 an optimal site for drug resistance in N2 NA.

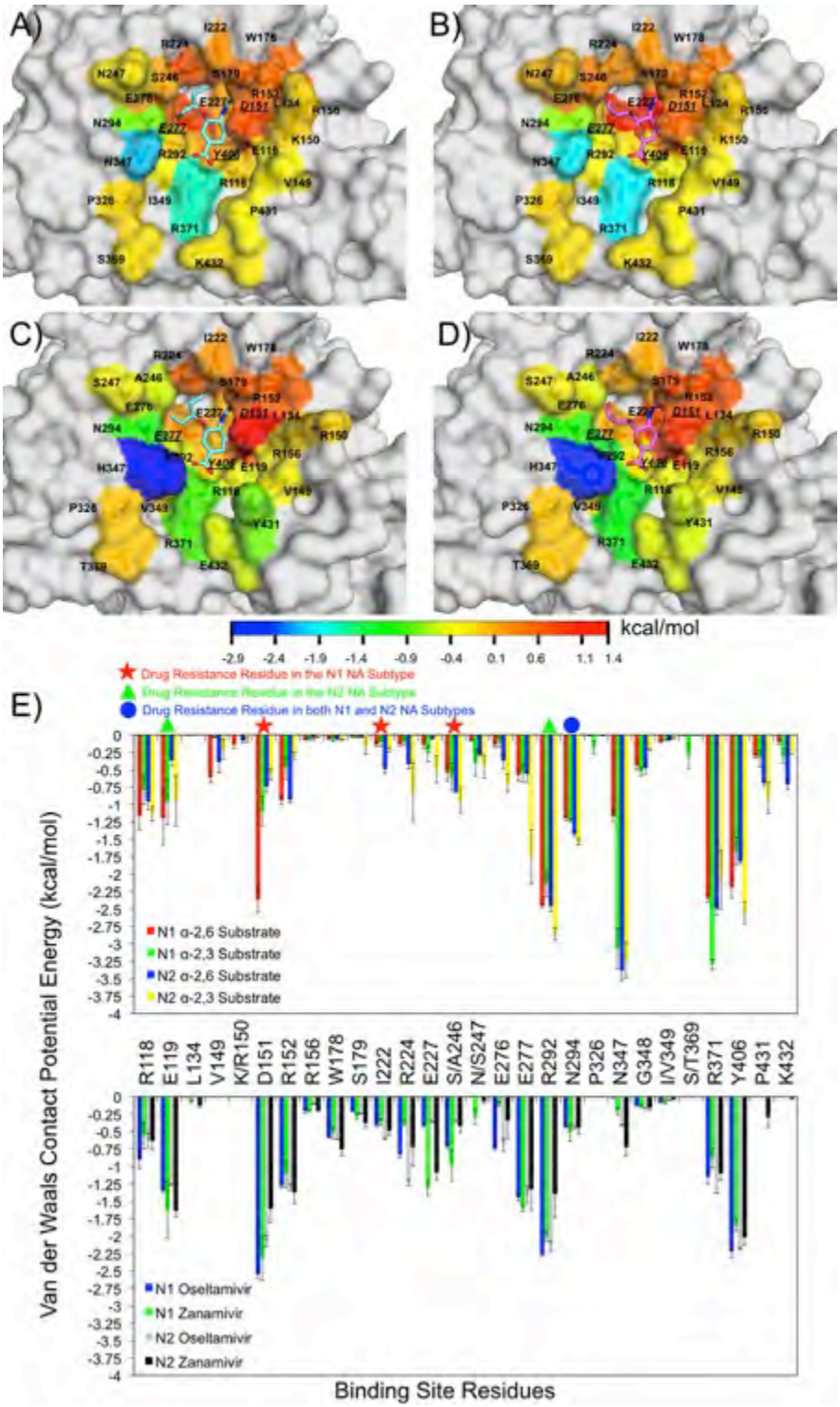


Figure 2.7 Van der Waals interactions in NA.

Figure 2.7 Van der Waals interactions in NA. In A) through D), NA residues that contact inhibitors are colored on the surface of NA according to differences in average van der Waals contact potential energies during MD simulations between substrates and the inhibitor. This calculation was performed for A) N1 NA in complex with oseltamivir, B) N1 NA in complex with zanamivir, C) N2 NA in complex with oseltamivir, and D) N2 NA in complex with zanamivir. Oseltamivir and zanamivir are shown in cyan and violet sticks, respectively, and the surface of NA is shown in gray surface representation. E) Average van der Waals contact potential energies for NA residues that contact substrates and inhibitors. Drug resistance residues in N1 NA, N2 NA, and in both subtypes are indicated with red stars, green triangles, and blue circles, respectively.

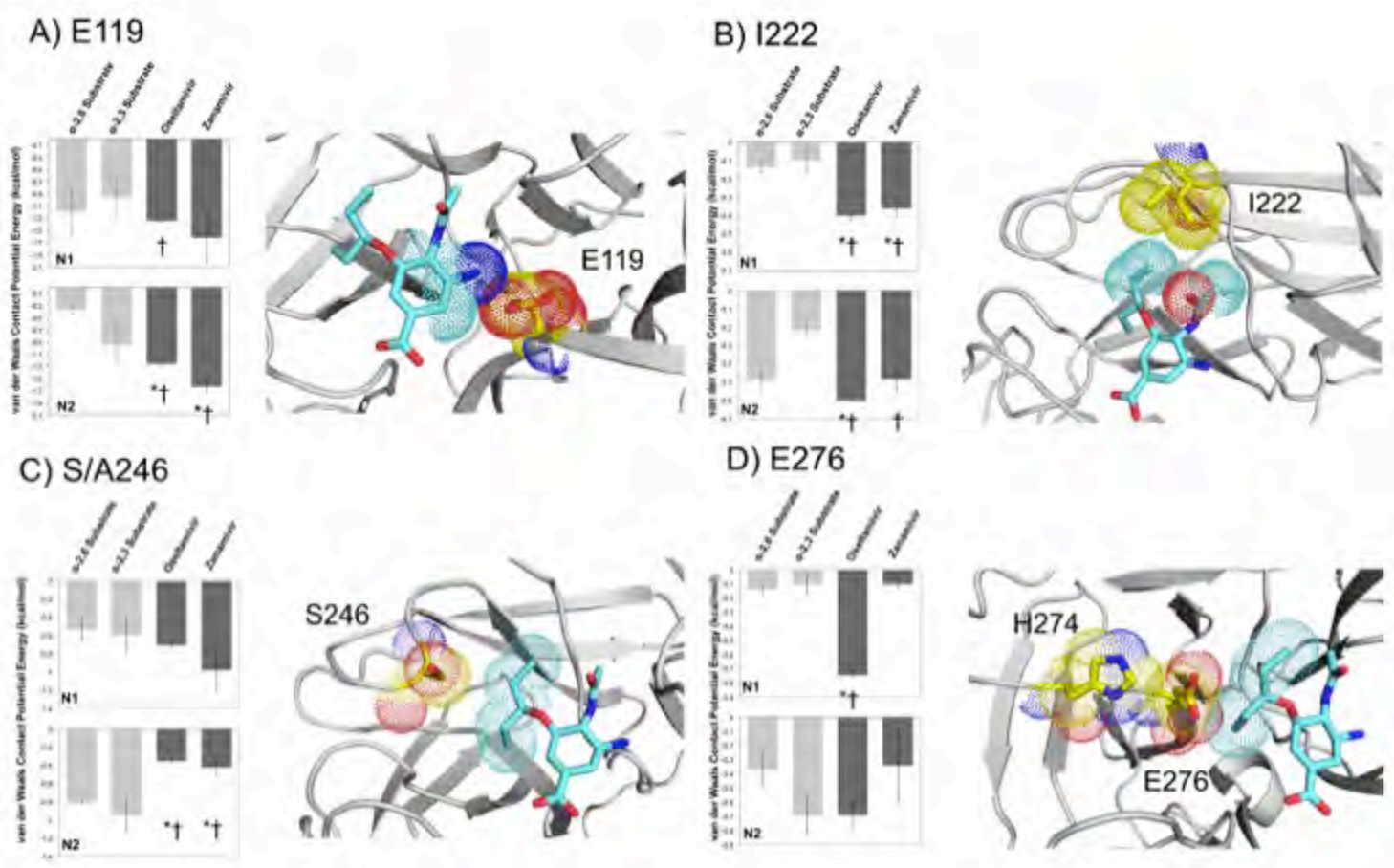


Figure 2.8 Van der Waals interactions for drug resistance residues in NA.

Figure 2.8 Van der Waals interactions for drug resistance residues in NA. Each figure panel corresponds to either a clinically observed drug resistance residue or a residue directly involved in drug resistance: A) E119, B) I222, C) S/A246, and D) E276. For panel C), residue 246 is S in N1 and A in N2. In each panel, histograms of average intermolecular van der Waals contact energies during MD simulations are shown for N1 NA on the top and for N2 NA on the bottom. Substrates are in light gray and inhibitors are in dark gray. On the right of each figure panel, the structure of NA is shown in gray cartoon representation, the indicated residue is shown in yellow sticks, and oseltamivir is shown in cyan sticks. Dots show van der Waals radii of atoms. Significant difference in the average van der Waals contact energies ($p < 0.05$) between the α-2,3 and α-2,6 substrates and inhibitors is indicated by an asterisk (*) and a cross (†), respectively.

A)

N1	Residue	α-2,6 Substrate	α-2,3 Substrate	Oseltamivir	Zanamivir
406	TYR	17%	25%	0%	0%
371	ARG NH2	89%	87%	78%	65%
371	ARG NH1	89%	66%	76%	38%
347	ASN	16%	39%	0%	0%
292	ARG	67%	76%	91%	71%
277	GLU	6%	4%	0%	49%
227	GLU	0%	0%	0%	84%
152	ARG	14%	11%	57%	34%
151	ASP	18%	5%	91%	40%
119	GLU	38%	25%	81%	85%
118	ARG	73%	46%	65%	36%

N2	Residue	α-2,6 Substrate	α-2,3 Substrate	Oseltamivir	Zanamivir
406	TYR	3%	23%	0%	3%
371	ARG NH2	90%	90%	90%	65%
371	ARG NH1	90%	90%	90%	67%
347	HIS	36%	3%	0%	10%
292	ARG	88%	93%	84%	16%
277	GLU	7%	23%	0%	24%
227	GLU	0%	0%	0%	61%
152	ARG	12%	7%	51%	48%
151	ASP	7%	10%	85%	24%
119	GLU	10%	15%	84%	81%
118	ARG	56%	58%	36%	43%

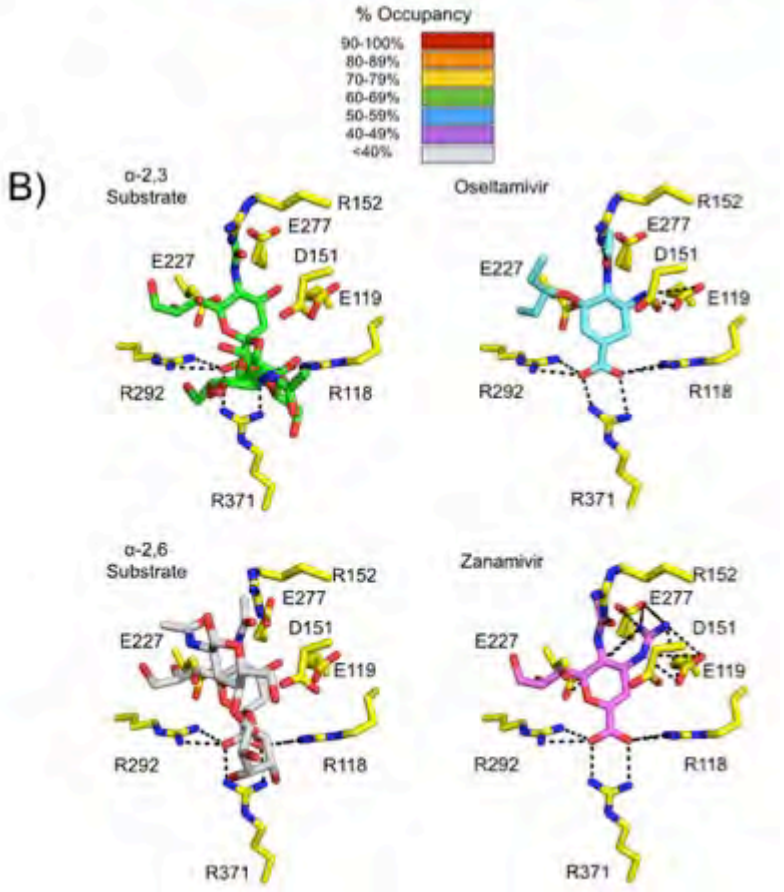


Figure 2.9 Intermolecular hydrogen bonding interactions during MD simulations.

Figure 2.9 Intermolecular hydrogen bonding interactions during MD simulations. A) Tables of the percentage of time that intermolecular hydrogen bonds are present during MD simulations in N1 and N2 NA. Drug resistance residues are in italics and catalytic residues are in bold. B) The residues involved in each intermolecular hydrogen bond to the four ligands are shown in yellow sticks, and hydrogen bonds are shown with black dotted lines. The α-2,3 substrate, α-2,6 substrate, oseltamivir, and zanamivir are shown in green, gray, cyan, and violet sticks, respectively.

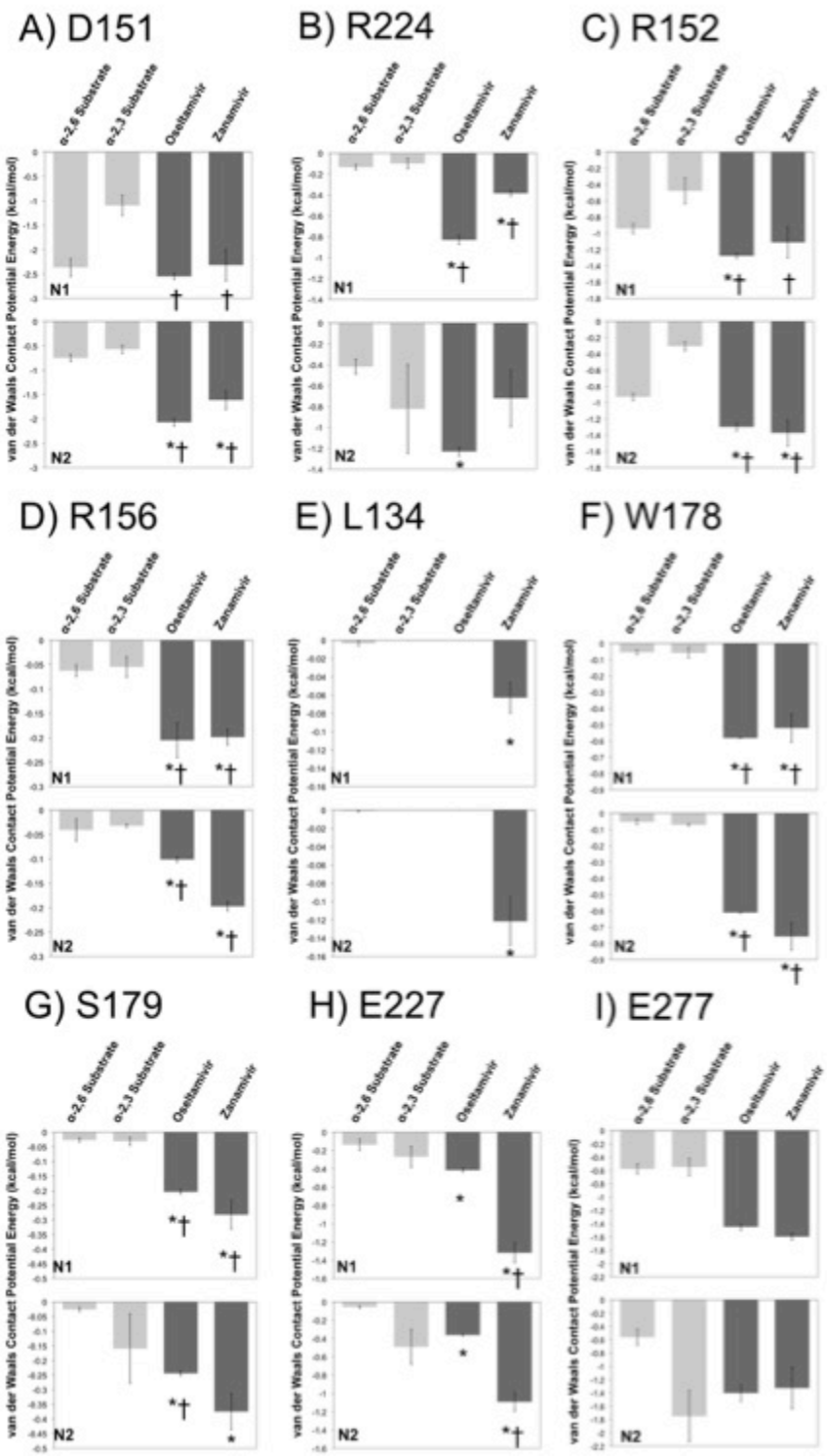


Figure 2.10 Increased van der Waals contact potential energies of inhibitors compared to substrates.

Figure 2.10 Increased van der Waals contact potential energies of inhibitors compared to substrates. Each figure panel A) through H) corresponds to a residue that has been associated with decreased neuraminidase inhibitor susceptibility. In each panel, histograms of average intermolecular van der Waals contact energies during MD simulations are shown for N1 NA on the top and for N2 NA on the bottom. Substrates are in light gray and inhibitors are in dark gray. Significant difference in the average van der Waals contact energies ($p < 0.05$) between the α-2,3 and α-2,6 substrates and inhibitors is indicated by an asterisk (*) and a cross (†), respectively.

Residue I222 is a primary drug resistance residue in N1 NA and a secondary permissive mutation in N2 NA, and differences in vdW contacts also explain this pattern (**Figure 2.8B**) (84). I222 contacts the N-acetyl group and the glycerol/pentyl-ether hydrophobic moieties in ligands. When comparing the vdW contacts for residue I222 in N1 and N2, I222 is makes extensive vdW contact with both substrate and inhibitor binding in N2, and therefore I222 is not an optimal site for drug resistance in N2. However, in N1, I222 is makes much more extensive vdW contact with inhibitors compared to substrates. A drug resistance mutation would be well tolerated at this site, as both oseltamivir and zanamivir have greater vdW contacts with N1 compared to substrates. Since I222 does not appear to be necessary for substrate binding in N1, mutations at this site confer strong fitness advantages in the presence of inhibitor, as shown experimentally in fitness experiments (97). In fact, I222 appears to be a “hotspot” location for drug resistance mutations in N1, where many drug resistance mutations are well tolerated and provide wild type like fitness for N1 in the presence of oseltamivir (97).

S246N is a primary drug resistance mutation in N1 NA, but not N2, in both in vitro experiments and clinical samples (90). When comparing vdW contacts in N1 and N2, residue 246 is interacts more extensively in substrate recognition than inhibitor binding in N2, and therefore this location is not an optimal site for drug resistance (**Figure 2.8C**). However, in N1, N246 is more important for

inhibitor binding compared to substrate recognition, especially for the α-2,6 substrate, indicating that a drug resistance mutation may be tolerated at this site in N1. Thus, the differences in vdW contacts in comparison to substrates correlate well with S246N being a resistance mutation in N1 only.

vdW contacts with residue E276 also explain the differential patterns of drug resistance observed for mutation H274Y. Binding of oseltamivir to NA requires that residue E276 rotate to create a larger pocket in the active site that can accommodate the pentyl-ether hydrophobic moiety on oseltamivir (98). Previous studies have shown that the drug resistance mutation H274Y in N1 NA prevents rotation of E276 so that the active site can no longer accommodate oseltamivir, but can still bind substrates and zanamivir. Since oseltamivir and peramivir both contain the pentyl-ether hydrophobic moiety, both inhibitors are impacted by the mutation H274Y. Although residue H274 does not directly contact ligands, H274 mediates the interaction of residue E276 with ligands that contain a bulky hydrophobic group at this position. When comparing vdW contacts for residue E276 in N1 and N2 N, hydrophobic interactions of E276 are much more important for oseltamivir binding in N1 NA than substrate or zanamivir binding. This result indicates that a residue that may impact the orientation of E276, such as drug resistance mutation H274Y, may prevent inhibitors with a bulky hydrophobic group, such as oseltamivir and peramivir, from binding (**Figure 2.8D**). In corresponding experimental fitness experiments in N1 NA in the presence or absence of oseltamivir as a selective pressure,

mutations at position E276 cause null-like fitness (97). However, mutation H274Y confers wild type like fitness in the presence of oseltamivir.

For remaining residues that make increased vdW contacts with inhibitors compared to substrates, these residues have been associated with either *in vitro* drug resistance or decreased NA inhibitor susceptibility (**Figure 2.10 and Tables 2.2-2.3**) (41, 99, 100). These residues are active site and framework residues that are well conserved, and mutations at these locations do not seem to be tolerated in experimental fitness measurements in N1 (97). The differential behavior of these residues for inhibitor versus substrate binding may explain some of the decreased NA inhibitor susceptibility that has been observed with mutations at these residues.

2.3.4 Hydrogen Bond Interactions

NA inhibitors were designed to optimize interactions in the active site, such as hydrogen bonds and electrostatic interactions, and therefore there are larger numbers of specific hydrogen bonds with NA active site residues compared to the substrates (**Figure 2.9**) (101). Substrates form the most prevalent hydrogen bonds with residues R371, R292, and R118 (**Figure 2.9A**). These residues form hydrogen bonds to the carboxylic acid adjacent to the scissile bond in the substrates, and may help stabilize the substrate in the active site during the cleavage reaction (**Figure 2.9B**) (102). These residues are also conserved evolutionarily in both N1 and N2 sequences; R118, R292, and R371

are 99% conserved. Inhibitors form additional hydrogen bonds with additional NA residues. Oseltamivir makes hydrogen bonds with R152 and E119. E119 is a drug resistance mutation in N1 NA, and R152 is a drug resistance mutation in type B influenza NA. Mutations at R152 have been associated with decreased NA inhibitor susceptibility (84). Zanamivir forms hydrogen bonds with E119 in both N1 and N2, R152 in N2, and E227 and E277 in both N1 and N2. E227 and E277 have been associated with decreased NA inhibitor susceptibility but not reported as major drug resistance sites (41). Oseltamivir makes significant hydrogen bonds with D151 in both N1 and N2 (93% and 94% of the simulation time, respectively), and these hydrogen bonds are also present in zanamivir (40% and 24% of the simulation time, respectively). These hydrogen bonds are not as prevalent with substrates even though D151 is a catalytic residue (less than 20% and 10% of the simulation time in N1 and N2, respectively). Thus, in addition to increased vdW contacts, additional hydrogen bonds to non-conserved NA active site residues beyond those of substrates correlate with sites of drug resistance mutations in the two subtypes.

2.4 Discussion

In this study, we used the *dynamic substrate envelope* to explain differential patterns of drug resistance between N1 and N2 subtypes. This method is useful for understanding drug resistance where vdW interactions play a significant role in the drug resistance mechanism. However, for understanding

more complex drug resistance mechanisms, such as those related to mutations that primarily involve electrostatic interactions, like mutations R292K and N294S, and for understanding double mutations and secondary permissive mutations, more in-depth study is required. Another mechanism of NA inhibitor drug resistance, which is even more complex, occurs when hemagglutinin binding to sialic acid is weakened, thereby decreasing the dependence of the virus on NA cleavage activity.

D151E is another complex resistance mutation involving a catalytic residue (102). However, hydrogen bonds between inhibitors and D151 may be more stabilizing for inhibitor binding than substrate binding (**Figure 2.9**), and the mutation D151E may be able to prevent inhibitor binding while still maintaining catalytic activity. The mutation D151G allows unhydrolyzed substrate to bind and be visualized in the active site through crystallography, which supports the concept that while D151 is necessary for catalysis, this side chain is not required for recognition (48). However, mixed populations of virus with drug resistance mutations D151D/E and D151V/D have also been observed clinically and during influenza surveillance, so further study is needed to fully understand the drug resistance mechanism of this residue (41, 85, 86).

This study examined models and crystal structures of influenza neuraminidase in complex with ligands that are similar to naturally occurring substrates, such as 3'-sialyl-N-acetyllactosamine and 6'-sialyl-N-acetyllactosamine. However, the solvent exposed GlcNac termini of these

ligands may be more flexible than naturally occurring glycans because sialic acids are tethered to polysaccharide chains that are longer than three sugar residues and can be found as terminating branches of N-glycans, O-glycans, glycosphingolipids (gangliosides), and side chains on glycophosphatidylinositol (GPI) anchors, which are glycolipids on the surface of cell membranes often associated with C-termini of proteins during post-translational modification (103-105). During the molecular dynamics simulations examined in this study, extensive motion in the GlcNac termini of substrates was observed, but this motion is likely to be reduced biologically because sialic acid is tethered to polysaccharides on glycoproteins and polysaccharides attached to glycolipids in the cell membrane. Although the motion of the GlcNac would likely be reduced as a result of being tethered, the motion from the tethering object, whether it is a glycoprotein or glycolipid, likely still contributes to the increased flexibility that is seen in sialic acid compared to the inhibitors oseltamivir and zanamivir, which are not tethered and have a tight subnanomolar binding affinity. Future analysis of this motion in tethered substrates may more accurately identify which molecular interactions in the active site are critical for substrate recognition.

The *dynamic substrate envelope* of NA can be incorporated into the design of novel NA inhibitors to prevent the development of drug resistance. Inhibitors can be designed to target evolutionarily conserved residues important for substrate binding and to fill the *substrate envelope* more optimally to take advantage of the remaining volume in the *substrate envelope* that is unfilled by

inhibitors. In addition, inhibitor rigidity may promote susceptibility to drug resistance mutations in the active site. Substrates can accommodate binding to an active site with drug resistance mutations because they are more flexible, so inhibitors can be designed to better mimic these substrate dynamics. However, the high rigidity in inhibitors may also be important for tight binding interactions in the active site. Inhibitors must be optimized to balance tight binding interactions that contribute to high potency while sharing critical features of substrates, such as substrate flexibility. Following these guidelines may heighten the barrier to the development of drug resistance.

2.5 Methods

2.5.1 Influenza Neuraminidase Substrate and Inhibitor Complex Structures

We chose prototypic N1 and N2 sequences for this study based on three criteria: 1) existence of high quality high resolution crystal structures for MD simulations 2) presence of a “typical” 150-loop in the active site based on previous reports for N1 and N2, and 3) high percent identity to N1 and N2 consensus sequences based on multiple sequence alignments (72, 106). All of the crystal structures used in this study are of the globular head domain. Alignments were performed using the multiple sequence alignment tools available on the Influenza Research Database (www.fludb.org) and accessed on February 18, 2016. The strain of N2 NA used is A/Tanzania/205/2010 H3N2 NA. This strain has 94% sequence identity and 96% sequence similarity to a

consensus sequence determined from an alignment of 8,745 complete and unique sequences. The strain of N1 NA used is A/Brevig Mission/1/1918 H1N1 NA. This strain has 92% sequence identity and 96% sequence similarity to a consensus sequence determined from an alignment of 7,370 complete and unique N1 NA sequences

For N2 NA, crystal structures were available in complex with α-2,6 and α-2,3 substrates (PDB ID: 4GZX and 4GZW, respectively) (48). A crystal structure of the same strain of N2 NA was also available in complex with oseltamivir (PDB ID: 4GZP). For N1 NA, the highest quality crystal structure available of N1 NA is in complex with zanamivir (PDB ID: 3B7E) (107). Additional models were created using these four structures for co-complex crystal structures that were unavailable (**Table 2.4**).

2.5.2 Structure Preparation

Crystallographic waters and calcium ions were retained, and all buffer salts were removed. The substrate co-crystal structures have a D151G substitution in the active site to prevent catalytic activity and allow substrates to be captured in the active site. To more accurately model the interactions of ligands with this residue, the back-mutation G151D was modeled *in silico* using the software Maestro and Prime from Schrodinger (108, 109). Crystal structures were prepared using the Protein Preparation Wizard from Schrodinger (110).

2.5.3 Molecular Dynamics Simulation Protocol

We performed 100 nanosecond MD simulations for each tetrameric complex using Desmond and the OPLS2005 force field (111, 112). Each system was solvated with a 10 Å pad of TIP3P waters in a truncated octahedron solvent box. Sodium (Na^+) or chloride (Cl^-) counterions were added to neutralize the overall charge of the system. Each system was energy minimized using a relaxation protocol to relieve steric clashes before initiating production stage MD calculations. After minimization, each system was equilibrated using a sequence of four short MD stages, following the default relaxation process for an NPT ensemble published in the Desmond User Manual with modifications (113). For the production stage, MD simulations were performed for 100 ns and 1 atm in the NPT ensemble using a Nose-Hoover thermostat and a Martyna-Tuckerman-Klein (MTK) barostat. Long-range electrostatics were calculated using the Particle Mesh Ewald method with a cutoff radius of 9 Å. For each system, the trajectories of each monomer were concatenated to provide 400 ns of sampling.

2.5.4 Root Mean Squared Deviation (RMSD)

For each system, trajectories were sampled at intervals of 200 ps. Root mean squared deviation (RMSD) calculations were performed using a custom TCL-TK script in the Visual Molecular Dynamics software package (VMD) (114). Before RMSD values were calculated, the frames from each interval were aligned to the first frame of the trajectory, and RMSD values were calculated

using all backbone alpha carbon atoms. In addition, five N-terminal and C-terminal residues were omitted from the additional RMSD calculations to show that the approximately 375 central amino acids of each monomer were highly stable and equilibrated rapidly.

2.5.5 Root Mean Squared Fluctuation (RMSF) and Simulation-Derived Temperature Factors

Root mean squared fluctuations (RMSF) and simulation-derived temperature factors (B factors) were calculated for all alpha carbons in the trajectory and averaged over 400 ns using VMD. RMSF was calculated using the built in rmsf command in VMD, and simulation-derived B factors were calculated using a TCL-TK script on the VMD website (115). For crystal structures with more than one molecule in the asymmetric unit, temperature factors were averaged over all molecules for comparison.

2.5.6 Dynamic Substrate Envelope

All trajectories were aligned based on the alpha carbons of a subset of structurally rigid residues using VMD. The van der Waals (vdW) volumes of ligand conformers in the active site of each monomer from each trajectory were mapped onto a three-dimensional grid, and a probability distribution was calculated for each grid point and plotted using in-house Fortran scripts. The vdW radii were defined by the OPLS2005 force field. The mathematical details have been reported previously (94).

2.5.7 Van der Waals Contact Potential

The van der Waals contact potential energies between ligands and NA were calculated over an MD trajectory and averaged using a simplified Lennard-Jones potential function defined by the following equation, where r_{ij} is the distance between NA atom *i* and ligand atom *j,* ε is the energy well depth, and σ is the collision diameter:

$$V(r_{ij}) = 4\varepsilon[(\frac{\sigma}{r_{ij}})^{12} - (\frac{\sigma}{r_{ij}})^{6}]$$

Van der Waals contact energies were calculated for all intermolecular atom pairs within a 6 Å cutoff of the binding interface using an in-house Fortran script. The nonbonded parameters were determined using the OPLS2005 force field, and values were averaged over 400 ns. Further details of this computation have been described previously (94).

2.5.8 Hydrogen Bond Calculations

The percentage of time that a hydrogen bond exists during a trajectory was calculated using the HBonds Plugin from VMD and averaged over 400 ns. A hydrogen bond was defined as having a donor-acceptor distance of a maximum of 3.5 Å and involving only polar atoms nitrogen, oxygen, sulfur, and fluorine. The donor-hydrogen-acceptor angle was also defined as being less than the cutoff of 30 degrees. Hydrogen bonds were summed over each residue and ligand except as indicated.

2.5.9 Plots and Figures

Microsoft Excel, Matlab, GraphPad Prism, PyMOL, and Geneious version 8.0.5 were used to create all plots and figures (116-119).

CHAPTER III

3 COMPARATIVE ANALYSIS OF HEPATITIS C NS3/4A INHIBITOR STRUCTURES

3.1 Abstract

Direct acting antivirals are increasing the efficacy and tolerability of hepatitis C virus treatment, but drug resistance has emerged with some of these inhibitors, including non-structural protein 3/4A protease inhibitors. Although many structures of PIs in complex with the NS3/4A protease have been reported, a systematic review of these crystal structures in the context of the substrate envelope hypothesis has not been performed. To provide a framework for designing better NS3/4A protease inhibitors with a higher barrier to resistance, we performed a quantitative structural analysis of PIs in clinical use or development using co-crystal structures and models of NS3/4A protease in complex with natural substrates and inhibitors. By comparing substrate structural motifs and active site interactions with inhibitor recognition, we observed that the development of drug resistance mutations correlates with how inhibitors deviate from viral substrates in molecular recognition.

3.2 Introduction

Hepatitis C is a blood-borne liver disease caused by the hepatitis C virus (HCV), that infects 3-4 million people each year (120). According to the WHO, approximately 150 million people are infected with HCV chronically, which is one of the most common reasons for liver transplants and can cause liver cirrhosis and hepatocellular carcinoma (120, 121). HCV is a genetically heterogeneous RNA virus, with six major genotypes and several subtypes within each genotype. HCV is also a member of the Flaviviridae virus family, which also includes West Nile Virus, dengue virus, and zika virus (122). Genotypes differ in sequence by approximately 30%, and genotype 1 is the most prevalent and until recently, was the most difficult to treat (123, 124). The error-prone RNA-dependent RNA polymerase generates high sequence variation, in addition to a high viral replication rate, making HCV a difficult target for designing effective antiviral drugs (125, 126). Although many direct acting antivirals (DAAs) against HCV have been developed, genetic diversity is a challenge for designing DAAs that effectively inhibit multiple genotypes and drug resistant variants of HCV (123, 125, 126).

While DAAs are increasing the efficacy and tolerability of HCV treatment, drug resistance has emerged with some of these inhibitors, including non-structural protein 3/4A protease inhibitors (NS3/4A PIs) (127). Although current PIs are effective against genotype 1 HCV, some PIs have less activity against other genotypes. In addition, first generation PIs, such as telaprevir (now

withdrawn (128)) and boceprevir, are limited in use due to their burdensome administration, considerable side effect profiles, drug-drug interactions, and low barriers to resistance (127). Second generation PIs, such as simeprevir, offer more benefits over first generation PIs, such as better side effect profiles and fewer drug-drug interactions, but the barrier to resistance is still low (123, 127).

By studying drug resistance in HIV-1 PIs and HCV NS3/4A PIs, we learned that the key to designing robust inhibitors with high barriers to resistance is to incorporate details of molecular recognition in substrate and inhibitor binding into drug design (33, 43, 129). Inhibitors that structurally mimic viral substrates are more likely to retain binding against drug resistant variants because these variant proteases must retain the ability to cleave viral substrates for survival. Inhibitors that mimic the dynamics of natural substrates are also more likely to maintain potency against drug resistant variants, and in general, additional flexibility allows inhibitors to better accommodate drug resistance mutations. To provide a framework for designing better NS3/4A PIs with a higher barrier to resistance, we performed a quantitative structural analysis of PIs in clinical use or development using co-crystal structures and models of NS3/4A protease in complex with natural substrates and inhibitors. By comparing substrate structural motifs and active site interactions with inhibitor recognition, we observed that the development of drug resistance mutations correlates with how inhibitors deviate from viral substrates in molecular recognition.

3.2.1 HCV NS3/4A Protease and Current Inhibitors

HCV NS3/4A helicase-protease is a 631 amino-acid protein with two domains, a N-terminal NTPase/helicase domain and a C-terminal serine protease domain, which is the target of NS3/4A PIs (**Figure 3.1**). The protease domain contains two beta-barrel subdomains and a zinc binding site, sharing a fold similar to chymotrypsin. To activate, HCV NS3 forms a heterodimer with the cofactor NS4A (123). The protease cleaves the HCV viral polyprotein, releasing proteins essential for viral maturation and infectivity. HCV protease also impairs host-mediated viral elimination by cleaving host proteins, including TRIF, which is involved in TRIF-mediated Toll-like receptor 3 (TLR3) signaling, and MAVS, which is involved in Cardiff-mediated retinoic acid-inducible gene 1 (RIG-1) signaling (130-133). The catalytic triad His 57, Asp 81, and Ser 139 in the active site of the protease domain hydrolyzes substrates and is located between the two beta-barrel subdomains. The protease active site is very shallow, which is another challenge for designing tight binding low molecular weight inhibitors (46).

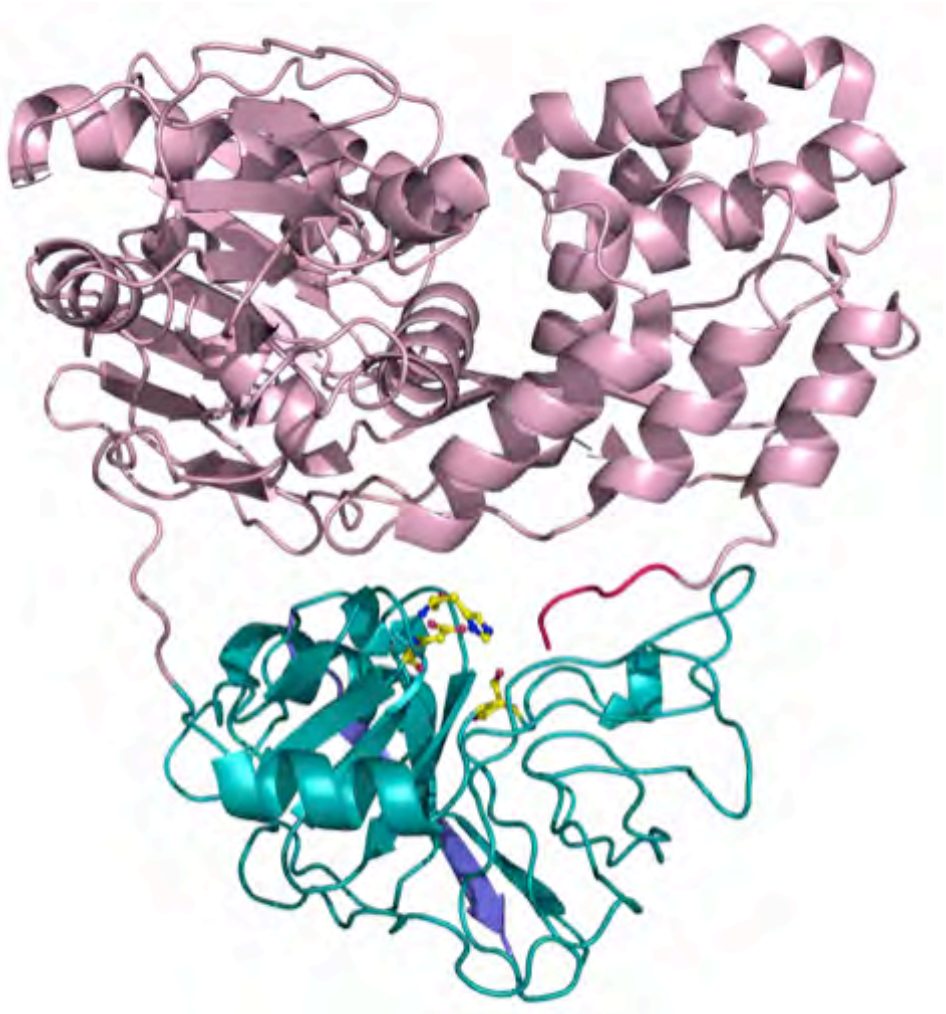


Figure 3.1 Cartoon representation of the hepatitis C NS3/4A helicase-protease structure.

Figure 3.1 Cartoon representation of the hepatitis C NS3/4A helicase protease structure. The N-terminal protease domain is green, the C-terminal helicase domain is pink, the last six amino acids in the C-terminus are in magenta, and the NS4A cofactor is blue. The catalytic triad H57, D81, and S139 is in yellow sticks.

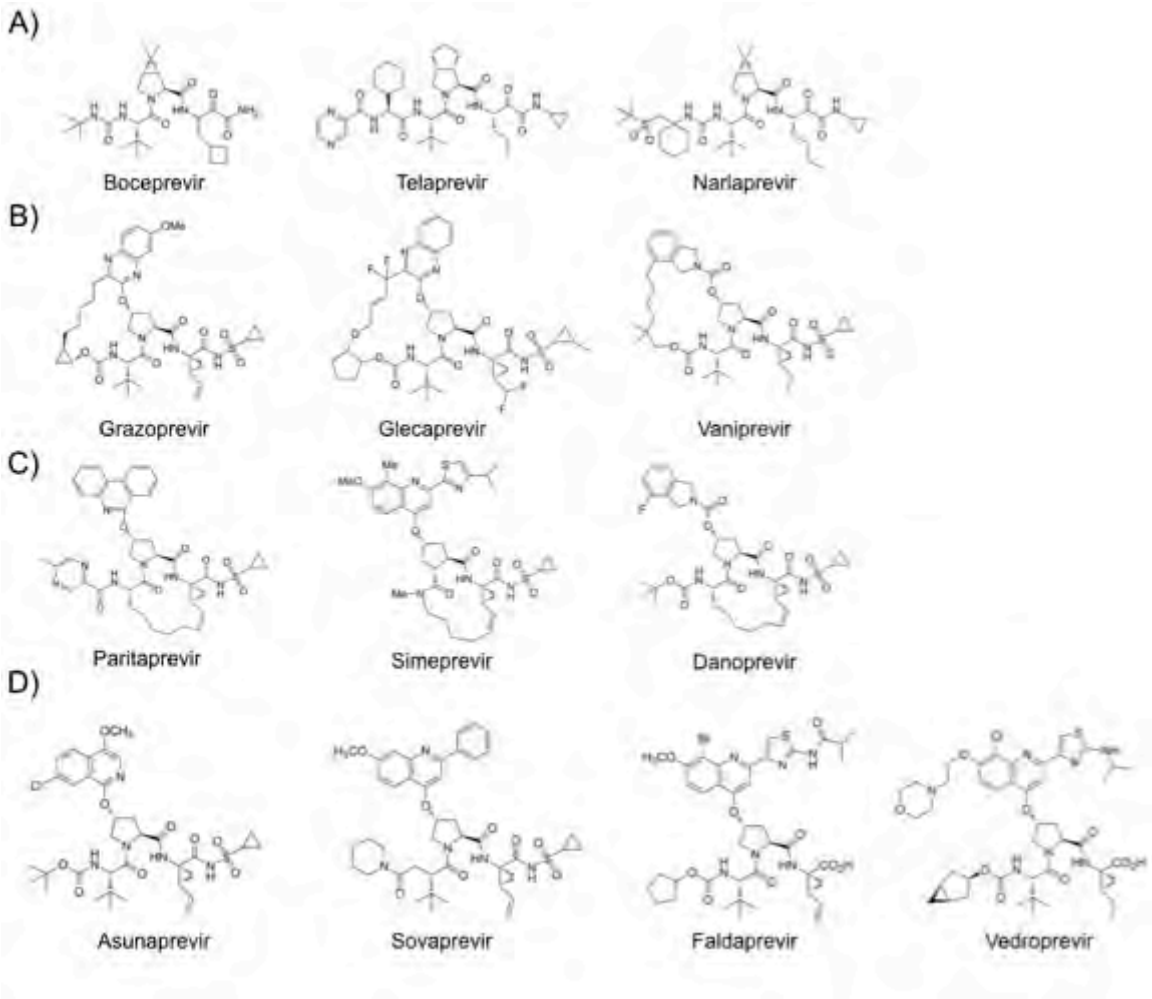


Figure 3.2 Hepatitis C NS3/4A protease inhibitors in clinical use or development.

Figure 3.2 Hepatitis C NS3/4A protease inhibitors in clinical use or development. A) Linear covalent ketoamide inhibitors, B) P2-P4 macrocyclic inhibitors, C) P1-P3 macrocylic inhibitors, and D) linear non-covalent inhibitors that have been FDA approved or are in Phase II or III clinical trials.

Five HCV PIs, telaprevir, boceprivir, simeprevir, paritaprevir, and grazoprevir, are FDA-approved, and several more PIs are in clinical development (**Figure 3.2**) (134). Faldaprevir, asunaprevir, vaniprevir, narlaprevir, glecaprevir, and GS-9857, are currently in Phase III clinical trials, while danoprevir, sovaprevir, and vedroprevir are in Phase II clinical trials (135). The structures of these inhibitors have been disclosed except for GS-9857. GS-9256 was also in Phase II clinical trials however inhibition of bilirubin transport and metabolism is preventing development (136, 137). Additional PIs with disclosed and undisclosed structures are in Phase I trials or pre-clinical development, such as ACH-2684 (deldeprevir), MK-2748, MK-6325, and GS-9857 (138-140). TG-2349 is in Phase II clinical trials in Taiwan and Phase I clinical trials in the United States (135).

NS3/4A PIs are prescribed in combination with other classes of DAAs and nonspecific antivirals, such as pegylated interferon-alpha and ribavirin (123). With many of the newly developed DAAs on the market, several interferon-free and ribavirin-free regimens are available, but many of these newer regimens are costly, so treatment choices are often limited by what is covered by an insurance provider.

NS3/4A PIs are peptidomimetics and competitive active site inhibitors, and they were designed based on the structure of natural substrates, whose cleavage products are weak inhibitors (141-143). Many of these inhibitors were co-crystallized with the protease domain and the cofactor NS4A. Similar to substrate

positions in the active site, inhibitor moieties are named based on the analogous substrate residue in the active site. For instance, an inhibitor moiety in a position analogous to the P1 substrate residue is called a P1 inhibitor moiety.

The FDA approved the first NS3/4A PIs, boceprevir and telaprevir, in 2011, two potent acyclic peptidomimetic ketoamide inhibitors (28, 144-146). These PIs contain an alpha-ketoamide group that reversibly covalently bonds with the catalytic residue Ser 139. In addition, these inhibitors form short-range electrostatic and van der Waals interactions with the binding site (46). Designed to target the active site of genotype 1 NS3/4A proteases, these inhibitors are most effective against genotype 1 and also inhibit genotypes 2, 5, and 6 *in vitro*, but they are least effective against genotype 3 (147, 148). Narlaprevir is another linear ketoamide inhibitor in Phase III clinical trials, which has improved potency, pharmacokinetic profile, and physicochemical properties (147, 149).

In addition to the linear covalent ketoamide PIs, there are linear non-covalent peptidomimetics, including faldaprevir, asunaprevir, sovaprevir and vedroprevir (26, 150-154). Faldaprevir and asunaprevir are in Phase III clinical trials, while sovaprevir and vedroprevir are in phase II clinical trials (135). All four of these inhibitors have a quinoline or isoquinoline moiety in the P2 position. Asunaprevir and sovaprevir are related linear acylsulfonamide inhibitors that only differ in structure at the P4 capping group and the P2 extension. In addition, sovaprevir demonstrates high potency against all HCV genotypes except genotype 3 and has a low side effect profile (139, 147). Asunaprevir is a highly

potent inhibitor with *in vitro* activity against genotypes 1 and 4, but drug resistance mutations have emerged in patients infected with HCV genotype 1b as well as associated with hepatotoxicity, limiting its use in combination with pegylated interferon alpha and ribavirin (123, 155, 156).

Faldaprevir and vedroprevir are related linear C-terminal carboxylic acid inhibitors that span the P4-P1 substrate sites, and they also only differ in structure at the P4 capping group and the P2 extension. These inhibitors were designed based on the observation that the C8 substituent (bromide in faldaprevir and chloride in vedroprevir) on the P2 quinoline B-ring improved the cell-based potency and pharmacokinetic profile of these inhibitors (147, 150). In particular, faldaprevir is a once-daily selective inhibitor with a good absorption, distribution, metabolism, and excretion (ADME) profile and favorable pharmacokinetics (123, 147).

The other group of non-covalent NS3/4A PIs is the acylsulfonamide macrocyclic inhibitors, which include vaniprevir, danoprevir, glecaprevir, grazoprevir, paritaprevir, and simeprevir (25, 147, 157-159). Grazoprevir, paritaprevir, and simeprevir are FDA-approved inhibitors. Vaniprevir and glecaprevir are in Phase III clinical trials, and danoprevir is in Phase II clinical trials (135). These inhibitors contain a macrocycle that connects either the P1 and P3 moieties (simeprevir, paritaprevir, and danoprevir) or the P2 and P4 moieties (grazoprevir, glecaprevir, and vaniprevir). In general, macrocyclic PIs demonstrate higher subnanomolar affinities compared to linear inhibitors, but

drug resistance mutations have also emerged. However, macrocyclic inhibitors are better able to accommodate drug resistance mutations compared to linear inhibitors. Some of these inhibitors have been designed for increased pan-genotypic efficacy, such as grazoprevir, paritaprevir, and glecaprevir (123, 160, 161).

Many of these macrocyclic inhibitors are related compounds with differences in specific substituents or moiety positions. Vaniprevir and danoprevir are related compounds with similar isoindoline P2 moieties, P4 capping groups, and carbamate linkages between the P4 and P3 moieties, but they differ in macrocyclization. Vaniprevir demonstrates excellent selectivity against a panel of 169 pharmacologically relevant receptors, enzymes, and ion channels (IC_{50} > 10 μM) except for chymotrypsin (IC_{50} = 520 nM), and danoprevir is unique with a low incidence of viral rebound after combined treatment with PEG interferon alpha-2a (162, 163). Grazoprevir and glecaprevir are related compounds that have similar quinoxiline P2 moieties and carbamate linkages between the P4 and P3 moieties, but they differ in macrocyclization, P4 capping groups, and P2 quinoxiline substituents. Grazoprevir is active against genotypes 1, 2, and 3, whereas many NS3/4A PIs lose potency against genotype 3 (164). Paritaprevir, which is an inhibitor developed by Abbvie, has a phenanthridine P2 moiety, and it was recently FDA approved in December 2014 in a combination medication with ombitasvir and ritonavir (165). Simeprevir is smaller than other macrocyclic inhibitors, has a P1-P3 macrocycle, and has a quinoline P2 group. Simeprevir

spans the P3-P1' sites compared to the other macrocyclic inhibitors, which span the P4-P1' sites, and has a high selectivity ratio of 5,785 *in vitro*, which is the ratio of the 50% cytotoxic concentration and the 50% effective concentration in cell-based assays (166). Over the past few years, pan-genotypic inhibitors have also emerged. Glecaprevir is a P2-P4 macrocylic inhibitor in Phase III clinical trials that is highly potent against many genotypes (EC_{50} = 0.85-2.7 nM), and in particular, it has high activity against genotype 3a (EC_{50} = 1.6 nM) (167). It is currently being investigated as an oral, ribavirin free, one-a-day combination therapy with NS5A inhibitor ABT-530 for patients with genotype 1-6 HCV infection (167).

3.2.2 Drug Resistance to HCV NS3/4A Protease Inhibitors

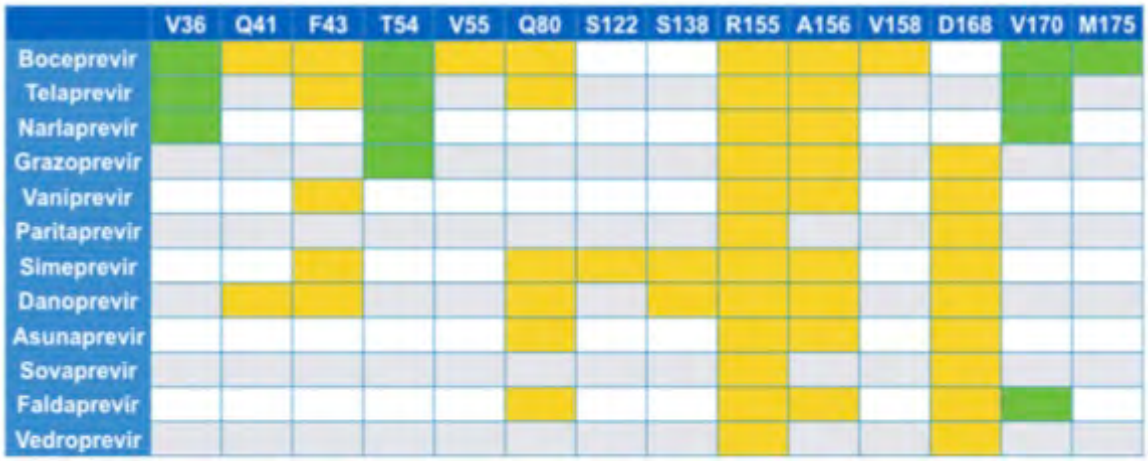
Drug resistance mutations have emerged against NS3/4A protease inhibitors both experimentally and clinically in 5-15% of patients (124, 160, 168-174). These mutations occur both inside and outside the protease active site (**Table 3.1**). Drug resistance mutations at residues R155, A156, and D168 impact the most inhibitors. These residues are located directly in the active site, and they are also the main sources of drug resistance clinically (124). In addition to drug resistant variants with single mutations, variants with multiple mutations have also been reported, such as V36M and R155K. Boceprevir and telaprevir are the oldest inhibitors and are most susceptible to drug resistance, with main drug resistance mutations at residues V36, T54, R155, and A156 (124, 168).

These resistance mutations cause three orders of magnitude changes in Ki and IC_{50} values (175). Macrocylic PIs are less susceptible to resistance, but are still impacted by mutations at residues R155, A156, and D168. Grazoprevir has increased activity against multiple genotypes compared to other PIs and a more favorable and flatter drug resistance profile. Grazoprevir is still susceptible to resistance mutations at R155, A156, and D168, but the fold changes in Ki and IC_{50} vary widely. For instance, the resistance mutation R155K increases the Ki for grazoprevir by 6 fold, but the resulting Ki and IC_{50} measurements are still subnanomolar. However, for the resistance mutations A156T and D168A, the fold changes in Ki and IC_{50} are much greater.

Using the *substrate envelope hypothesis* to understand drug resistance in HIV and HCV protease inhibitors, we discovered that substrates fill a conserved volume within the active site known as the *substrate envelope*, and primary drug resistance mutations occur where inhibitors protrude outside of the *substrate envelope* and contact residues that are not evolutionarily conserved or are not necessary for biological function (43, 44). Although many structures of PIs in complex with the NS3/4A protease have been reported, a systematic review of these crystal structures in the context of the *substrate envelope hypothesis* has not been performed. This type of a study would provide insight into how different patterns of resistance have emerged for different classes of NS3/4A PIs and may also help us predict resistance for inhibitors that are still in development. We used the *substrate envelope hypothesis* to analyze crystal structures of NS3/4A

protease in complex with inhibitors that have been FDA approved or are in Phase II or III clinical trials. We also modeled four additional PIs that have not yet been crystallized. Understanding how different patterns of drug resistance have emerged for different classes of NS3/4A PIs and applying this information to predict resistance would be useful for designing inhibitors that are less susceptible to resistance.

Table 3.1 Sites of drug resistance mutations to HCV NS3/4A protease inhibitors.



The top row lists residues where resistance mutations have occurred. Colored boxes indicate that a drug resistance mutation was observed. Green boxes are for residues outside of the active site, and yellow boxes for are for residues that contact inhibitors.

3.3 Results

3.3.1 Substrate Envelope and V_{IN} and V_{OUT}

High resolution crystal structures of the protease domain in complex with cleavage products of four viral polyprotein substrates demonstrate that viral substrates fill a conserved volume in the active site, known as the crystallographic substrate envelope (**Figure 3.3**) (43). Even though substrates are non-homologous, they adopt a conserved shape in the active site. Previous studies established that the most severe drug resistance mutations occur where inhibitors protrude beyond the substrate envelope and contact binding site residues more important for inhibitor binding compared to substrate binding. Three sites of highly prevalent drug resistance mutations, Arg 155, Ala 156, and Asp 168, do not contact natural substrates but are more critical for inhibitor binding.

The fit of inhibitors within the substrate envelope, represented by V_{IN} and $V_{OUT,}$ also correlates with loss of binding affinity to NS3/4A protease drug resistant variants (**Figures 3.4 and 3.5**). V_{IN} and V_{OUT} are the volumes of inhibitors that are contained within and protrude outside of the substrate envelope, respectively. The linear ketoamide inhibitors telaprevir, boceprevir, and narlaprevir do not protrude outside of the substrate envelope in the P2 moiety as much as the macrocylic inhibitors, and consequently they have lower V_{OUT} volumes. However, the macrocylic inhibitors have higher potency, with better Ki and IC_{50} values (175). For linear ketoamide inhibitors in the presence of drug

resistance mutations, fold changes in Ki and IC_{50} values range from 2-20 fold, whereas these same changes in macrocyclic inhibitors are 10-1000 fold (175). However, since macrocyclic inhibitors have much stronger subnanomolar or nanomolar binding affinity in wild-type protease, losses of affinity in drug resistant variants still result in relatively tight binding inhibitors. For instance, vaniprevir and grazoprevir have subnanomolar activity against wild-type NS3/4A protease, and these inhibitors still retain nanomolar activity in the presence of drug resistance mutations. (175)

However, many inhibitors do not take advantage of the remaining space within the substrate envelope to make additional contacts with the protease. Only narlaprevir and telaprevir extend into the P5 position and none of the inhibitors extend into the P6 position. Most inhibitors also fill the P1' volume. Full un-cleaved substrates in the active site have not yet been crystallized, so this portion of the substrate envelope remains undefined but is likely still useful for designing compounds that are less susceptible to drug resistance. Optimizing the balance between staying within and filling the substrate envelope, minimizing protrusions from the substrate envelope, and maintaining a high level of activity are important for designing potent inhibitors with flat resistance profiles.

A)

	P6	P5	P4	P3	P2	P1	P1´	P2´	P3´	P4´
5A5B	E	D	V	V	C	C	S	M	S	Y
4B5A	E	C	T	T	P	C	S	G	S	W
4A4B	D	E	M	E	E	C	S	Q	H	L
3-4A	D	L	E	V	V	T	S	T	W	V

B)

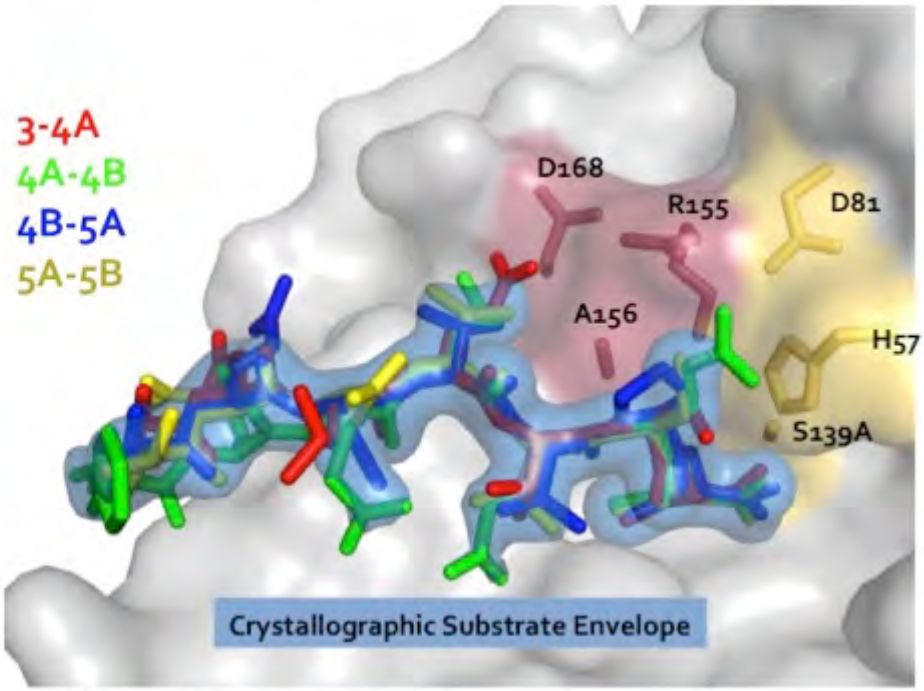


Figure 3.3 Viral substrates of HCV NS3/4A protease share a conserved binding mode in the active site.

Figure 3.3 Viral substrates of HCV NS3/4A protease share a conserved binding mode in the active site. (A) Amino acid cleavage site sequences of four HCV NS3/4A protease viral substrates. (B) Viral substrates fill a conserved volume in the active site known as the crystallographic substrate envelope. The NS3/4A protease domain is in gray surface representation. Substrate 3-4A is in red sticks, substrate 4A-4B is in green sticks, substrate 4B-5A is in blue sticks, substrate 5A-5B is in yellow sticks, and the substrate envelope is in blue surface representation. The catalytic triad D81, H57, and S139 is in yellow sticks, and sites of three reported drug resistance mutations are in burgundy sticks.

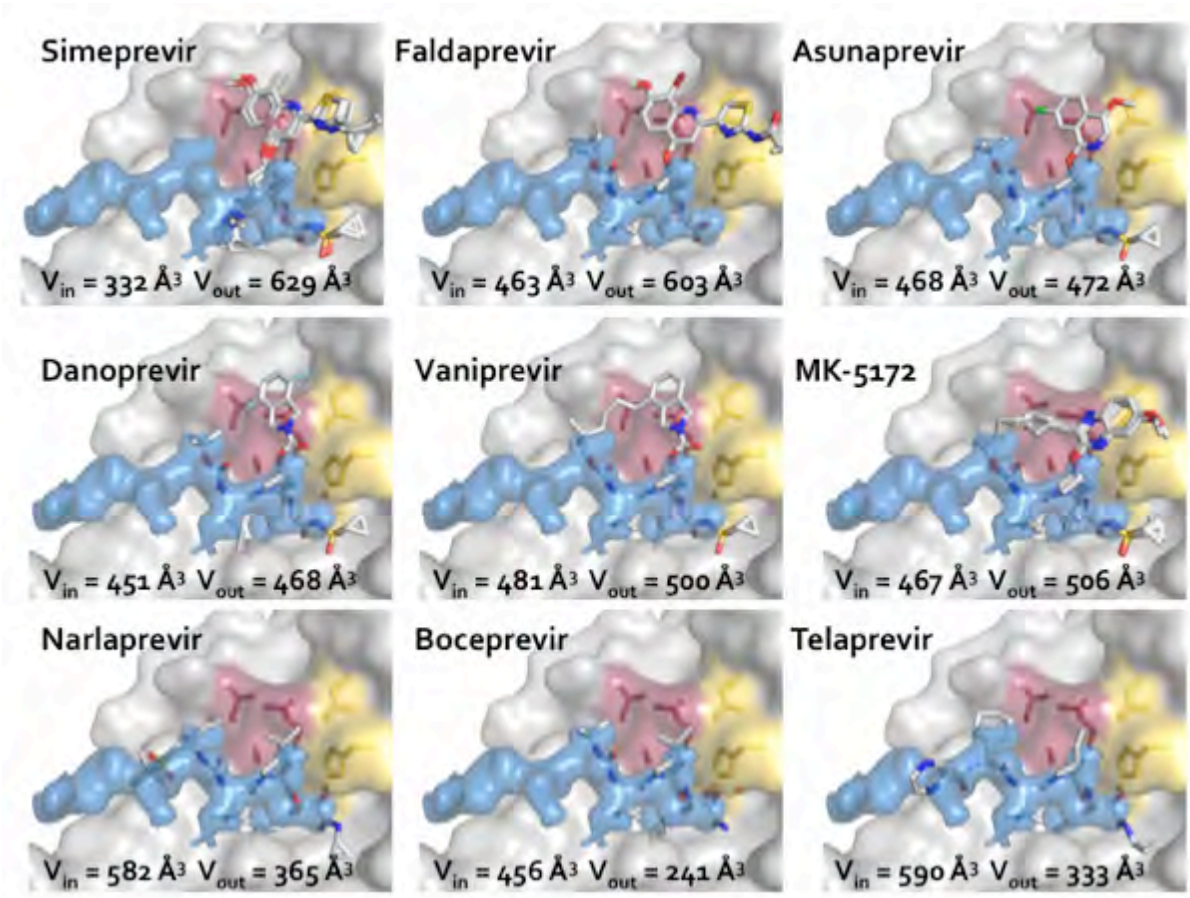


Figure 3.4 Inhibitors protruding outside of the substrate envelope in HCV NS3/4A protease.

Figure 3.4 Inhibitors protruding outside of the substrate envelope in HCV NS3/4A protease. Drug resistance mutations occur where inhibitors protrude beyond the substrate envelope and contact residues that are not important for substrate binding. The NS3/4A protease domain is in gray surface representation, and the substrate envelope is in blue surface representation. The catalytic triad D81, H57, and S139 is in yellow sticks, and sites of drug resistance residues R155, A156, and D168 are in burgundy sticks. Inhibitors are in stick representation with carbon atoms in gray and other atoms in standard CPK colors. The volume of each inhibitor inside (V_n) and outside (V_{out}) the substrate envelope is reported in $Å^3$. MK-5172 is grazoprevir.

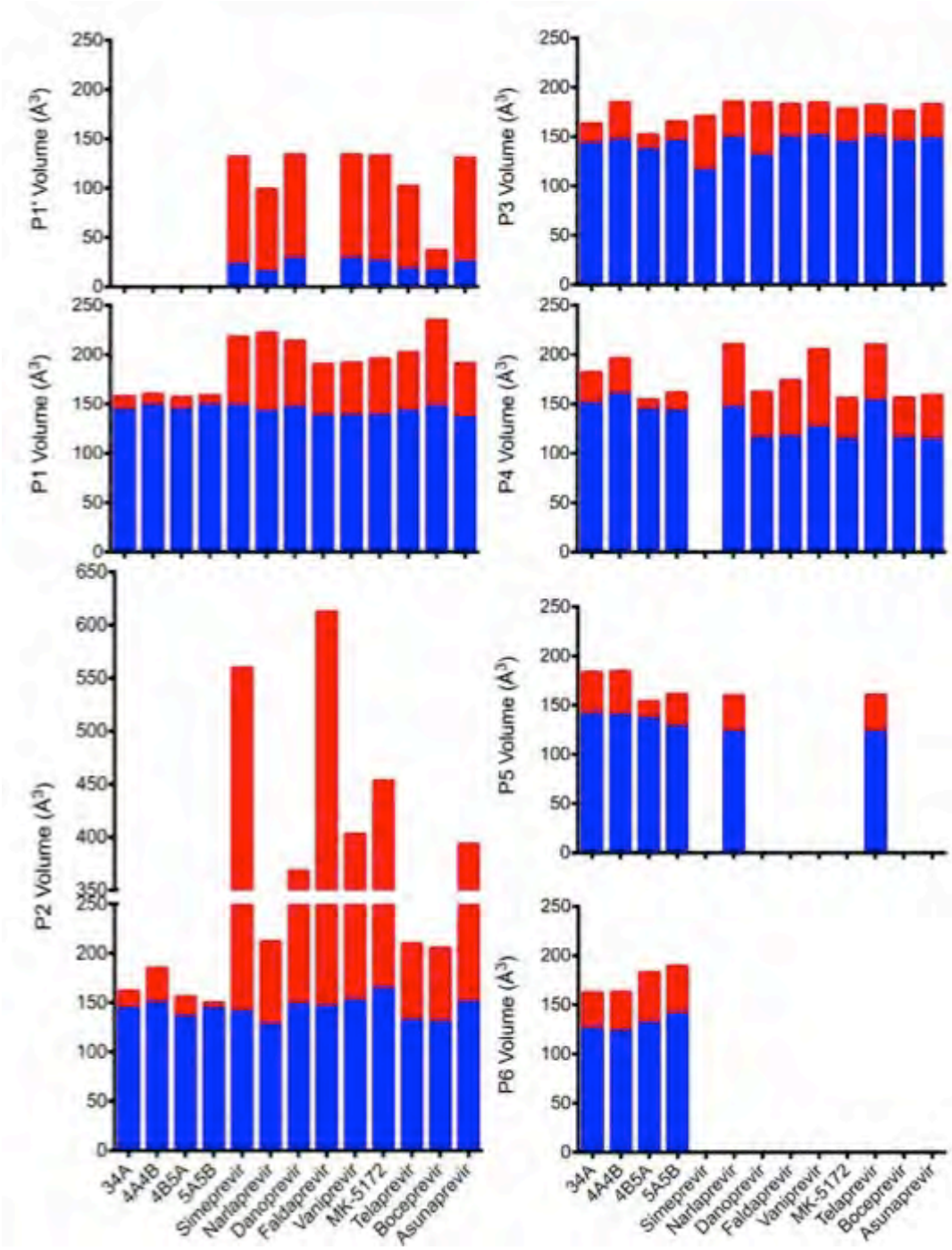


Figure 3.5 Volume of substrates and inhibitors inside (V_{IN}) and outside (V_{OUT}) of the substrate envelope by substrate moiety.

Figure 3.5 Volume of substrates and inhibitors inside (V_{IN}) and outside (V_{OUT}) of the substrate envelope by substrate moiety. V_{IN} is blue and V_{OUT} is in red. If inhibitors did not occupy a particular substrate moiety, V_{IN} and V_{OUT} were not calculated (such as with the P6 moiety). Ligands are listed on the x-axis, and volume is on the y-axis in $Å^3$. MK-5172 is grazoprevir.

3.3.2 Inhibitor-Protease Contacts at the Active Site

Van der Waals contact energies of substrates and inhibitors with the protease were calculated and mapped onto the surface of the active site (**Figures 3.6 and 3.7**). These calculations were also performed for four inhibitors without available crystal structures that were modeled: glecaprevir, paritaprevir, sovaprevir, and vedroprevir (**Figure 3.8**). All surface van der Waals figures were colored on the same energy scale from 0 to a maximum of approximately -7.4 kcal/mol according to the per residue van der Waals contact energies (**Figure 3.9**).

Overall, inhibitors demonstrate different distributions of van der Waals contacts based on the type of moiety that extends from the P2 position and which types of macrocycles are present in inhibitors. These different patterns of van der Waals contacts also correspond to the different patterns of drug resistance that have emerged for these inhibitors. In addition, all inhibitors generally make increased contacts with catalytic residues H57, D80, and S139 compared to substrates, and these interactions are ideal because these residues are evolutionarily conserved and required for biological function, so they are less likely to mutate without compromising catalysis and they are less likely to confer drug resistance.

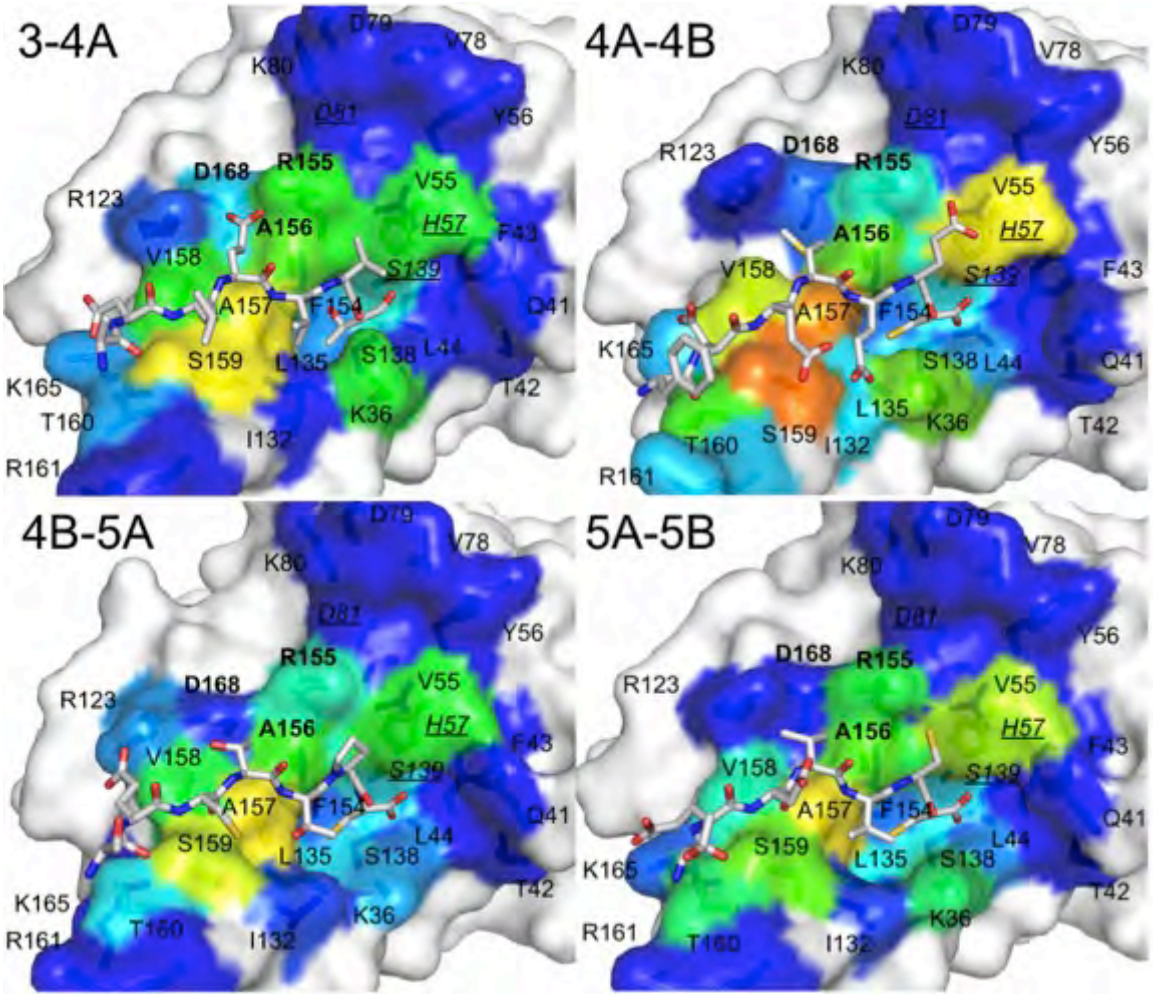


Figure 3.6 Substrate van der Waals contacts mapped on the surface of the NS3/4A protease active site for each substrate.

Figure 3.6 Substrate van der Waals contacts mapped on the surface of the NS3/4A protease active site for each substrate. The protease domain is in gray surface representation, and the contact surface is in rainbow spectrum colors where tangential contacts are in blue and warmer colors indicate stronger van der Waals contact interactions. Substrates are in stick representation with carbon atoms in gray and other atoms in standard CPK colors. Residues in bold are drug resistance residues R155, A156, and D168. Underlined italicized residues are the catalytic residues H57, D81, and S139.

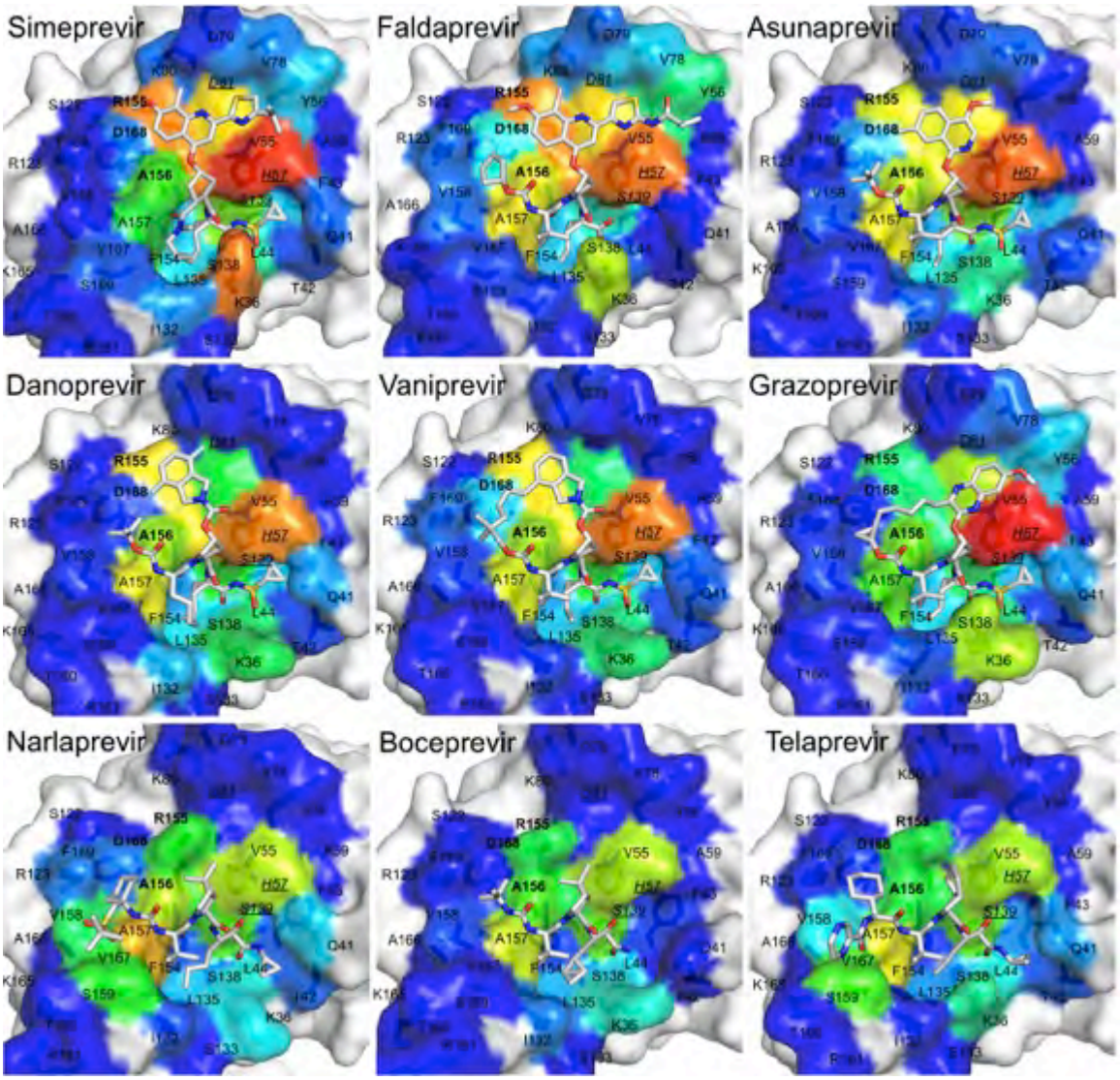


Figure 3.7 Inhibitor van der Waals contacts mapped on the surface of the NS3/4A protease active site.

Figure 3.7 Inhibitor van der Waals contacts mapped on the surface of the NS3/4A protease active site. The protease domain is in gray surface representation, and the contact surface is in rainbow spectrum colors where tangential contacts are in blue and warmer colors indicate stronger van der Waals contacts interactions. Inhibitors are in stick representation with carbon atoms in gray and other atoms in standard CPK colors. Residues in bold are drug resistance residues R155, A156, and D168. Underlined italicized residues are the catalytic residues H57, D81, and S139.

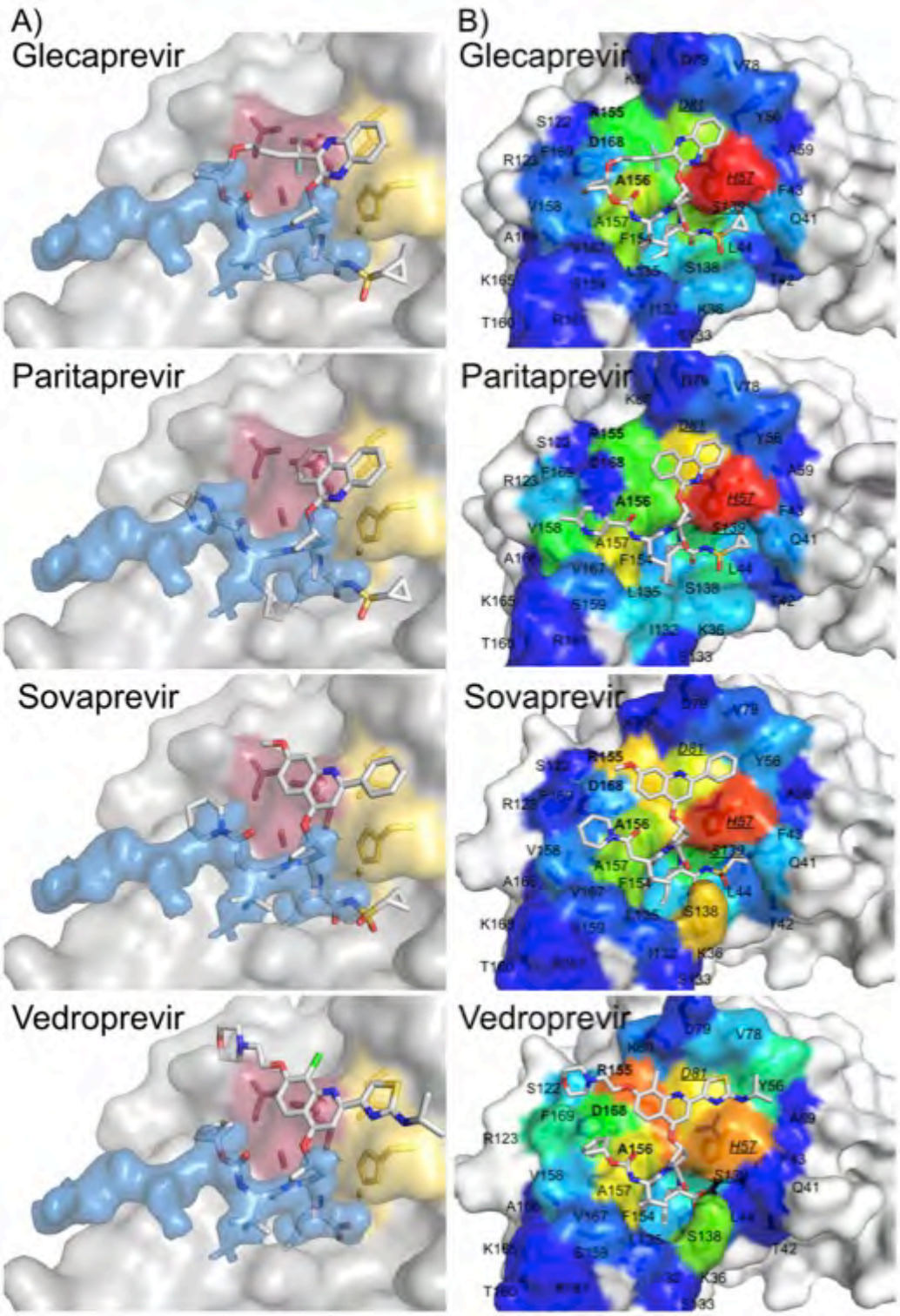


Figure 3.8 Substrate envelope and van der Waals surface representations for modeled HCV NS3/4A protease inhibitors.

Figure 3.8 Substrate envelope and van der Waals surface representations for modeled HCV NS3/4A protease inhibitors. A) The NS3/4A protease domain is in gray surface representation, and the substrate envelope is in blue surface representation. The catalytic triad D81, H57, and S139 is in yellow sticks, and sites of drug resistance residues R155, A156, and D168 are in burgundy sticks. B) Inhibitor van der Waals contacts mapped on the surface of the NS3/4A protease active site. Residues in bold are drug resistance residues R155, A156, and D168. Underlined italicized residues are the catalytic residues H57, D81, and S139.

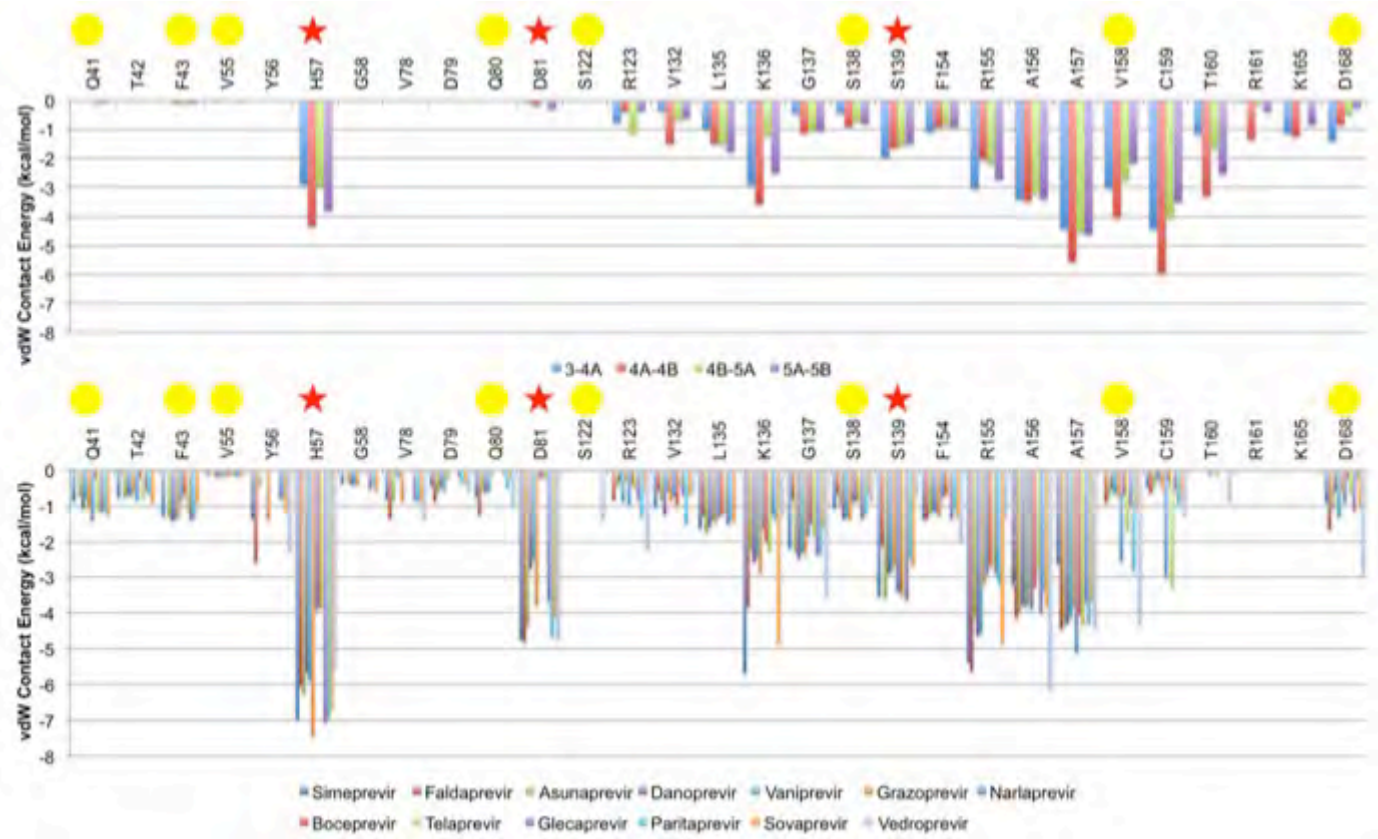


Figure 3.9 Differential van der Waals interactions on the binding surface of HCV NS3/4A protease.

Figure 3.9 Differential van der Waals interactions on the binding surface of HCV NS3/4A protease. (A) Per residue van der Waals contacts interactions for each viral substrate. (B) Per residue van der Waals contacts for inhibitors. Red stars indicate catalytic residues, and yellow circles indicate drug resistance residues.

Grazoprevir has the best activity and also is able to accommodate drug resistance mutations at residues R155 and D168. The quinoxiline group in grazoprevir stacks most favorably on the catalytic triad H57, D80, and S139. Out of all the macrocyclic inhibitors with extended P2 moieties, grazoprevir appears to make the least contacts with R155 and D168. None of the other inhibitors that have been crystallized have a quinoxiline group, but glecaprevir has a quinoxiline group that would also be predicted to stack against the catalytic triad. However, the P2-P4 macrocycle in glecaprevir has two additional fluorine atoms, and the P4 moiety has a five membered ring instead of a three membered ring, so these differences may impact how glecaprevir behaves in the presence of resistance mutations at residues R155 and D168. Paritaprevir has a phenanthridine group in the extended P2 position, and this moiety also seems to stack well on the catalytic triad. Paritaprevir also has a P1-P3 macrocycle instead of a P2-P4 macrocycle, and this difference may allow paritaprevir to better accommodate resistance mutations at A156. It also appears that paritaprevir makes slightly less contacts with A156 compared to glecaprevir and grazoprevir. In addition, compounds without a P2-P4 macrocycle are better able to accommodate resistance mutations at R155 and D168 (46, 176).

Compounds with quinoline, isoquinoline, and isoindoline at the P2 extended position have worse antiviral profiles against main drug resistance residues R155, A156, and D168 compared to compounds like grazoprevir with a quinoxiline at this position. These inhibitors include asunaprevir, danoprevir, and

vaniprevir, and the moieties at the P2 extended position make increased contacts with drug resistance residues R155, A156, and D168 (175, 176). Rather than stacking on the catalytic triad, these moieties stack on these primary drug resistance residues, which increases the susceptibility of these inhibitors to resistance mutations at these locations.

Inhibitors with a P1-P3 macrocycle, such as simeprevir and danoprevir, are slightly more susceptible to resistance mutations at S138, which is located underneath the P1-P3 macrocycle binding site. Drug resistance at S138 has also been reported for simeprevir and danoprevir. Based on the van der Waals analysis, other inhibitors with a P1-P3 macrocycle may also be susceptible to resistance at this location, such as paritaprevir.

Simeprevir and faldaprevir have P2 moieties that extend very broadly into the active site in both the direction of the catalytic triad and also the direction of drug resistance residues. Simeprevir is susceptible to mutations as S122, but faldaprevir is not susceptible to this drug resistance mutation even though both inhibitors are similar at this position. However, faldaprevir has additional contacts in the S4 pocket whereas simeprevir only extends to the S3 pocket. These additional interactions in S4 may help stabilize faldaprevir in the presence of mutations at S122 even though S122 is located distal from the S4 pocket. Sovaprevir is similar to faldaprevir and simeprevir at this location, with increased contacts in the S4 pocket, so it may be able to accommodate resistance mutations at S122. Vedroprevir has the largest P2 moiety of all the inhibitors

analyzed in this study, and it makes many contacts with R155, D168, and S122, so vedroprevir may also be susceptible to drug resistance mutations at these locations.

Boceprevir, telaprevir, simeprevir, danoprevir, asunaprevir, and faldaprevir are susceptible to mutations at Q80. Q80K is a polymorphism that impacts the activity of PIs. The mechanism of resistance as a result of this polymorphism is not clear, especially for inhibitors that do not directly contact this residue, such as boceprevir and telaprevir, but the change in charge at this residue likely impacts the electrostatic network of residues around residue 80. Residue 80 is located at the edge of the active site, and it makes tangential contacts with inhibitors that have a P2-moiety that extends in the direction of this residue, such as simeprevir, danoprevir, asunaprevir, and faldaprevir. Therefore, mutations at residue 80 may have some direct impact on binding of these inhibitors. In addition, sovaprevir and vedroprevir make tangential contacts with residue 80 and may also be susceptible to resistance mutations at this location.

3.3.3 Hydrogen Bond Interactions

Hydrogen bonds were calculated for inhibitors and substrates in complex with the protease for backbone and side chain atoms (**Figures 3.10 and 3.11**). Overall, inhibitors have similar hydrogen bonds as substrates to backbone atoms in the protease. Inhibitors have increased hydrogen bonds to side chain atoms of the conserved catalytic residues H57 and S139 compared to substrates.

Substrates make hydrogen bonds with residues K165 and K136, and hydrogen bonds to these residues are not as common in inhibitors. In a comparison of the crystal structures analyzed in this study, K136 appears to occupy different conformations in the active site. In addition, acidic residues at P6 in the substrate interact with K165, but this interaction has been shown to be highly dynamic (46), so designing inhibitors to maintain these hydrogen bonds may be challenging.

		G137:N	G137:N	S138:N	S139:N	S139:N	R155:O	A157:N	A157:O	A157:O	S159:N	S159:O
Simeprevir	A	3.2	2.8	-	-	3.7	1.9	2.8	-	-	-	-
	B	3.1	2.9	-	-	3.7	2	2.9	-	-	-	-
	C	3.2	3.1	-	-	3.6	2	3	-	-	-	-
	D	3	2.9	-	-	3.7	2.2	3.1	-	-	-	-
Faldaprevir	A	2.9	-	-	-	-	2.9	2.9	3	-	-	-
	B	2.9	-	-	-	-	2.8	2.9	3	-	-	-
Asunaprevir		2.9	3.1	-	-	3.5	2.9	2.9	2.8	-	-	-
Danoprevir		3.1	2.9	3.5	3.5	3.5	3	2.9	2.9	-	-	-
Vaniprevir		3.1	3	3.5	3.5	3.6	3	2.9	3	-	-	-
MK-5172	A	2.9	2.9	3.6	3.5	3.7	3	3.2	3.1	-	-	-
	B	2.7	2.9	3.5	3.3	3.4	2.9	3	3.2	-	-	-
	C	2.9	3.1	3.5	3.2	3.6	2.9	3	2.8	-	-	-
	D	3	3.3	3.5	3.3	3.7	3	3	2.9	-	-	-
Narlaprevir		2.8	-	-	3.3	-	3.2	3.1	2.8	2.8	3.3	-
Telaprevir		2.7	-	-	3	-	3	2.9	3	-	3	3.1
Boceprevir		2.8	-	-	3	-	3	3	3	2.9	-	-
3-4A		-	-	-		-	3.7	3	2.8	-	2.7	2.9
4A-4B		2.9	-	-	3.1	-	3.3	2.9	2.6	-	2.8	3
4B-5A		-	3.1	-	3.2	-	3.1	2.9	2.8	-	2.9	3.1
5A-5B		2.9	-	-	3.1	-	3	2.9	2.8	-	2.9	3.1

Figure 3.10 Hydrogen bond interactions between ligands and backbone atoms in the HCV NS3/4A protease active site.

Figure 3.10 Hydrogen bond interactions between ligands and backbone atoms in the HCV NS3/4A protease active site. Backbone atoms are listed on the top row with the residue name and number and the name of the backbone atom. Backbone atoms listed twice make more than one hydrogen bond with ligand atoms. Crystal structures with more than one ligand-protease complex in the asymmetric unit are listed by chain letter. Distances less than or equal to 2.5 Å are red, distances greater than 2.5 Å and less than or equal to 3.0 Å are yellow, and distances greater than 3.0 Å and less than or equal to 3.5 Å are green. MK-5172 is grazoprevir.

		H57:NE2	R123:N	K136:NZ	K136:NZ	S139:OG	S159:OG	K165:NZ
Simeprevir	A	3.3	-	2.6	3.7	2.7	-	-
	B	3.2	-	2.7	2.8	3.1	-	-
	C	3.2	-	2.6	3.7	2.7	-	-
	D	3.2	-	-	-	2.9	-	-
Faldaprevir	A	2.8	-	3	-	-	-	-
	B	2.8	-	2.9	3.6	-	-	-
Asunaprevir		3.1	-	MS	MS	2.9	-	-
Danoprevir		3.1	-	-	-	-	-	MS
Vaniprevir		3	-	-	-	-	-	MS
MK-5172	A	3	-	3.6	-	-	-	MS
	B	3.3	MS	-	-	-	-	MS
	C	3.2	-	-	-	-	-	MS
	D	3.1	MS	-	-	-	-	MS
Narlaprevir		2.6	-	-	-	-	-	-
Telaprevir		2.6	-	5.1	-	-	2.9	MS
Boceprevir		2.6	-	-	-	-	-	-
3-4A		2.7	3.3	2.5	-	3.9	-	-
4A-4B		2.8	-	2.9	-	-	-	2.7
4B-5A		2.8	2.8	-	-	-	-	-
5A-5B		2.6	-	-	-	-	-	2.8

Figure 3.11 Hydrogen bond interactions between ligands and side chain atoms in the HCV NS3/4A protease active site.

Figure 3.11 Hydrogen bond interactions between ligands and side chain atoms in the HCV NS3/4A protease active site. Side chain atoms are listed on the top row with the residue name and number and the name of the backbone atom. Side chain atoms listed twice make more than one hydrogen bond with ligand atoms. Crystal structures with more than one ligand-protease complex in the asymmetric unit are listed by chain letter. Distances less than or equal to 2.5 Å are red, distances greater than 2.5 Å and less than or equal to 3.0 Å are yellow, and distances greater than 3.0 Å and less than or equal to 3.5 Å are green. MK-5172 is grazoprevir. MS is a missing side chain.

3.4 Discussion

In general, inhibitors with a P2-P4 macrocycle and a quinoxiline-related P2 moiety have greater activity and stack well against the conserved catalytic triad in the active site, but inhibitors with a P1-P3 macrocycle may maintain a better balance between high potency and increased flexibility to accommodate resistance mutations and have a flatter resistance profile. Therefore, an inhibitor with a P1-P3 macrocycle and a flexible quinoxiline-related P2 moiety may be an optimal strategy, and QSAR studies are underway to develop compounds with flexible P2 quinoxilines.

In addition, extending inhibitors in the P1' and P4-P6 regions to make increased contacts with conserved residues in the S1' and S4-S6 pockets rather than extending outside of the S2 pocket may increase inhibitor potency while decreasing susceptibility to drug resistance mutations (46, 95). For instance, simeprevir does not have any contacts in the S4 pocket and is susceptible to the resistance mutation S122 while faldaprevir makes additional contacts in the S4 pocket and is not susceptible to this resistance mutation even though P2 extended moieties in both compounds are similar. Increased conserved active site contacts throughout an inhibitor may allow the inhibitor to better accommodate specific resistance mutations if they emerge. In addition, there are many basic residues in the S6 pocket, such as R119, R123, R161, and K165, which make conserved interactions with the conserved acidic D or E residue at the P6 position of the substrate. Making additional electrostatic interactions with

these residues may also increase the potency of inhibitors but may also be challenging because this region is dynamic (46).

Another interesting strategy for designing a PI with high potency while maintaining a flat resistance profile would be to use a bis-macrocycle with both P1-P3 and a P2-P4 macrocycles. This approach was used to design the inhibitor MK-6325, which was in Phase I clinical trials that have been completed. This inhibitor has better activity than grazoprevir and also has improved potency compared to grazoprevir against drug resistance variants R155K (0.07 vs. 0.013 nM), A156T (5.3 vs. 0.42 nM), and D168Y (0.14 vs. 0.036 nm), so it would be interesting to see how this inhibitor performs compared to other inhibitors in clinical use and development (176-178).

This analysis has shown that differences in how inhibitors fit within the substrate envelope and interact with residues in the active site are correlated with how different patterns of drug resistance have emerged for these inhibitors. Incorporation of the substrate envelope hypothesis into structure based drug design would facilitate the development of robust inhibitors with greater potency and a higher barrier to resistance.

3.5 Methods

3.5.1 Substrate Envelope and V_{IN} and V_{OUT} Calculations

We analyzed crystal structures of NS3/4A protease in complex with inhibitors that have been FDA approved or are in Phase II or III clinical trials.

These structures include the protease in complex with telaprevir (3SV6), boceprevir (5EBQ), narlaprevir (3LON), faldaprevir (3P8N), simeprevir (3KEE), asunaprevir (4WF8), danoprevir (3M5L), vaniprevir (3SU3), and grazoprevir (3SUD) (179-182). The substrate envelope was calculated using crystal structures of NS3/4A protease in complex with substrate cleavage products 4A-4B (3M5M), 4B-5A (3M5N), and 5A-5B (3M5O) (43). The structure of the substrate cleavage product 3-4A in complex with NS3/4A protease was modeled from the full-length crystal structure as previously described (1CU1) (43, 183). The PDB IDs for each of these structures is in parentheses. The models of glecaprevir, paritaprevir, sovaprevir, and vedroprevir were generated based on similar fragments of inhibitors from the existing crystal structures listed, and models were minimized using Maestro in the Schrodinger Software Suite (108). The substrate envelope was generated in PyMOL as previously described (43).

Details of the V_{IN} and V_{OUT} volume calculations have been described previously and were performed using in house Fortran scripts (94). V_{OUT} is the volume of inhibitor that protrudes outside of the substrate envelope, and V_{IN} is the volume of inhibitor inside the substrate envelope. Briefly, to perform these calculations, the active site is divided into a three dimensional grid, and each grid cell is defined by the indices *i,j,k*. Then, an initial value of zero is assigned to each grid cell. The variable $g_{ijk\ 1}$ is the total grid for the inhibitor and $g_{ijk\ 2}$ is the total grid for the substrates. A grid cell value is increased by 1 when a substrate or inhibitor non-hydrogen atom based on the OPLS2005 vdW radius occupies

the grid cell. N_1 and N_2 are equal to the number of static crystal structure calculations for the inhibitor (1) and substrates (4), respectively. The volume of an individual inhibitor outside the substrate envelope, V_{OUT}, is calculated using the following equation, where the resulting sum of grid point values is normalized by the product of N_1 and N_2, and the overall sum is multiplied by the volume of a single grid cell d^3.

$$V_{out} = d^3 \sum_{i,j,k}^{\text{all grid points}} \frac{N_2 - g_{ijk,2}}{N_2} \times \frac{g_{ijk,1}}{N_1} = \frac{d^3}{N_2} \sum_{i,j,k}^{\text{inside}} \left(N_2 - g_{ijk,2}\right) \times \frac{1}{1}$$

The volume of inhibitor inside the substrate envelope, V_{IN}, is calculated using the following equation.

$$V_{in} = d^3 \sum_{i,j,k}^{\text{all grid points}} \frac{g_{ijk,2}}{N_2} \times \frac{g_{ijk,1}}{N_1} = \frac{d^3}{N_2} \sum_{i,j,k}^{\text{inside}} g_{ijk,2} \times \frac{1}{1}$$

3.5.2 Van der Waals Contact Potential Energy

The van der Waals contact potential energies between ligands and the protease were calculated in each structure using a simplified Lennard-Jones potential function defined by the following equation, where r_{ij} is the distance between NA atom *i* and ligand atom *j*, ε is the energy well depth, and σ is the collision diameter:

$$V(r_{ij}) = 4\varepsilon[(\frac{\sigma}{r_{ij}})^{12} - (\frac{\sigma}{r_{ij}})^{6}]$$

Van der Waals contact energies were calculated for all intermolecular atom pairs within a 6 Å cutoff of the binding interface using an in-house Fortran

script. The nonbonded parameters were determined using the OPLS2005 force field (112). Further details of this computation have been described previously (94).

3.5.3 Hydrogen Bond Interactions

Hydrogen bonds were determined using Maestro. A hydrogen bond was defined as having a donor-acceptor distance of a maximum of 3.5 Å and involving only polar atoms nitrogen, oxygen, sulfur, and fluorine. The donor minimum angle was 120 degrees and the acceptor minimum angle was 90 degrees, according to the default settings.

3.5.4 Plots and Figures

Microsoft Excel, Matlab, Prism, and PyMOL were used to create all plots and figures (116, 118, 119).

CHAPTER IV

4 IDENTIFICATION OF INFLUENZA A VIRUS CANDIDATE RESISTANCE MUTATIONS TO BROADLY NEUTRALIZING ANTIBODY

4.1 Abstract

Influenza A virus (IAV), a major cause of morbidity and mortality, is continually evolving in response to selective pressures. Universal influenza vaccines that target multiple strains and stem-directed, broadly neutralizing antibodies are promising therapeutic strategies, but neutralization escape mutants can develop. We used an integrated approach combining viral passaging, deep sequencing methods, computational and biophysical analysis to ascertain the impact of a broadly neutralizing antibody directed against the stem of hemagglutinin (HA) on the generation of potential escape mutants. Human influenza A/Brisbane/59/2007 (H1N1) was grown in Madin-Darby canine kidney cells with escalating concentrations of the broadly neutralizing antibody F10 over serial passages. Sequence analysis revealed a mutation profile in the IAV genome that included three non-synonymous mutations in HA, as well as a distinct mutation in neuraminidase that was previously identified to be associated with antibody escape for A/Brisbane/59/2007. Structural analysis revealed that the HA mutations are located away from the highly conserved antibody epitope and may impact conformational changes important for fusion with the endosome during the viral life cycle, suggesting a novel molecular mechanism of resistance

where mutations away from the binding epitope are selected. Thus, whole genome population sequencing is a powerful method for identifying viral genome responses to antibody selective pressure to identify candidate resistance mutations.

4.2 Introduction

Influenza A virus (IAV) causes a highly contagious acute respiratory illness in humans that is responsible for significant morbidity and mortality. Approximately 30,000 people die from influenza each year in the United States (184). During the past century, three major influenza pandemics resulted in the deaths of 50–100 million people (185). Individuals at the extremes of age and those with certain underlying medical conditions are at particularly high risk for serious complications. IAV evades the body's immune system through small changes in the viral genome that occur continuously over time through antigenic drift. On occasion, *antigenic shift* occurs, in which the segmented genomes of at least two distinct influenza viruses reassort to produce a novel strain to which individuals are antigenically naïve. A pandemic strain can arise following antigenic shift, particularly when the segment encoding hemagglutinin is derived from a non-human influenza virus. Antigenic shift led to the pandemic flu of 1918-1919, which affected 25% of the world population, as well as the recent 2009 H1N1 pandemic flu (186).

IAV's unique evolutionary mechanisms, including high mutation rate, segment reassortment, and shifts between multiple host species, pose significant challenges for controlling the disease and developing effective vaccinations. Therefore, a detailed understanding of influenza virus genome sequence evolution is imperative. The influenza virion consists of eight negative-strand RNA segments which form protein-RNA complexes enveloped in a lipid membrane (187). These eight segments encode at least ten proteins known to be essential for infectivity and replication. Within a given influenza strain, sequence evolution proceeds by mutation, selection, and genetic drift, all of which are affected by the environment, host, and drug treatment. High mutation rates, together with rapid development of influenza epidemics, make tracing the evolutionary history of the virus and discovering the principles governing IAV's evolution complex.

Influenza virus has two surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA). HA binds to sialic acid on host cells, a process that is critical for initial attachment and infection. NA cleaves sialic acid from the host cell membrane during the release of newly formed viral progeny, thus reducing viral affinity for previously infected cells (37). Seventeen different subtypes of influenza A HA (H1–H17) exist, which are divided into two distinct phylogenetic groups, group 1 (H1, H2, H5, H6, H8, H9, H11–H13, H16 and H17) and group 2 (H3, H4, H7, H10, H14 and H15) (188, 189). Only H1, H2, and H3 are present in human influenza strains. HA is synthesized as a single polypeptide (HA0) that is

cleaved by host proteases into HA1 and HA2 subunits. The HA trimer is composed of an HA1 globular head (the receptor binding site) and a stem (or “stalk”) region composed of HA2 and HA1 which has the fusion machinery (see (190) for review). HA is the primary target of the humoral immune response during infection or vaccination. However, influenza vaccines generally elicit strain specific responses, thereby limiting their efficacy and necessitating administration of new vaccines when a novel strain becomes dominant.

Broadly neutralizing antibodies (BnAbs) have conserved epitopes on HA and can neutralize a wide spectrum of influenza viruses (12). BnAbs against viruses such as hepatitis C and HIV have also been discovered (59, 191). In influenza, the BnAb epitopes are involved with receptor binding as well as with the fusion machinery and are functionally conserved and less prone to mutation. BnAbs are potential therapeutic agents when used as passive immunotherapy and can also be integrated into the design of universal vaccines, which could be much more effective than current vaccines. BnAbs against the influenza receptor-binding site include CH65 (192), S139/1 (62, 193), and C05 (62).

In addition, several BnAbs directed against highly conserved epitopes on the stem of HA have been characterized, including F10, C179, CR6261, CR8020, and CR9114 (12, 194). The F10 antibody, derived from the *IGHV1-69* germline by panning immobilized HA using phage-display libraries generated from healthy donors, broadly neutralizes all group 1 viruses (194). An intraperitoneal injection of F10 one hour before challenge with a lethal dose of

H1N1 or H5N1 viruses protected approximately 80-100% of mice. F10 was also administered to mice 48 or 72 hours after inoculation with virus, and viral replication in the lungs significantly decreased. In addition, the A/Vietnam/1203/04 H5N1 strain was passaged three times *in vitro* in the presence of F10 antibody, but no resistant viruses developed at that time (194). Therefore, F10 is an important potential candidate for future therapeutic use. Interestingly, CR9114 and CR6261 are also derived from the *IGHV1-69* germline. C179 was the first antibody discovered to neutralize more than one subtype of influenza (58). CR6261 and F10 neutralize all group 1 viruses, and CR8020 neutralizes all group 2 viruses. Notably, CR9114 neutralizes all group 1 and group 2 viruses, and the CR9114 epitope is conserved across influenza A and B viruses (58). CR9114 neutralized H1-H12 and H14 viruses in a microneutralization assay and protected mice from lethal challenge with certain strains of influenza B viruses, making it one of the most effective broadly neutralizing influenza antibodies identified to date (12, 58).

Stem-directed antibodies prevent fusion of the host and virus membranes in the low pH of the endosome by locking HA in a pre-fusion conformation and preventing the extensive conformational changes in HA required for membrane fusion, blocking entry of viral RNA into the infected cell. Despite the high conservation of BnAb epitopes in the HA stem region, neutralization escape mutations by stem-directed BnAbs have occurred in and around these epitopes (12, 58, 61, 63, 66, 67). Many of these mutations cause neutralization escape by

directly reducing antibody binding affinity, but additional escape mechanisms that impact HA function or viral fitness may also emerge. Understanding mechanisms of escape and identifying novel escape mechanisms are critical steps for evaluating BnAbs that may be incorporated into future therapies and vaccines.

The goal of this study was to identify potential IAV escape mutants for the BnAb F10 through high throughput sequencing (HTS) analysis of samples generated through *in vitro* trajectory experiments. Mutations in the influenza genome have been structurally characterized to determine possible mechanisms for resistance. In HA, no escape mutations emerged in the highly conserved antibody binding epitope region, but potential resistance mutations that may impact fusion were selected, suggesting a possible novel mechanism for resistance. By screening candidate BnAbs through such new methods, we may be able to identify which BnAbs are least likely to generate escape mutants and thereby be most effective in the long term.

4.3 Results

4.3.1 Serial passage of influenza in the presence of F10 monoclonal antibody

We tested F10, which broadly neutralizes all group 1 influenza HAs (194), against influenza A/Brisbane/59/2007 (H1N1) virus in MDCK cells in our experimental trajectories (**Figure 4.1**). F10 is a broad-spectrum antibody directed against the stem of influenza HA (194) that was originally identified by panning immobilized HA using phage-display libraries. Rather than blocking cell

attachment, this antibody binds to a highly conserved pocket in the stem region of HA that contains the fusion peptide so that membrane fusion of IAV is blocked. We passaged influenza virus A/Brisbane/59/2007 (H1N1) under escalating doses of F10 antibody, starting at 1X the ED_{50}, or 0.3 μg/mL, at passage 4 and escalating to ≥5 μg/mL in MDCK cells to select for a resistant virus population in two independent trajectories, designated Experiments 1 & 2. Each experiment included a complete no-antibody control arm. In Experiment 2, we included an additional control that included escalating amounts of irrelevant monoclonal antibody 80R specific to severe acute respiratory syndrome coronavirus (195) (**Figure 4.1A**). The variation in output viral titers over time is displayed in **Figure 4.1B**.

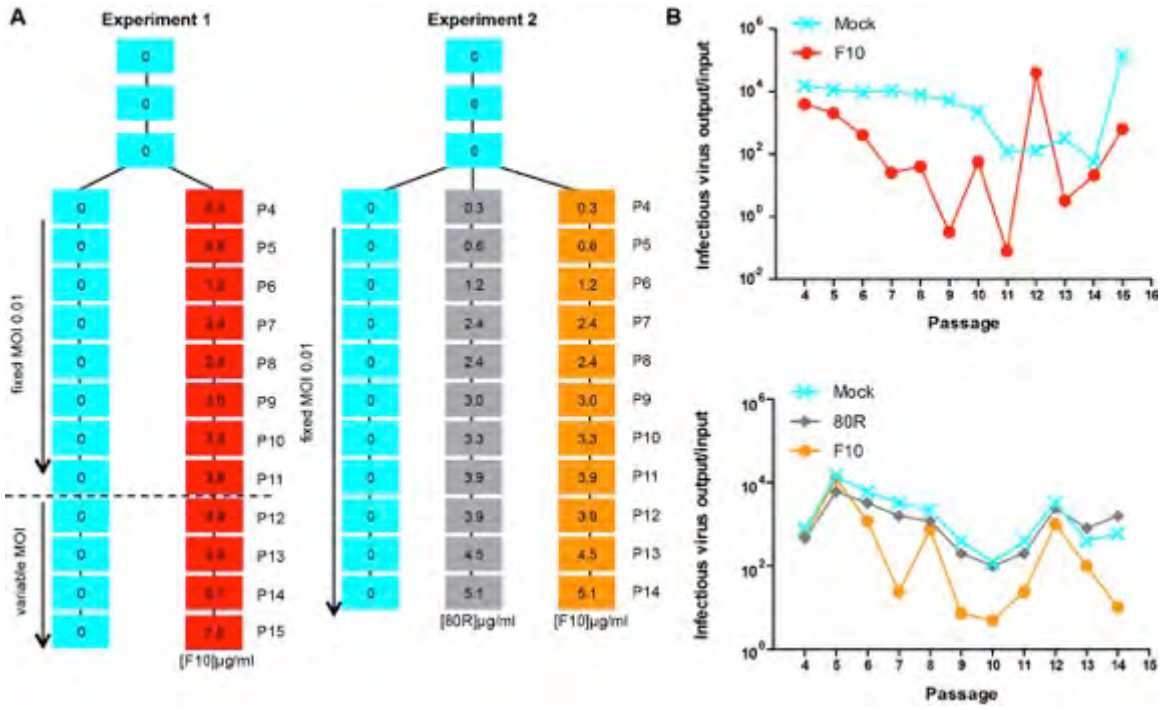


Figure 4.1 Experimental design and viral titers for F10 trajectories.

Figure 4.1 Experimental design and viral titers for F10 trajectories. (A) Schematic of Experiment 1 and Experiment 2 for F10 trajectories. Cyan boxes indicate virus that was passaged without the presence of antibody, with the top three passages as P1, P2, and P3, and additional passages as labeled. Red boxes indicate virus that was passaged in the presence of F10 broadly neutralizing antibody (Experiment 1). Orange boxes indicate virus that was passaged in the presence of F10 broadly neutralizing antibody (Experiment 2). Grey boxes indicate virus that was passaged in the presence of 80R control antibody (Experiment 2). (B) Ratios of viral titers (output/input) plotted against passage number. Experiment 1, upper panel. Experiment 2, lower panel.

4.3.2 Sequence analysis reveals candidate F10 escape mutations

All viral samples were processed for HTS as previously described (196-198). Analysis of HTS data from Experiments 1 & 2 using the Wright-Fisher ABC (WFABC) model identified viral mutations with a 99% posterior probability of being under positive selection (**Table 4.1**) (196, 199). These candidate F10 escape mutations included three non-synonymous mutations in segment 4 (HA): N203V, N460S, and S123G (H1 numbering system) and one non-synonymous mutation in segment 6 (NA), E329K, which was previously identified as important for antigenic drift (200). In addition, one non-synonymous mutation in segment 2 (PB1), A643T, one non-synonymous mutation in segment 3 (PA), L28P, and two synonymous mutations in segments 4 and 5 were elicited. The allele frequencies increased as a function of passage number, and none of these mutations was elicited with the irrelevant control 80R antibody (**Figure 4.2A)**. Selection coefficients are shown in **Figure 4.2B**. Of note, segment 4 mutations A638G and A639T generate a double mutant in perfect linkage to encode the HA N203V amino acid substitution.

4.3.3 Structural mapping of candidate F10 hemagglutinin escape mutants

To further investigate the mutations in HA and identify possible structural changes and escape mechanisms, we mapped the non-synonymous mutations N203V, N460S, and S123G onto existing crystal structures. We leveraged the knowledge on HA structure, conformational changes in HA that occur during

fusion, and residues that are important for HA receptor binding. Notably, all three mutations are located away from the F10 binding epitope. Hence, instead of directly affecting F10 antibody binding, these distal mutations may impact the function of HA by conformational changes important for fusion with the endosome during the viral life cycle and by modulating receptor binding affinity (**Figure 4.3**). Such detailed analysis of the location of these mutations in the HA structure enabled us to generate hypotheses on how these mutations may impact conformational changes that occur in both subunits of HA, HA1 and HA2.

4.3.4 Potential impact of specific mutations on hemagglutinin function

Residue 203 is located at the receptor binding site of HA and forms a hydrogen bond with the human receptor analog LSTc in H2 HA (**Figure 4.4)** (201). In the H3 subtype crystal structures, this residue also interacts with sialic acids through hydrogen bonds (202, 203). In influenza A/Brisbane/59/2007 (H1N1) HA, the selected mutation N203V could result in the loss of this hydrogen bond between the receptor and its binding site because the side chain changes from a carboxamide to a nonpolar group. Residue N203 (or N190 by the H3 numbering system) has been implicated in conferring receptor binding specificity (204), and mutations modulating receptor affinity are a known mechanism of antibody escape. Interestingly, this residue is one of the 17 contact residues for the BnAb C05 and is 99% conserved in human H1, H2, and H3 (62).

Table 4.1 Top-ranking polymorphic sites (posterior probability s>0>99%)

Experiment 1	**Segment**	**Protein**	**Nucleotide change**	**Amino acid change**
	3	PA	T106C	L28P
	4	HA	A398G	S123G
	4	HA	G1147A	Synonymous
	4	HA	A1410G	N460S
	5	NP	T1148C	Synonymous
	6	NA	G1004A	E329K
Experiment 2	**Segment**	**Protein**	**Nucleotide change**	**Amino acid change**
	2	PB1	G1950A	A643T
	4	HA	A638G	N203V
	4	HA	A639T	
	6	NA	G1004A	E329K

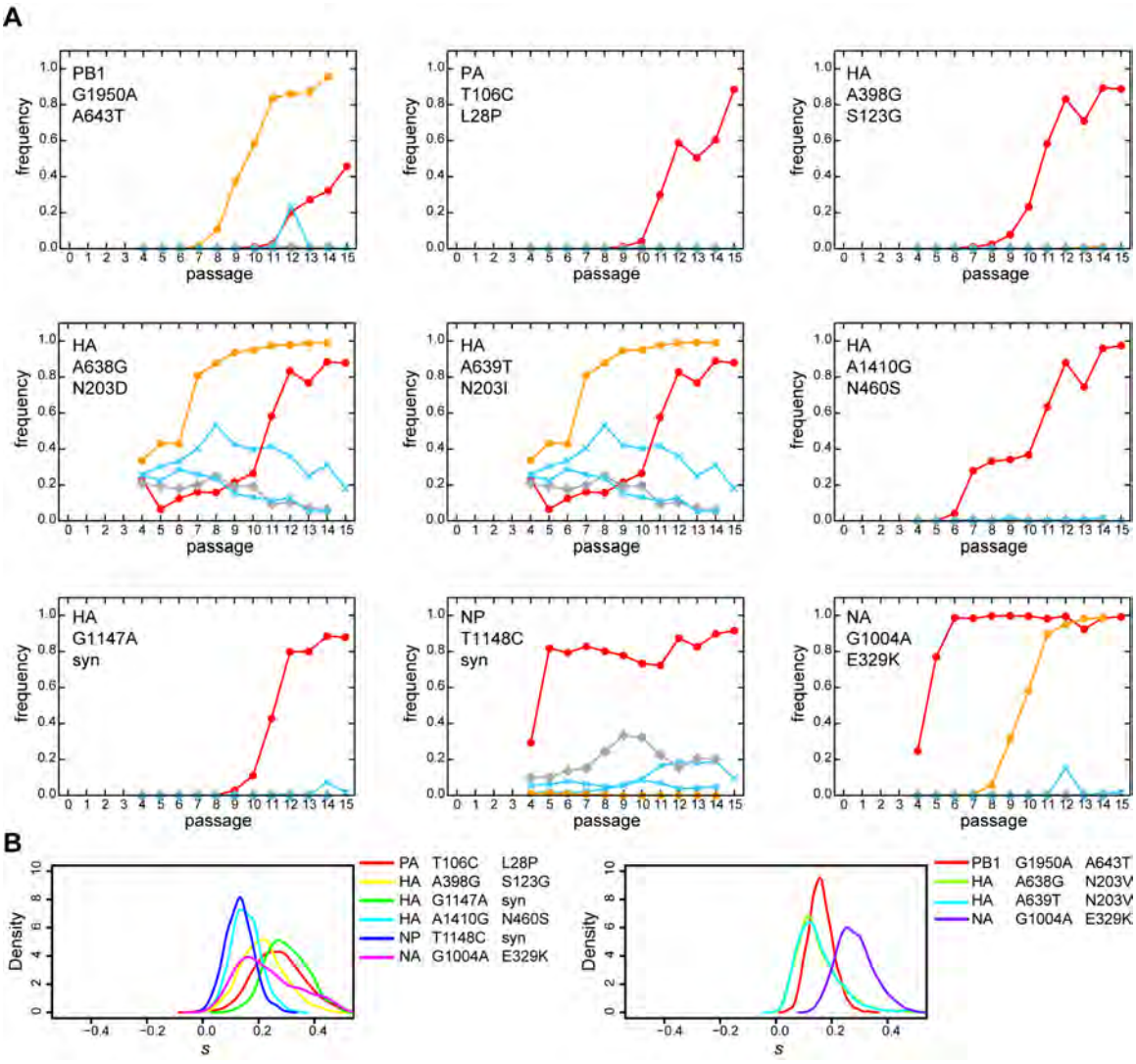


Figure 4.2 Significant mutations arising in influenza A virus under selection with the broadly neutralizing antibody F10.

Figure 4.2 Significant mutations arising in influenza A virus under selection with the broadly neutralizing antibody F10. (A) Trajectories of significant mutations elicited by viral passaging with F10, with 80R control antibody, or without antibody in terms of allele frequency. Red = F10, Experiment 1; orange = F10, Experiment 2; grey = 80R control; and cyan = no antibody control. *Mutations individually marked as N203D and N203I are in perfect linkage and yield N203V, as the wild type sequence is GGT AAC CAA (AAC = positions 638/639/640), protein: GNQ. The mutant sequence is GGT GTC CAA (GTC = positions 638/639/640), protein: GVQ (see second row, columns 1 and 2). (B) The posterior probability (*s*) distribution of selection coefficients for the mutations. Left panel, Experiment 1. Right panel, Experiment 2. Specific mutations are listed by influenza viral protein, nucleotide change, and amino acid change. Syn = synonymous.

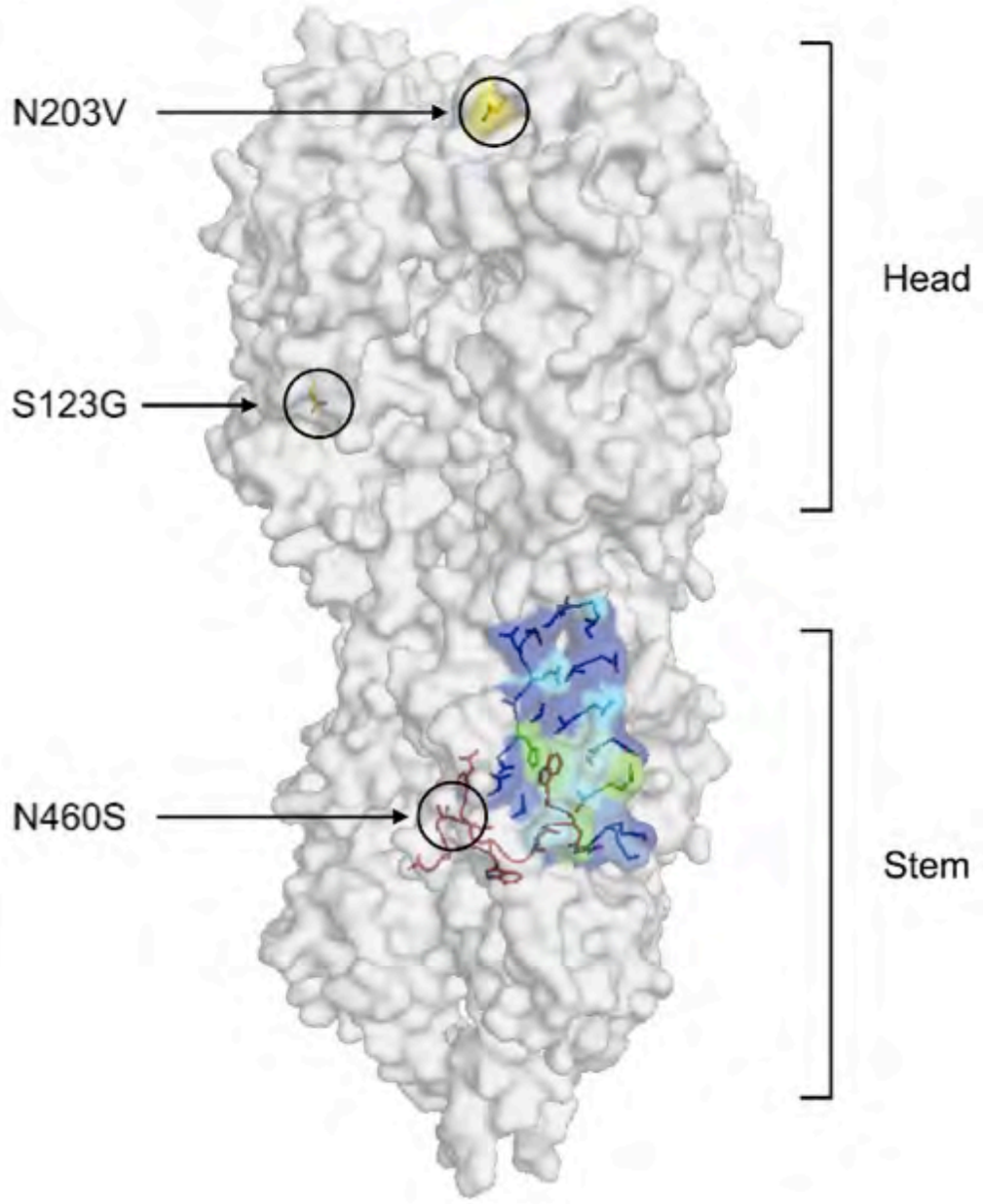


Figure 4.3 Candidate escape mutations identified in the F10 trajectories are mapped onto the structure of HA.

Figure 4.3 Candidate resistance mutations identified in the F10 trajectories are mapped onto the structure of HA. The F10 epitope on the HA stem, in surface and stick representation, is colored in blue, cyan, and green (PDB ID: 3FKU). Van der Waals contacts between HA and F10 are colored on the surface of HA using a rainbow color spectrum, where residues with the greatest van der Waals contacts are in green, intermediate contacts are in cyan, and smallest contacts are in navy blue. The fusion peptide is in red stick representation to the left of the F10 epitope, and candidate escape mutations are labeled.

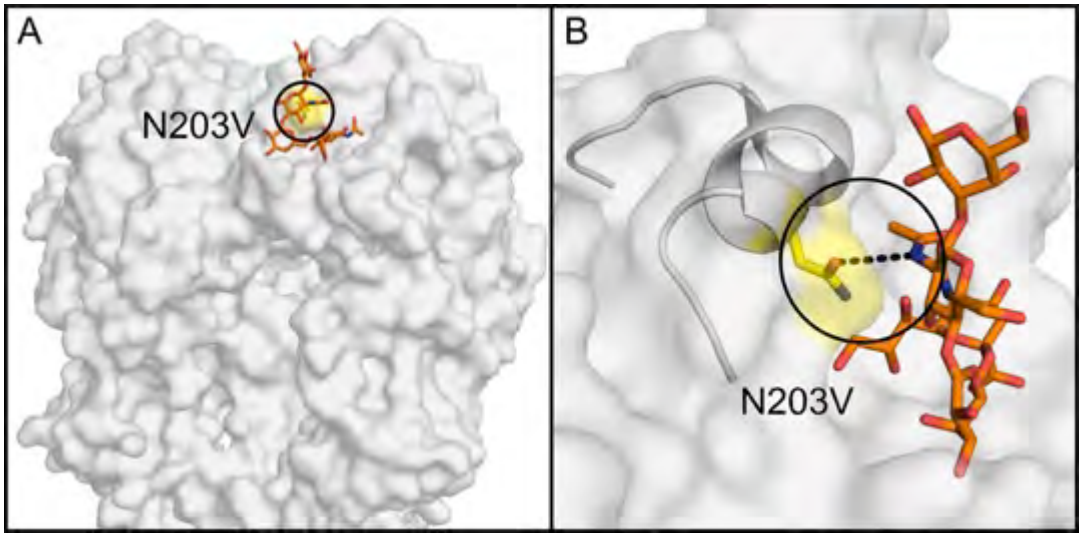


Figure 4.4 The N203V mutation in HA is located in the HA receptor binding site.

Figure 4.4 The N203V mutation in HA is located in the HA receptor binding site. (A) The head of HA is represented by a grey surface, and the location of the mutation N203V is labeled with a circle and a yellow surface. The human receptor analog LSTc is shown in gold sticks (PDB ID: 2WRG). (B) N203V is located in the HA receptor binding site. In the crystal structure, this residue forms a hydrogen bond with the human receptor analog LSTc. The hydrogen bond is shown with a black dashed line connecting the side chain oxygen atom of N203 with a nitrogen atom on LSTc.

The other two non-synonymous HA mutations selected by the F10 antibody, N460S and S123G, are located at key positions that may influence the conformational changes needed to facilitate membrane fusion. Influenza RNA enters the host cell when the viral envelope and the endosomal membrane fuse. The N-terminal fragment of the HA2 subunit, the fusion peptide, mediates fusion. At neutral pH, the fusion peptide is buried in a negatively charged pocket in the stem of HA, but at acidic pH, the fusion peptide dissociates from the HA stem and inserts into the endosomal membrane to promote fusion between the viral membrane and the endosomal membrane (190, 205, 206). Residue N460 hydrogen bonds with the backbone nitrogen of a glycine on the fusion peptide at neutral pH, and the mutation N460S results in the loss of this stabilizing hydrogen bond. Daniels *et al.* reported that mutations that destabilize the neutral pH conformation of the fusion peptide can alter the conformational change in HA that occurs at acidic pH (207). Furthermore, residue 460 hydrogen bonds with glycine 4 (H3 numbering) on the N-terminus of the fusion peptide, and substitution of a different residue that also results in the loss of a stabilizing hydrogen bond with this glycine has been reported to impact the pH of membrane fusion (208). The mutation N460S may also alter the F10 epitope on the HA stem because residues 17-21 of the fusion peptide form the center of the F10 epitope. Residue N460 is located adjacent to the fusion peptide in HA, forming a hydrogen bond (**Figure 4.5).** The fusion peptide is critical for the

conformational changes in HA that occur at acidic pH, where the fusion peptide dissociates from the stem of HA and inserts into the endosomal membrane.

The crystal structures of an early fusion intermediate of HA at neutral and acidic pH show the conformational changes in HA1 that occur during early fusion (**Figure 4.6**) (209). These structures reveal that residue S123 is located in a hinge region of the HA1 subunit. HA1 acts as a clamp on HA2 and stabilizes the metastable pre-fusion state of HA (209). To remove this clamp during fusion, HA1 undergoes conformational changes, and one of these conformational changes occurs at a hinge region around S123. In this region, an alpha helix that is present at neutral pH begins to unfold at acidic pH, causing additional conformational changes in the adjacent antiparallel beta-sheet that connects to the receptor binding subdomain (209). The mutation S123G introduces a flexible glycine residue into this hinge region, which may promote the unfolding of this alpha helix and consequently promote the early conformational changes in HA1 during membrane fusion.

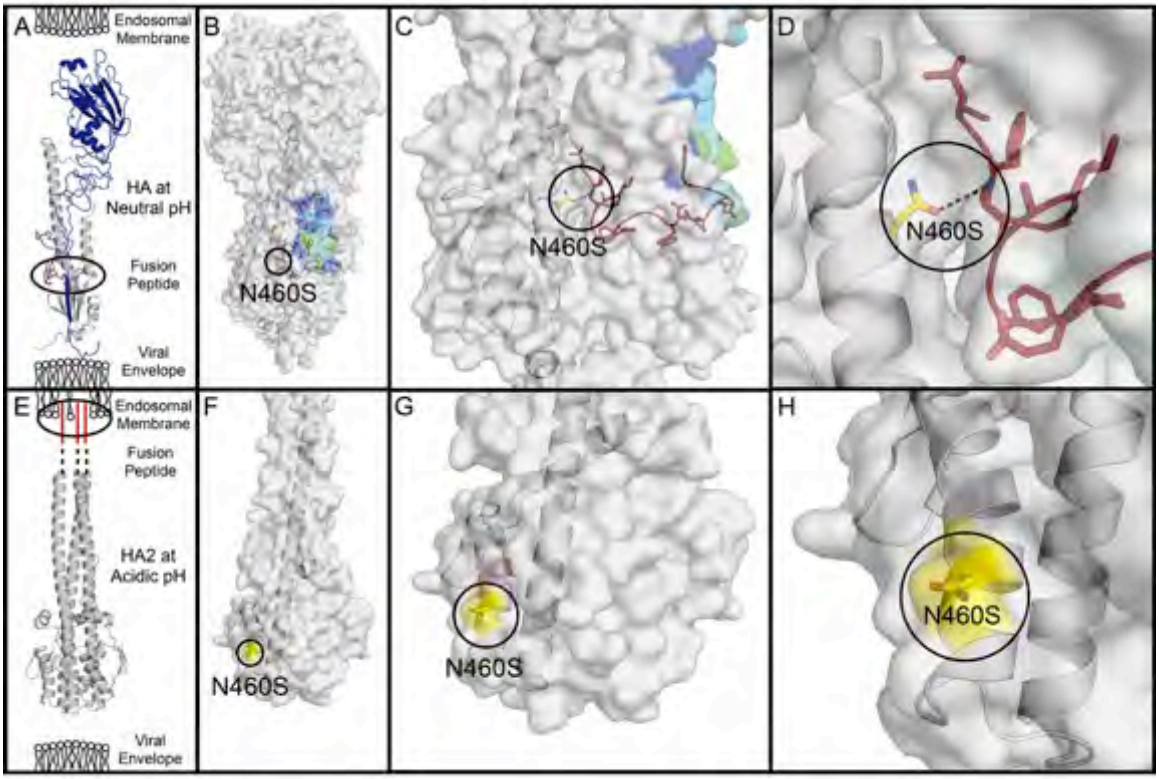


Figure 4.5 The N460S mutation in HA is located adjacent to the fusion peptide in HA.

Figure 4.5 The N460S mutation in HA is located adjacent to the fusion peptide in HA. (A) The structure of the HA monomer at neutral pH is shown with respect to the viral envelope and endosomal membrane (PDB ID: 3FKU). The HA1 subunit, which forms the head of HA, is shown in blue, and the HA2 subunit, which forms the stem of HA, is shown in grey. The fusion peptide is shown in red. The fusion peptide is important for the conformational change that occurs in HA2 at acidic pH. (B) The location of the mutation N460S is circled on the structure of HA at neutral pH with the epitope colored by van der Waals contacts and the fusion peptide shown in red (PDB ID: 3FKU). (C) and (D) The stem is shown in more detail, and the hydrogen bond between residue N460 and the fusion peptide is shown with a black dashed line. (E) At acidic pH, the fusion peptide dissociates from the stem of HA and inserts into the endosomal membrane (PDB ID: 1HTM). (F) The structure of HA2 at acidic pH is shown, and residue N460 is exposed to the surface and is in yellow (PDB ID: 1HTM). (G) and (H) show this residue in more detail.

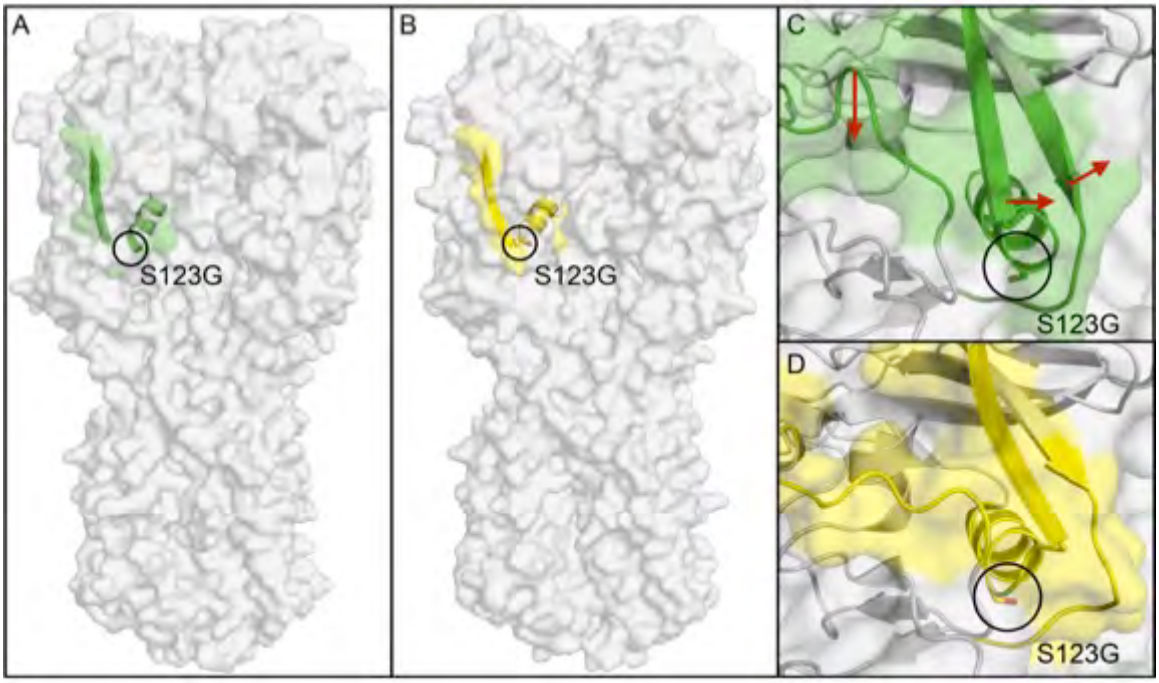


Figure 4.6 S123G is located in a hinge region of conformational change in an early fusion intermediate of HA1.

Figure 4.6 S123G is located in a hinge region of conformational change in an early fusion intermediate of HA1. (A) The structure of HA at neutral pH is shown with residue S123 circled and surrounding residues 115-129 shown in green (PDB ID: 3QQB). (B) The structure of an early fusion intermediate of HA at acidic pH with residue S123 circled and surrounding residues 115-129 shown in yellow (PDB ID: 3QQO). (C) A detailed view of S123 at neutral pH with surrounding residues colored in green to shown the early conformational changes that occur in HA1 during fusion. S123 is located in a hinge region of conformational change in HA1, and the direction of the conformational changes that occur at acidic pH is shown with red arrows (PDB ID: 3QQB). (D) The resulting structure of the early fusion intermediate of HA is shown in yellow (PDB ID: 3QQO).

4.3.5 Neuraminidase mutant

In addition to mutations in HA, the development of NA E329K under selection with F10, but not control antibody 80R, suggests an epistatic adaptation of the virus to circumvent the broadly neutralizing antibody as an escape mechanism. The E329K mutation in NA of influenza A/Brisbane/59/2007 was described in the antigenic evolution of proteins in H1N1 viruses used in vaccine formulations during the last 15 years through analysis of inhibition titers and antigenic cartography (200). This single point mutation was found to be primarily responsible for the lack of inhibition by polyclonal antibodies specific for an earlier influenza vaccine antigen, impacting NA drift. Although antigenic change and drift in NA is often due to antibody selection, antigenic change in NA may also result from a functional change in HA (210). The E329K change is located on a loop on the surface of NA, opposite from the tetramer interface, and the glutamic acid to lysine substitution suggests that a change in a charged interaction is involved in the development of this mutation (67). Amino acid changes at residue 329 were also reported in earlier studies of NA in response to selection with monoclonal antibodies (211-213).

4.3.6 PA, PB1, and synonymous mutants

One non-synonymous PB1 mutation significantly emerged in Experiment 2 and one non-synonymous PA mutation emerged in Experiment 1, as well as two synonymous mutations in HA and NP in Experiment 1 only. The potential

relevance of these mutations in the presence of F10 as a selective pressure needs further evaluation. Synonymous mutations can affect viral fitness, such as with RNA structure, optimum codon usage, translation efficiency, and mutational robustness (214-216).

4.4 Discussion

We identified and characterized several candidate F10 escape mutations for a vaccine strain of influenza virus, A/Brisbane/59/2007, by combining viral passaging and HTS methods with computational and structural analysis. We previously applied such methods to understand the temporal evolution of oseltamivir resistance (197) and viral reassortment (198). Using a similar approach, we defined three novel mutations in influenza HA that may impact viral binding or fusion. In addition, our data highlight a mutation in NA (E329K) previously identified to drive antigenic drift for this particular viral strain, further validating our approach (200).

The three novel mutations in HA that we identified are located in regions of HA that modulate receptor binding specificity or the fusion pH of HA (217). Mutations at residue 203 (190 in H3 numbering) were reported to impact receptor specificity for substrates with an α-2,3 or α-2,6 glycosidic linkage between the terminal sialic acid and the adjacent carbohydrate (200, 218-221). For instance, the mutation E190D in combination with G225D (H3 numbering) in H1 HA increases specificity for α-2,6 linked sialic acids and reduces affinity for α-2,3

linked sialic acids (221). Such mutations that alter receptor binding affinity are selected in the response to other neutralizing antibodies as well (222, 223).

In addition, both mutations at residue 460 and 123 are located in regions of HA that modulate the pH of fusion (217). These regions include the HA1-HA1 interface, the region in the stem surrounding the fusion peptide, and the HA1-HA2 interface that includes the 110-helix (H3 numbering). Residue 460 (residue 117 of HA2 in H3 numbering) is located in the stem region surrounding the fusion peptide, and mutations at residues 111, 112, and 114 (H3 numbering) increase the pH of fusion in H3, H5, and H7 subtypes (201, 217, 224, 225). Many other mutations in the fusion peptide or the surrounding pocket have also been shown to significantly affect the fusion activity of HA or the pH of membrane fusion (201, 217, 224-227). Residue 123 (residue 113 in H3 numbering) is located in the 110-helix, which is involved in the reorganization of the HA1-HA2 interface that occurs during membrane fusion, and mutations at residues 104, 110, and 115 (in H3 numbering) also impact the pH of fusion due to changes at the HA1-HA2 interface (217, 228). Similarly, the three candidate F10 escape mutations discovered in this study may also play an important role in modulating receptor binding specificity and the pH of membrane fusion.

The stem region of HA is highly conserved, especially in the region of the F10 epitope (**Figure 4.7**), suggesting an important role in membrane fusion (60). Despite this conservation, previously reported mutations in HA which cause neutralization escape by stem-directed BnAbs have occurred in regions in and

around the BnAb epitopes. For BnAb CR8043, the two escape mutations reported, R25M and Q34R on HA2, are located on the HA stem near the center of the CR8043 epitope (63). BnAb CR6261 neutralizes H1, H2, H5, H6, H8, and H9 viruses, and after ten in vitro passages, an escape variant of H5N1 was generated with a H111L mutation in HA2 (12, 66, 67). This residue is also buried under the fusion peptide in the HA stem and is located under the CR6261 epitope. The BnAb C179 was the first BnAb reported to neutralize more than one influenza subtype, and two escape mutations were also selected in the HA stem region, T318K in HA1 and V52E in HA2, which are both extremely rare variants (58). In addition, escape mutations against the BnAb CR8020, such as D19N and G33E in HA2, also occur in the CR8020 epitope and are relatively rare variants, especially in human isolates (61). Interestingly, the three candidate F10 escape mutations that appeared in this study, N203V, N460S, and S123G, also occur at low levels in the alignment of the group 1 HA sequences from the NCBI database, at 2.4%, 1.1%, and 6.8%, respectively (229, 230). In addition, the wild-type residues N460 and S123 are both highly conserved in group 1 HA, at 98% and 91%, respectively.

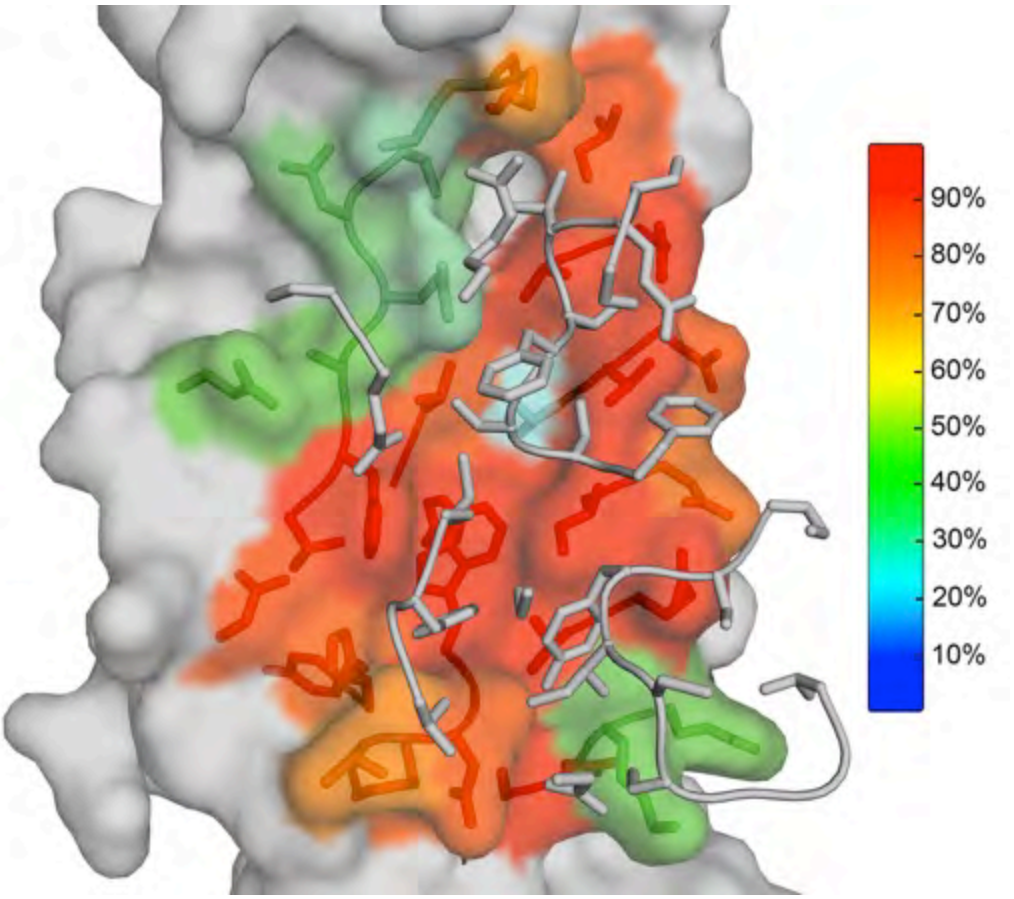


Figure 4.7 The F10 epitope on the stem region of HA is highly conserved.

Figure 4.7 The F10 epitope on the stem region of HA is highly conserved. The F10 epitope on the HA stem is colored based on the percentage of sequence conservation in Group 1 HA. Colored sticks show the residues that form the F10 epitope, and grey sticks show the residues of the F10 antibody that make direct van der Waals contact with the F10 epitope. The percentage of sequence conservation on the epitope is indicated by the color bar using a rainbow color spectrum, where red indicates 90% sequence conservation, yellow indicates 60% sequence conservation, green indicates 40% sequence conservation, and so on. PDB ID: 3FKU.

We conducted these experimental trajectories using influenza A/Brisbane/59/2007 to characterize the effect of a BnAb on a contemporary vaccine strain. However, thus far we have not been able to maintain a continuously productive recombinant A/Brisbane/59/2007 virus using a plasmid-based system. Performing similar trajectories with BnAb on a virus such as pandemic influenza A/California/7/2009 (H1N1), followed by expression of candidate escape mutants (using reverse genetic systems) and monitoring for alterations in antibody ED_{50} measurements, would help identify which mutations are the most daunting for resistance in circulating influenza strains (231, 232). This will help define the impact of the polymerase mutations and perhaps epigenetic effects of the synonymous mutations.

4.5 Conclusion

An in-depth understanding of genome-wide effects of BnAbs on IAV will yield insights on which “universal” influenza vaccines may be the most effective and least likely to induce escape mutants. Furthermore, additive and synergistic effects of antiviral drugs and BnAb can be monitored to define and quantify the impact of multiple selective pressures on the evolution of resistance over time. Given that these will be “real world” pressures faced by IAV, these combination studies may assist in determining which combinations may serve as optimal strategies in future epidemics and pandemics.

4.6 Methods

4.6.1 Cells, virus stocks, and chemicals.

Madin-Darby canine kidney (MDCK) cells were obtained from American Type Culture Collection (Manassas, VA) and propagated in Eagle's minimal essential medium (MEM) with 10% fetal bovine serum (FBS; Hyclone, Logan, UT) and 2 mM penicillin/streptomycin. Influenza virus A/Brisbane/59/2007 (H1N1), grown in chicken egg allantoic fluid, was obtained through the NIH Biodefense and Emerging Infections Research Resources Repository, NIAID, NIH (NR-12282; lot 58550257) and passaged three times in MDCK cells (passages 1–3).

4.6.2 Viral titer determination by plaque assay

Viruses were quantified on MDCK cells to determine infectious titer (plaque forming units per mL, or PFU/mL) as previously described (233). In brief, six 10-fold serial dilutions were performed on the viral samples followed by 1 h of binding at 37°C on confluent MDCK cells in 12-well plates. After washing off unbound virus with phosphate buffered saline (PBS), the cells were overlaid with agar (0.5%) in DMEM-F12 supplemented with penicillin/streptomycin, L-glutamine, bovine serum albumin, HEPES, sodium bicarbonate, and 20 μg/mL acetylated trypsin (Sigma, St. Louis, MO). After the agar solidified, the plates were incubated for ~48 h at 37 °C. Cells were fixed and stained with primary antibody anti-H1 (MAB8261, Millipore, Billerica, MA). Plaques were visualized

with anti-mouse horseradish peroxidase-conjugated secondary antibody (BD Biosciences, San Jose, CA) and developed with peroxidase substrate kit (Vector Laboratories, Burlingame, CA).

4.6.3 Determination of the ED_{50} for F10 antibody

The 50% effective dose (ED_{50}) value was defined as the concentration of antibody that reduced plaque number to 50% of no drug control. In brief, the ED_{50} was determined by seeding 2.5 x 10^5 MDCK cells/well in a 24-well plate and incubated overnight at 37 °C, 5% CO_2. Virus was added to cells at a multiplicity of infection (MOI) of 0.01 in 100 µL of influenza virus growth medium [EMEM/10% FBS with 2 mM penicillin/streptomycin, 7.5% bovine serum albumin, and 1 µg/mL TPCK-treated-trypsin (Sigma)] plus serial dilutions of F10 antibody. After incubation at 37 °C for 1 h, cells were washed once with PBS; 500 µL of influenza virus growth medium with the appropriate concentration of antibody was added and cells were again incubated at 37 °C for several days. Supernatants were collected when >90% cytopathic effect (CPE) was achieved for at least one antibody concentration. Supernatants were centrifuged for 15 min at 300 × *g* at 4 °C and stored at -80 °C. The viral titer for each sample was determined by plaque assay.

4.6.4 Viral culture

Viruses were serially passaged in MDCK cells (2.5 x 10^5 cells/well). The MOI for passages was 0.01 except for late passages in the first experiment, for

which output virus was briefly low and MOI was adjusted to accommodate. Trajectories were prepared both in the presence and absence of escalating doses of F10 antibody or equivalent amounts of the control monoclonal antibody 80R. In passage 4, the antibody concentration was 1X the ED_{50}. For the next passage, the concentration was increased to 4X the ED_{50}, and then doubled for each subsequent passage as long as >50% CPE was present. If <50% CPE was present, the dose of antibody was escalated at a slower rate.

4.6.5 High-throughput sequencing

We developed a high-throughput sample processing workflow, carried out in 96-well format, including RNA purification, reverse transcription, whole genome PCR, followed by DNA barcoding and library preparation, as previously described (197). Libraries were sequenced on the Illumina HiSeq2000 platform to generate 100 nucleotide reads.

4.6.6 Bioinformatics analysis

An integrated bioinformatics pipeline was developed to trim and bin the raw read data based on barcode, align reads to the reference IAV genome, and quantify the level of nucleotide and amino acid variability within the viral population, as previously described (196, 197). To streamline the processing of large numbers of IAV samples, an SQL database with a web interface was developed, integrating sample growth conditions with DNA barcoding information. The database was directly accessed using the analysis pipeline,

eliminating the potential of human error when correlating experimental conditions with large scale IAV genomic data.

Short reads from the Illumina platform were filtered for quality scores >20 throughout the read and aligned to the strain's reference genome using BLAST. Over 95% of the selected reads could be mapped to the IAV reference genome obtained from GenBank (accessions CY030232, CY031391, CY058484-CY058486, CY058488- CY058489, CY058491). Only alignments longer than 80 nucleotides were retained. The median sequencing depth was 14,400. Amino acid frequencies were calculated after aligning translated reads to the corresponding positions in the reference proteins. Unfolded SNP frequencies were generated using the IAV reference genome and used for the population genetics analyses and the amino acid frequencies were used for the structural analysis. The sequencing datasets generated in this study are available at http://bib.umassmed.edu/influenza.

4.6.7 Population genetic analysis

To perform population genetic analysis and identify the SNPs under selection from random drift, we applied a Wright-Fisher ABC approach (see **http://jensenlab.epfl.ch/page-86730-en.html**) to estimate effective population size and selection coefficient based on the allele frequency trajectories through time (196, 197, 199). We generated the trajectories of all SNPs once the newly derived allele frequency rose to greater than 2% in any passage. We first

estimated N_e, which determines the level of genetic drift, based on the trajectories of the entire genome. Then we estimated the selection coefficient of each SNP and the posterior probability given the genome-wide estimated N_e. If the posterior probability of the given trajectory with positive selection coefficient was greater than 99%, we considered these sites to be significantly under positive selection.

4.6.8 Structural analysis methods

The amino acid sequence of influenza A/Brisbane/59/2007 (H1N1) HA was obtained from UniProt using the accession number B0VX46, which is associated with the GenBank accession number CY030232. This HA sequence was aligned to the amino acid sequences of published crystal structures to determine the location of specific mutations on the structure of HA, and the possible impact of these mutations was determined based on what has been reported in the literature about HA structure, conformational changes in HA that occur during fusion, and HA receptor binding. The published crystal structures used in this analysis include a structure of F10 in complex with H5 HA (PDB ID: 3FKU), a structure of H1 HA bound to the human receptor analog sialylneolacto-N-tetraose c (LSTc) (PDB ID: 2WRG), a structure of solubilized trimeric H3 HA at the pH of membrane fusion (PDB ID: 1HTM), and structures of H2 HA at neutral and acidic pH (PDB ID: 3QQB, 3QQO). The mutagenesis wizard in PyMOL was

used to mutate residue 203 to an asparagine in two crystal structures to match the A/Brisbane/59/2007 (H1N1) HA sequence (PDB ID: 3FKU, 2WRG) (116).

Van der Waals contacts between F10 and H5 HA were calculated using an in-house script that applies a simplified Lennard-Jones potential as described previously (94). The van der Waals contact energy was calculated for all possible HA-F10 atom pairs within 6.0 Å of each other in the structure (PDB ID: 3FKU). Sequence conservation in group 1 HA was calculated using a previously published sequence alignment of all full-length non-redundant influenza group 1 HA sequences downloaded from the National Center for Biotechnology Information Influenza database in August 2011 (229, 230). To determine the percent conservation at each position in the alignment, the alignment was viewed using the Protein Family Alignment Annotation Tool editor (PFAAT) (234).

CHAPTER V

5 STRUCTURAL BASIS OF AN INFLUENZA HEMAGGLUTININ STEM-DIRECTED ANTIBODY RETAINING THE G6 IDIOTYPE

5.1 Abstract

Stem-targeting broadly neutralizing antibodies (sBnAbs) against hemagglutinin can neutralize a wide range of influenza viruses and are promising in developing strategies for broad protection against seasonal and pandemic infections. Many sBnAbs utilize the *IGHV1-69* germline gene, and the anti-idiotypic antibody G6 has been used to study antibodies derived from the *IGHV1-69* germline gene because G6 preferentially binds to *51p1* F-allele V segments of these antibodies. Typical potent *IGHV1-69* sBnAbs include a rearranged germline gene with distinctive V-segment substitutions that optimize binding to the conserved epitope on the HA stem. However, these substitutions have been shown to either reduce or abrogate G6 binding with only a few sBnAbs retaining the G6 idiotype. Here we report the crystal structure of D80, a G6-binding sBnAb, both alone and in complex with the anti-idiotypic Ab G6. The structures reveal only subtle changes in the sBnAb upon binding G6, suggesting a pre-organized binding epitope, including a pre-arranged CDR3 loop stabilized by extensive pi-stacking interactions that may be key for retaining the G6 idiotype while still being able to potently target the HA stem region.

5.2 Introduction

Seasonal influenza is a global health problem with millions of people infected and thousands dead due to infection every year (79). Pandemic influenza is a major threat and has occurred in the past when a new strain against which there is little or no immunity emerged, such as the 1918 “Spanish” flu, 1968 “Hong Kong” flu, and more recently the 2009 swine flu. Vaccines are the primary method of prevention against influenza. However, as a result of antigenic drift and a high mutation rate, vaccine strains need to be updated annually as new strains become dominant, and some years, the vaccine is not effective because circulating strains do not match the vaccine strains. In addition, vaccines take time to manufacture and may not be ready if a pandemic occurs (12). Although direct-acting antivirals such as neuraminidase inhibitors exist, they are only effective if administered early during infection, and emerging drug resistance limits their use and efficacy (5). Hence, we need improved prevention and therapeutic strategies against influenza that are effective and broadly protective, and understanding and applying broadly neutralizing antibodies (BnAbs) against influenza are promising for addressing some of these needs.

Influenza A is the most prevalent type of influenza, classified by two main glycoproteins on the surface of the virus, hemagglutinin (HA) and neuraminidase (NA). Various combinations of 16 HA and 9 NA subtypes define the strains of influenza A, which are further divided into two phylogenetic groups based on HA. HA binds to sialic acid on glycoproteins on the surface of cells to initiate viral

infection. HA also mediates fusion between the viral and endosomal membranes during the viral life cycle. Most antibodies (Abs) isolated against influenza and utilized in seasonal vaccines target the head region of HA, which is highly variable among various strains and easily mutates to escape Ab neutralization. Influenza BnAbs have been discovered that neutralize a broad spectrum of influenza viruses, and they target highly conserved regions of HA that are important for function, such as the stem region (sBnAbs). sBnAbs that neutralize all group 1, all group 2, or both groups of influenza A, and even cross-reactive against influenza B have been reported (11, 12, 58, 60-62, 64, 66, 67, 192). Therapeutic and prophylactic studies of these sBnAbs in mouse and ferret models are promising toward developing a “universal” therapeutic for broad protection against all influenza viruses.

Many influenza HA sBnAbs are derived from the immunoglobulin heavy chain variable region germline gene 1-69 (*IGHV1-69*) on chromosome 14 (12). These Abs target a conserved epitope on the stem region of HA and display robustness against neutralization escape (12). Previous studies identified distinct V-segment amino acid substitutions within the rearranged germline gene that are preferentially used to enable an Ab to become a potent sBnAb (235). As few as two V-segment mutations including I52S in CDR2 and a critical Tyr in CDR3 are enough for an sBnAb to mature from the *IGHV1-69* germline gene and bind to HA (235). Almost all identified *IGHV1-69* sBnAbs contain a hydrophobic residue

at position 53 and belong to the *51p1* F-allele group, which encodes a critical Phe54 in CDR2.

IGHV1-69 germline-derived Abs from the *51p1* allele group are specifically and uniquely recognized by the anti-idiotypic Ab G6 (236). Anti-idiotypic antibodies bind to the variable region of another antibody, and can be used as a diagnostic tool to monitor expression levels of Abs they recognize or as potential immunotherapy (237). G6 is a murine anti-idiotypic antibody that was discovered in a screen against rheumatoid factors and has been used to study many different targets, including rheumatoid factors, fetal splenic tissue, germinal centers in human tonsils, and B cell receptors in patients with B cell chronic lymphocytic leukemia (B-CLL) (236, 238-244). G6 has been used as a diagnostic tool for monitoring expression levels of *IGHV1-69* antibodies from B cells (245, 246). Although anti-idiotypic antibodies have been studied for several years, there are only a few structures of these antibodies in complex with their targets in the protein structure database.

Due to G6's potential therapeutic applications, G6 has been humanized using a structure-based CDR grafting approach, and some of the humanized G6 antibodies, especially HuG6.3, bind as tightly or better to their targets compared to the parental murine G6 (239). Humanized G6 could potentially be used as a "primogen" immunogen during influenza vaccination to promote proliferation of naive *IGHV1-69* B cells and memory *IGHV1-69* BnAb B cells that could then produce sBnAbs (12). However, we found that in most cases the same

substitutions in the V-segment that allow mature potent sBnAbs to evolve from the *IGHV1-69* germline gene reduce or abrogate G6 binding. We identified only a few *IGHV1-69* BnAbs that retain binding to both G6 and HA.

5.2.1 An example of a stem-directed broadly neutralizing antibody that neutralizes influenza and binds G6

D80 is an sBnAb that neutralizes all Group 1 influenza HAs and also binds to G6. D80 was discovered after two rounds of panning with a “non-immune” human phage-display antibody library against HA (194, 195). D80 can neutralize H5N1, and shares the same heavy chain as D8, which has been shown to be effective therapeutically and prophylactically in mice infected with influenza (194). D8 has subnanomolar binding affinity to HA and a very slow dissociation rate. In addition, prophylaxis with 10 mg/kg of D8 IgG one hour before lethal challenge with H5 and H1 viruses protected 80–100% of mice from death, and treatment with 15 mg/kg of D8 IgG 24-72 h following lethal challenge with H5 viruses also protected 80–100% of mice from death. D8 can bind to H2, H6, and H11 pseudotyped viruses.

In this study the crystal structures of one of these unique antibodies that retain binding to both G6 and HA, D80 Fab both alone and bound to humanized G6 is described. Both structures were refined to around 2.5 Å resolution and allowed detailed comparative structural analysis to identify the determinants of the G6 idiotype. There are no major conformational differences in the CDR loops between the bound and unbound structures of D80, suggesting a pre-organized

binding epitope. We discovered an extensive pi-stacking interaction network mainly involving the CDR3 loop that may contribute to the tight subnanomolar binding that we observe between D80 and G6 (239) and allow retaining the G6 idiotype while still binding to HA. Our structures provide key insights that may be scaffolds for the design of therapeutics or vaccines that utilize BnAbs and anti-idiotypic antibodies such as G6.

5.3 Results

5.3.1 Elucidation of the G6 Idiotype: Preferential binding to germline-like V-segments

Previous studies that interrogated the murine G6 (mG6) binding epitope suggested that the binding epitope is located on the CDR-H2 loop of *IGHV1-69* 51p1 F-allele V-segments, G6 does not bind to *IGHV1-69* Abs of the L-allele family that are primarily defined by Leu54, and binding does not seem to be restricted by CDR-H3 composition and light chain type (236, 238, 247). Since these studies analyzed only a limited number of antibodies, these observations were expanded and validated by performing 1 round of panning with a 27 billion-member naive human scFv (single chain variable region fragment) antibody library against beads coupled with mG6 (**Figure 5.1**). Immunogenetic analysis of the pool of scFvs that were both in-frame and where the VH and VL domains could be deciphered (n = 133) indicated that 89% of the scFvs are composed of *IGHV1-69* V-segments and all of these belong to the 51p1 F-allele group. In accordance with the lack of any antibody recovered in the library from the L-allele

family, when F54 is mutated to L54, G6 binding is lost. No preferential binding is observed to particular D-segments, J-segments, light chains and CDR-H3 length, in agreement with previous reports (236). However, preferential binding was observed to non-mutated germline configuration V-segments (**Figure 5.1G**). Most of the antibodies included no or only one or two substitutions relative to the germline gene, sharing more than 95% identity in amino acid sequence. Hence, almost all antibodies deviating more than 5% with respect to the germline gene lost the G6 idiotype.

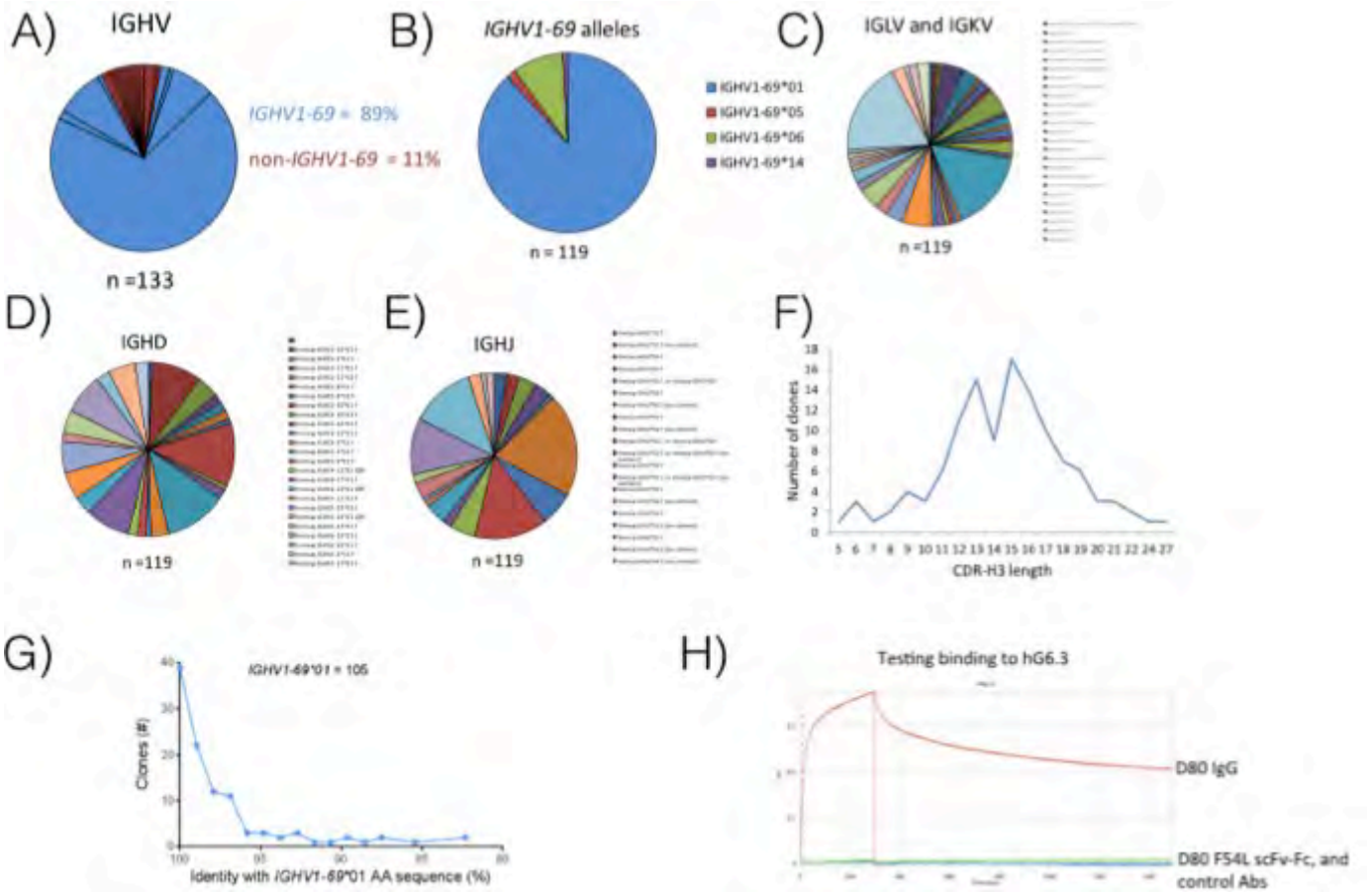


Figure 5.1 The G6 idiotype is defined by the CDR2 loop of the antibody heavy chain.

Figure 5.1 The G6 idiotype is defined by the CDR2 loop of the antibody heavy chain. Sequencing results were analyzed from 133 bacterial colonies from a phage display panning experiment where beads coated with G6 were incubated with a phage display library. A) Of the 133 colonies that were sequenced, 89% were of the antibodies were derived from the *IGHV1-69* germline gene. B) All of the antibodies from the phage display experiment that were derived from the *IGHV1-69* germline gene belong to the 51p1 F-allele group. C, D, E) G6 binding does not seem to be influenced by C) any particular light chain, D) D-segment, or E) J-segment. F) G6 binding also does not seem to be influenced by the length of the CDR3 loop. G) G6 preferentially binds to antibodies characterized by non-mutated V-segments similar to the *IGHV1-69* germline gene. H) D80 is an sBnAb that binds to G6, but when the mutation F54L is introduced to resemble L-allele *IGHV1-69* antibodies, binding to G6 is lost, confirming that the G6 idiotype is located in the CDR2 loop of the antibody heavy chain.

5.3.2 BCR Cross-Linking: F10 antibody maturation abrogates the G6 idiotype

The ability of the bivalent G6 (IgG) to bind to B-cells bearing *IGHV1-69* F-alleles (239) suggests that G6 can lead to cross-linking of the B-cell receptor (BCR), which in turn can initiate B-cell proliferation. As influenza stem targeted broadly neutralizing antibodies are mostly composed of *IGHV1-69* F-alleles (HV1-69-sBnAbs), theoretically, G6 can serve as a primogen that would bring to the proliferation of HV1-69-sBnAbs memory B-cells and precursor HV1-69-sBnAbs from the naive *IGHV1-69* F-allele B-cell pool.

Therefore, the ability of mG6 and humanized G6.3 (hG6.3) to cross-link B-cell receptors was tested by utilizing an artificial BCR display system. In this system, BCR cross-linking events were monitored by using an intracellular probe that exhibits fluorescence when calcium ions are released as a result of BCR cross-linking induced signal transduction events. To simulate conditions in which mG6 and hG6.3 target B-cells bearing HV1-69-sBnAbs, B cells were designed to display the mature F10, or F10 variants in which the V-segment was replaced by a non-mutated germline configuration *IGHV1-69*01* V-segment (VH1-69/F10), or F10 variants with 1-to-4 amino acid substitutions back introduced (**Figure 5.2**). Both mG6 and huG6.3 were able to cross-link the F10 variant with the germline-like V-segment, but not the mature F10. In addition, back-mutations to the germline gene increased cross-linking in the presence of G6, but in contrast, decreased cross-linking in the presence of HA. Curiously, enhancement of the G6 idiotype mirrored the decrease in HA binding as back-substitutions to the

germline gene were introduced. In addition, the mutations I52S and I53M either increase or do not impact B cell cross-linking in the presence of G6.

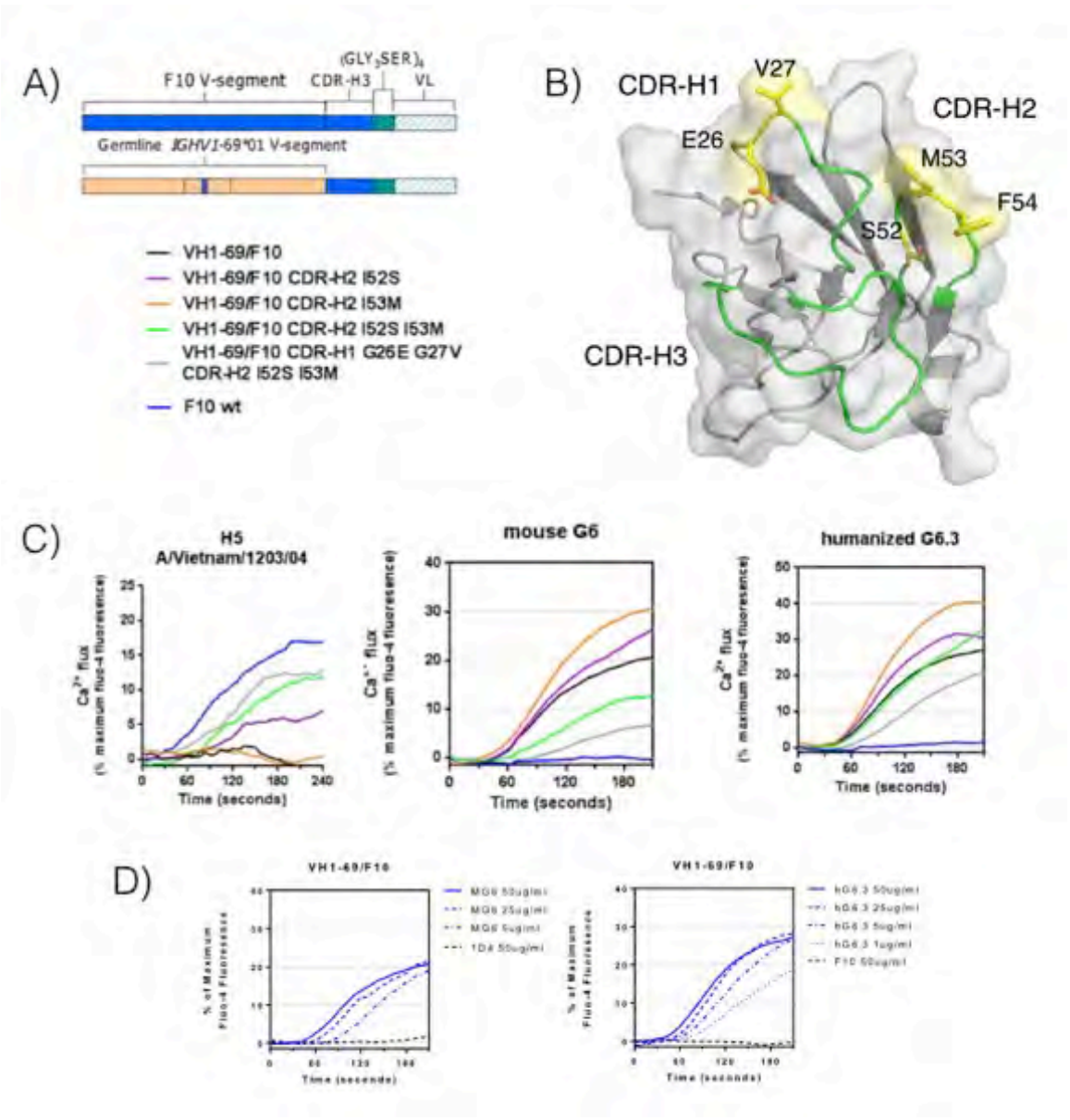


Figure 5.2 Murine G6 and humanized G6.3 cross-link B cell receptors that bear the G6 idiotype.

Figure 5.2 Murine G6 and humanized G6.3 cross-link B cell receptors that bear the G6 idiotype. As mutations are introduced to increase B cell cross-linking to HA, however, B cell cross-linking to G6 is decreased. A) Schematics showing the constructs of the scFvs of B cell receptors used in this experiment, illustrating the variable regions of the mature sBnAb F10 and the *IGHV1-69* germline gene V segment F10 hybrid (VH1-69/F10). VH1-69/F10 chimeras with specific mutations are shown in different colors. B) The locations of the introduced mutations are mapped onto the structure of the antibody VH domain and are shown in yellow. The remaining sections of the three VH CDR loops are shown in green. C) Back mutations to the germline gene increased cross-linking to G6, but in contrast, decreased cross-linking to HA. Mutations I52S and I53M either increase or do not impact B cell cross-linking to G6. D) Dose-dependent binding curves generated for the VH1-69/F10 chimeras and for the respective isotype controls.

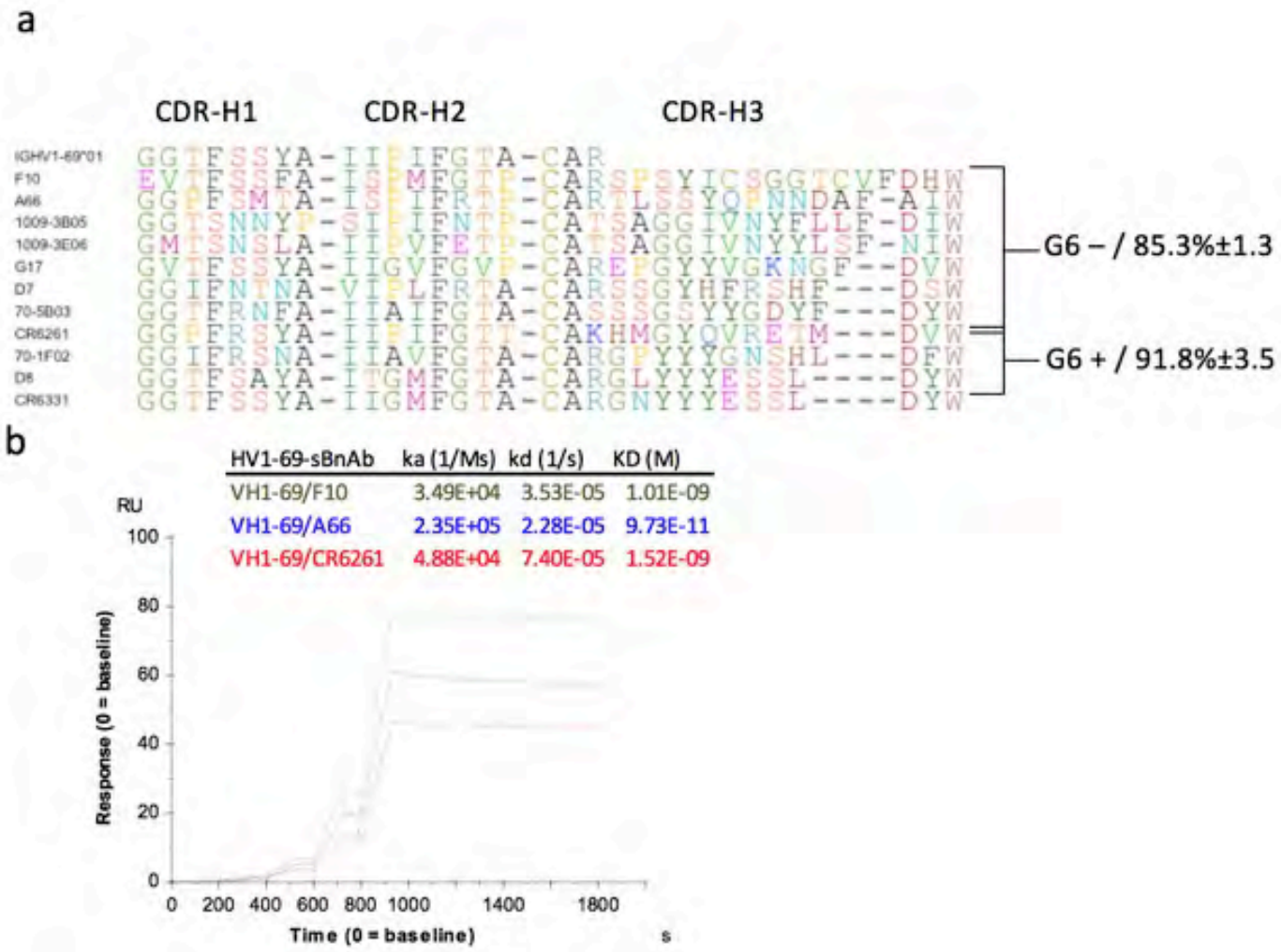


Figure 5.3 The G6 idiotype appears on select influenza stem-directed broadly neutralizing antibodies.

Figure 5.3 The G6 idiotype appears on select influenza stem-directed broadly neutralizing antibodies. A) Various HV1-69-sBnAbs were tested for G6 binding using ELISA or Biacore. Shown are the CDR-H1, H2, H3 amino acid sequences as well as the amino acid germline identity mean and STDEV of the G6 reactive and nonreactive groups. B) Converting the V-segment of the non-G6 reactive HV1-69-sBnAbs F10, CR6261 and A66 restores the G6 binding epitope.

5.3.3 Influenza stem-directed broadly neutralizing antibody binding to G6 versus HA

The lack of mature F10 binding to G6 led us to interrogate other HV1-69-sBnAbs binding to G6. With a panel of 11 different HV1-69-sBnAbs, whether maturation abrogated the idiotype was tested. The G6 binding epitope was observed to be lost on most influenza BnAbs from the IGHV1-69 germline gene. Of 11 different HV1-69-sBnAbs, only 3 were G6 reactive (**Figure 5.3**). These are: D8 (same as D80 but with a different light chain), CR6331 and 70-1F02. These three antibodies are defined by a low V-segment amino acid substitution frequency, and these antibodies also have smaller CDR3 loops and a CDR3 loop motif composed of three tyrosines at positions 97, 98 and 99.

To further confirm that the mature V-segment of HV1-69-sBnAbs prevents G6 binding, the V-segments of F10, A66, and CR6261 were replaced with the non-mutated *IGHV1-69* germline gene and binding kinetics to G6 were analyzed. In all three cases, G6 binding was rescued in the VH1-69/F10, VH1-69/A66 and VH1-69/CR6261 chimeras, confirming the substitutions in the V segment of these mature sBnAbs are responsible for the loss of the G6 idiotype. As seen with F10 above, gain of G6 binding was accompanied by a loss of binding to HA, as the chimeras did not bind to the H5VN04 HA (235).

Therefore, although G6 binds to antibodies derived from the *IGHV1-69* germline gene, maturation of antibodies into influenza HV1-69-sBnAbs results in a loss of the G6 idiotype in most cases. The substitutions in HV1-69-sBnAbs that

optimize binding to the conserved epitope on the HA stem either reduce or abrogate G6 binding, with only a few sBnAbs retaining the G6 idiotype.

To investigate the structure of G6 and the critical binding interactions in one of these few cases where an influenza sBnAb binds to both HA and G6, we solved crystal structures of one of these antibodies: D80 in complex with G6 and D80 unbound.

5.3.4 The crystal structures of the D80 Fab fragment and the D80-HuG6.3 Fab fragment complex

The crystal structures of the D80 Fab apo and bound to the HuG6.3 Fab were solved. The D80 Fab fragment was solved at 2.7 Å resolution in a P 21 21 21 space group with 2 molecules in the asymmetric unit. The D80 Fab fragment bound to the HuG6.3 Fab fragment was solved at 2.53 Å resolution in a C 1 2 1 space group with 1 molecule in the asymmetric unit (**Table 5.1**).

An overview of these structures is shown in **Figure 5.4**. The angle of binding between the two Fabs is 149 degrees. The elbow angle, which is the pseudo-dyad angle between the variable and constant domains, is 174 and 176 degrees in each asymmetric unit of the D80 apo structure. In the D80 bound structure, the angle is 173 degrees, which is similar to the unliganded structure, and in the G6 structure, this angle is 172 degrees. These elbow angles are consistent with results for Fabs with kappa light chains and for unliganded versus liganded forms of Fabs (248).

No major structural differences in the CDR loops were observed between the bound and unbound structures of the D80 Fab, suggesting a pre-organized binding epitope. To quantitatively measure the differences between the bound and unbound structures of D80, the distances were calculated between all of the alpha carbon atoms within each structure, the differences of these distances between the bound and unbound structures were computed. These differences were then graphed onto a contour plot, and the average differences were mapped onto the structure as previously described (**Figure 5.5**) (249). This method reduces bias that can be introduced by comparing structures using a structural alignment. The distance difference plots highlight regions of differences between bound and unbound structures of D80. (This calculation was performed only for the heavy chain of D80 because the light chain makes peripheral weak contacts with G6; discussed below). When looking at the VH domain of D80 in both the bound and unbound structures, the most structural differences were seen in the CDR1 loop, whereas the CDR2 and CDR3 loops had only a few regions of ~ 2 Å difference displacement. There are also small differences between the CH1 domains of the bound and unbound structures. In the comparison of the two domains VH and CH1 relative to each other in the bound and unbound structures (**top left and bottom right of Figure 5.5A**), there are some differences that propagated throughout the plots because the angles between the VH and CH1 domains are slightly different between the two domains.

Table 5.1 Crystallographic statistics for the D80-G6 complex crystal structure and D80 unbound crystal structure.

Tab e 5.1: Data Co ect on and Ref nement Stat st cs		
Structure	D80 Fab Fragment	D80 G6 Fab Fragment Comp ex
PDB Code	5JQD	5JO4
Reso ut on Range	46.12 2.59 (2.68 2.59)	49.61 2.53 (2.62 2.53)
Space Group	P 21 21 21	C 1 2 1
Un t Ce	89.06 91.58 106.76 90 90 90	143.16 50.78 139.74 90 90.9 90
Tota Ref ect ons	132762 (9946)	148763 (14956)
Un que Ref ect ons	26405 (2602)	33988 (3370)
Mu t p c ty	5.0 (3.8)	4.4 (4.4)
Comp eteness (%)	0.95 (0.95)	0.97 (1.00)
Mean I/S gma (I)	11.16 (2.42)	14.35 (1.99)
W son B Factor	46.81	39.51
R merge	0.09 (0.52)	0.08 (0.81)
R meas	0.10 (0.61)	0.10 (0.92)
CC1/2	0.996 (0.746)	0.998 (0.718)
CC*	0.999 (0.925)	0.999 (0.914)
Ref ect ons Used n Ref nement	26404 (2602)	33148 (3110)
Ref ect ons Used for R Free	1273 (127)	1968 (192)
R work	0.2152 (0.2891)	0.1966 (0.2766)
R free	0.2438 (0.2981)	0.2478 (0.3480)
CC(work)	0.93 (0.75)	0.95 (0.83)
CC(free)	0.96 (0.69)	0.91 (0.65)
Number of Non Hydrogen Atoms	6431	6757
Macromo ecu es	6338	6530
L gands	0	5
Prote n Res dues	851	866
RMS(Bonds)	0.005	0.010
RMS(Ang es)	0.76	1.36
Ramachandran Favored (%)	97	91
Ramachandran A owed (%)	3.2	7
Ramachandran Out ers (%)	0.24	1.8
Rotamer Out ers (%)	3.5	5.3
C ash score	7.11	10.23
Average B Factor	48.57	56.05
Macromo ecu es	48.70	56.48
L gands	N/A	63.23
So vent	40.05	43.21

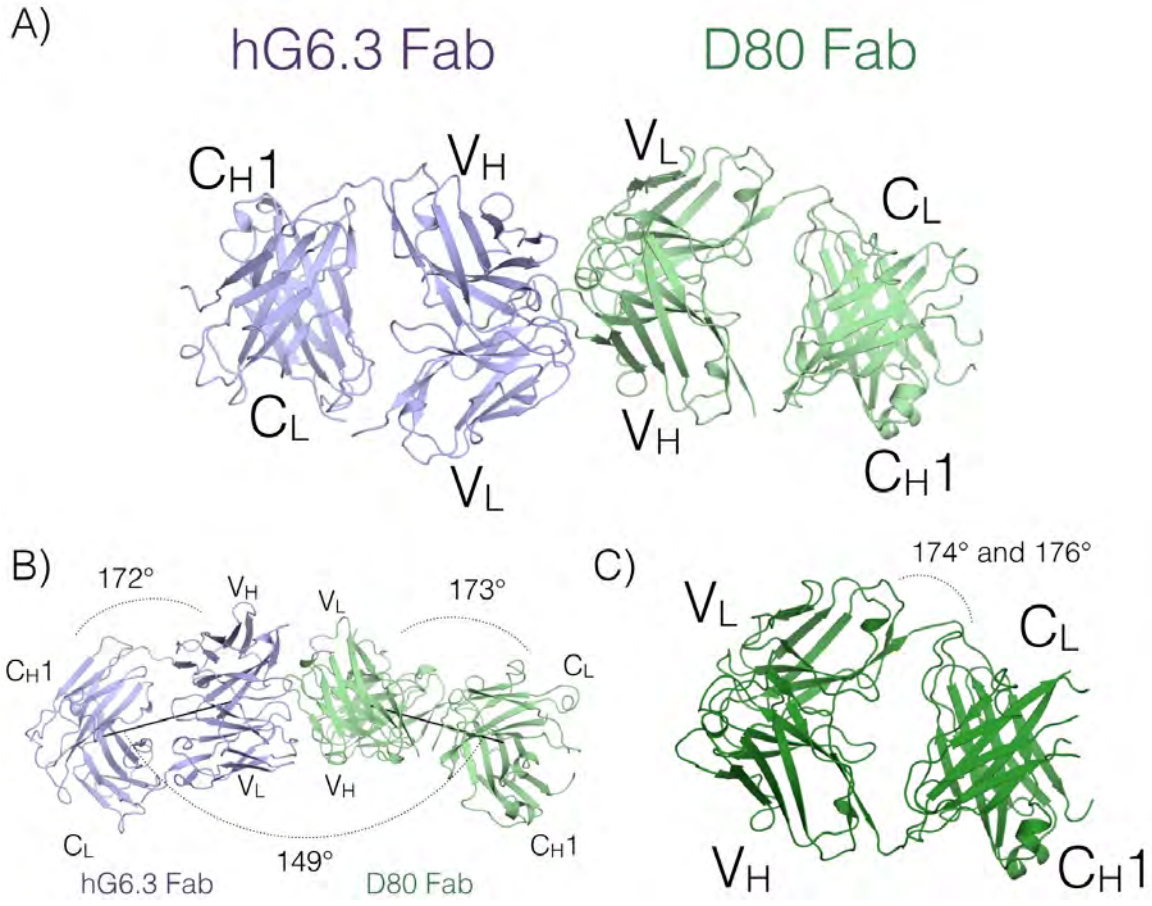


Figure 5.4 Crystal structures of a stem-directed broadly neutralizing antibody that bears the G6 idiotype: D80 in complex with G6.

Figure 5.4 Crystal structures of a stem-directed broadly neutralizing antibody that bears the G6 idiotype: D80 in complex with G6. A) Overview of the crystal structure of the D80 Fab fragment bound to the HuG6.3 Fab fragment. Each immunological domain on the Fab is labeled. D80 is in light green and G6 is in purple. The elbow angles in the G6 and D80 structures are 172 and 173 degrees, respectively. B) The binding angle between the two Fab fragments is 149 degrees. C) Overview of the crystal structure of the unbound D80 Fab fragment is shown in dark green. One molecule from the asymmetric unit is shown. The elbow angle is 174 and 176 degrees in each molecule of the asymmetric unit, respectively.

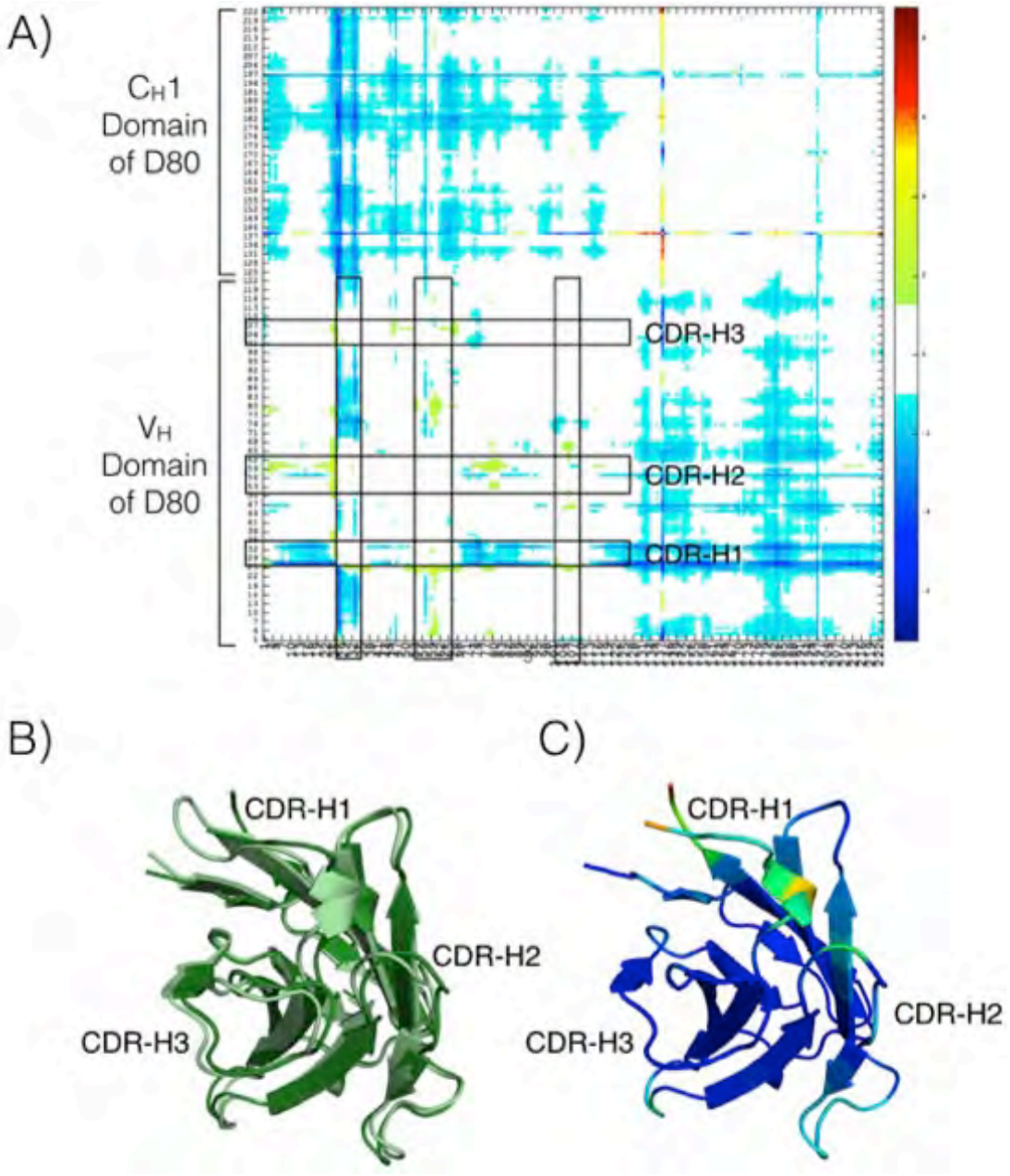


Figure 5.5 The bound and unbound structures of the D80 Fab fragment do not show significant differences.

Figure 5.5 The bound and unbound structures of the D80 Fab fragment do not show significant differences. A) A double difference plot of the heavy chain alpha carbons of the D80 fragment was calculated between the bound and unbound structures of D80. The areas on the plot that represent the three CDR loops in the VH domain of D80 are in boxes. B) A structural alignment of the bound VH domain of D80 (in dark green) and the unbound VH domain of D80 (in light green). C) The average double difference values for each residue were mapped onto the unbound structure of D80 and colored on a rainbow scale, where the greatest differences are shown in red and the least differences are shown in dark blue.

5.3.5 Intermolecular binding interface

Van der Waals contact energies of the residues participating in the binding interaction were calculated using a simplified Lennard-Jones potential and mapped onto the structures of D80 and G6 (**Figure 5.6**) (94). The two Fabs interact with each other through their variable domains, and the CDR loops of the VH region of the D80 Fab directly interact with their complementary CDR loops in the VH and VL domains of the G6 Fab. The VL region of D80 has only peripheral contacts with the G6 Fab. The heavy chain variable region of D80 is most important for binding, and CDRs 2 and 3 make the greatest contact with G6, especially residues M53, F54, G55, T56, N58 in CDR2 and Y98 and Y99 in CDR3. These residues insert into a pocket formed by the CDR loops of the heavy and variable chain regions of G6. The residues on G6 that make the most contact with D80 include $Y99_H$ and $W31_H$ from the heavy chain of G6, and $F94_L$, $Y91_L$, and $N32_L$ from the light chain of G6. The pocket in G6 is not entirely filled by D80, but could play a role in the diversity of antibody binding to G6. Using the POVME pocket volume analyzer (250, 251), the pocket was analyzed, and there is a remaining pocket volume of 32 $Å^3$ (**Figure 5.7)**. In general, water molecules cannot be visualized reliably at ~ 3.0 Å resolution, and the complex structure was solved at 2.52 Å resolution. However, we did see 1-2 waters buried in the interface between D80 and G6. In addition, hydrophobic residues and protein backbone atoms line this pocket. G6 binds to a large family of antibodies that are

derived from the 51p1 alleles of the *IGHV1-69* germline gene, so this pocket may be important for accommodating these different antibodies.

In addition to van der Waals interactions, hydrogen bonding and pi-stacking interactions were analyzed in the interface between D80 and HuG6.3 (**Figure 5.8**). In CDR1 of D80, hydrogen bonds exist between the side chain atoms of T28, S30, and Y32 with $W31_H$ and $D55_H$ of G6 (**Figure 5.8A)**. In CDR2 of D80, more extensive hydrogen bond interactions exist between side chain atoms of T52, T56, and N58 of D80 with $N32_L$, $S92_L$, and $Q93_L$ of G6. $Y99_H$ of G6 is also involved in a pi-stacking interaction with the backbone carbon of G55 in D80 (**Figure 5.8B**). In CDR3, there are no direct intermolecular hydrogen bond interactions, but van der Waals packing interactions with $F94_L$ and $Q93_L$ were observed (**Figure 5.8C**). In addition, an extensive pi stacking interaction network exists that involves the CDR2 and CDR3 loops of D80 with $F94_L$ and $W31_H$ in the pocket of G6, which are also residues in G6 that are involved in extensive van der Waals contacts with D80 (**Figure 5.8D)**. This pi-stacking interaction network is conserved between the unbound and bound structures of D80, and may contribute to the tight subnanomolar binding that observed between D80 and G6 (239).

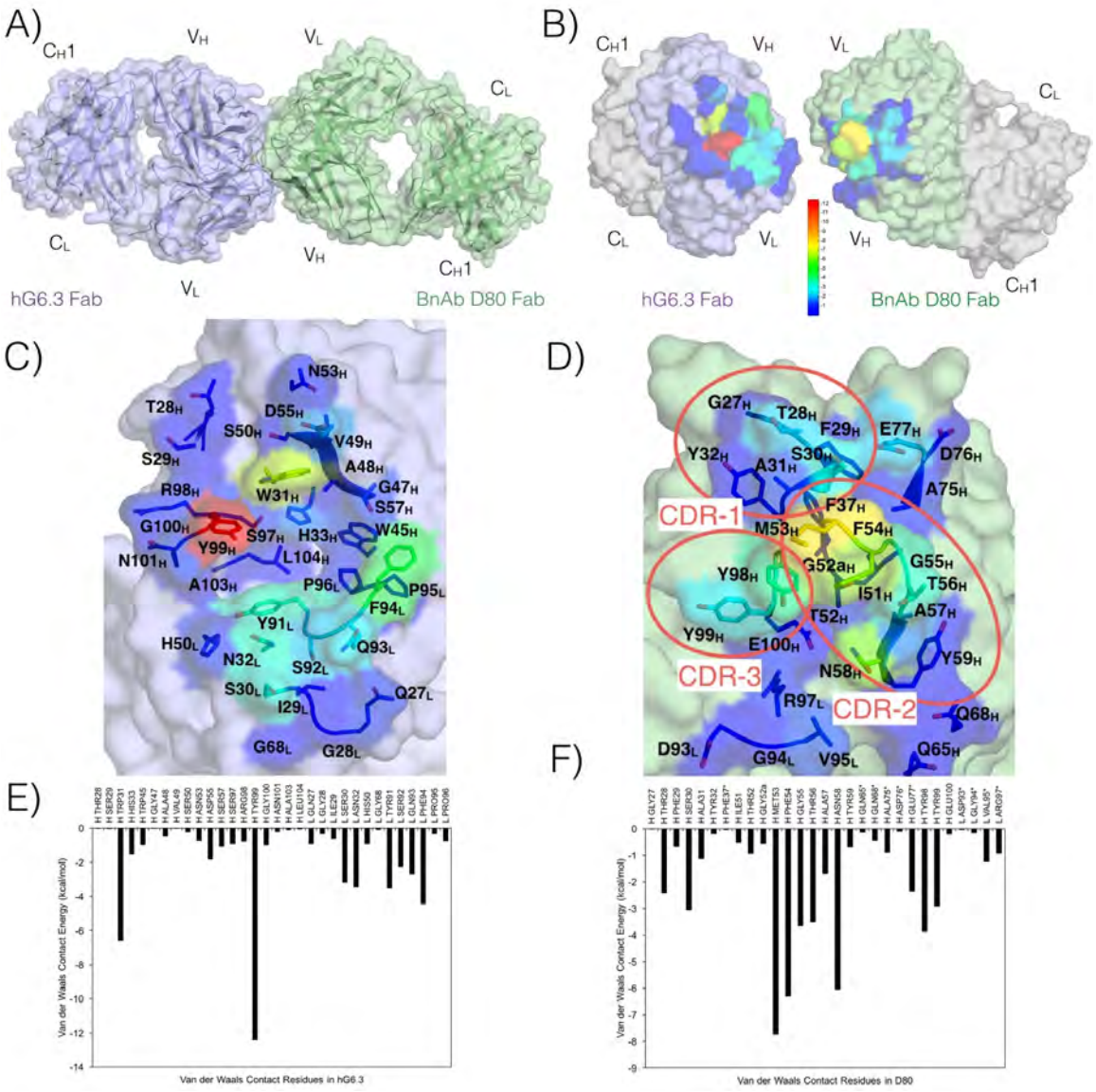


Figure 5.6 Overview of the binding epitopes and intermolecular interactions between D80 and G6.

Figure 5.6 Overview of the binding epitopes and intermolecular interactions between D80 and G6. A) Overview of the D80-G6 complex crystal structure shown in surface representation. The D80 Fab fragment is in light green, and the G6 Fab fragment is in purple. B) The D80-G6 complex crystal structure splayed open to view the residues in the binding site that interact with each other. The residues in the binding site are colored on a rainbow scale by the Van der Waals energy contacts, where residues with the greatest contacts are shown in red and the least contacts are shown in dark blue. The VH of D80 is shown in green and the variable region of G6 is shown in purple. The constant domains are shown in gray. C) A zoomed in view of the splayed open structure of G6 with interaction residues labeled, and E) a histogram of the per residue van der Waals contact energies for G6. D) A zoomed in view of the splayed open structure of D80 with interaction residues and CDR loops labeled, and F) a histogram of the per residue van der Waals contact energies for D80.

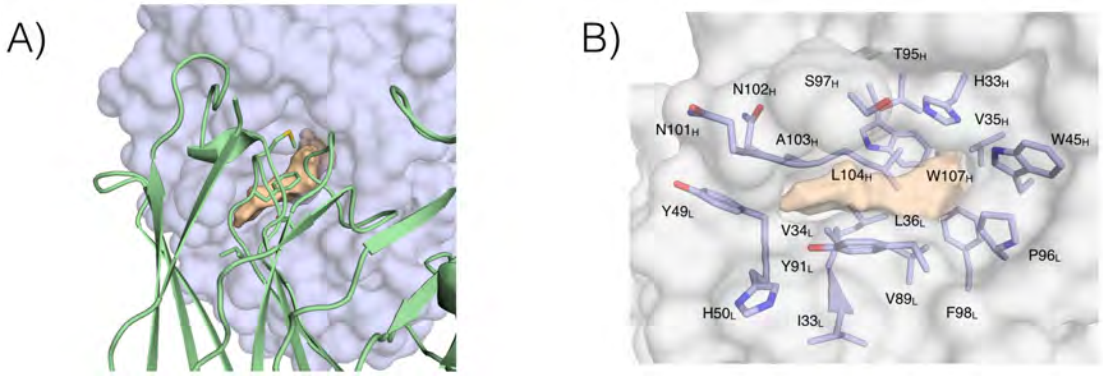


Figure 5.7 Characterizing the pocket in the binding site of G6.

Figure 5.7 Characterizing the pocket in the binding site of G6. A) The pocket within the binding site of G6 that remains upon binding D80 is shown in orange with G6 shown in purple and D80 shown in light green. B) The residues that line the pocket in G6 are shown in purple and the pocket is shown in orange.

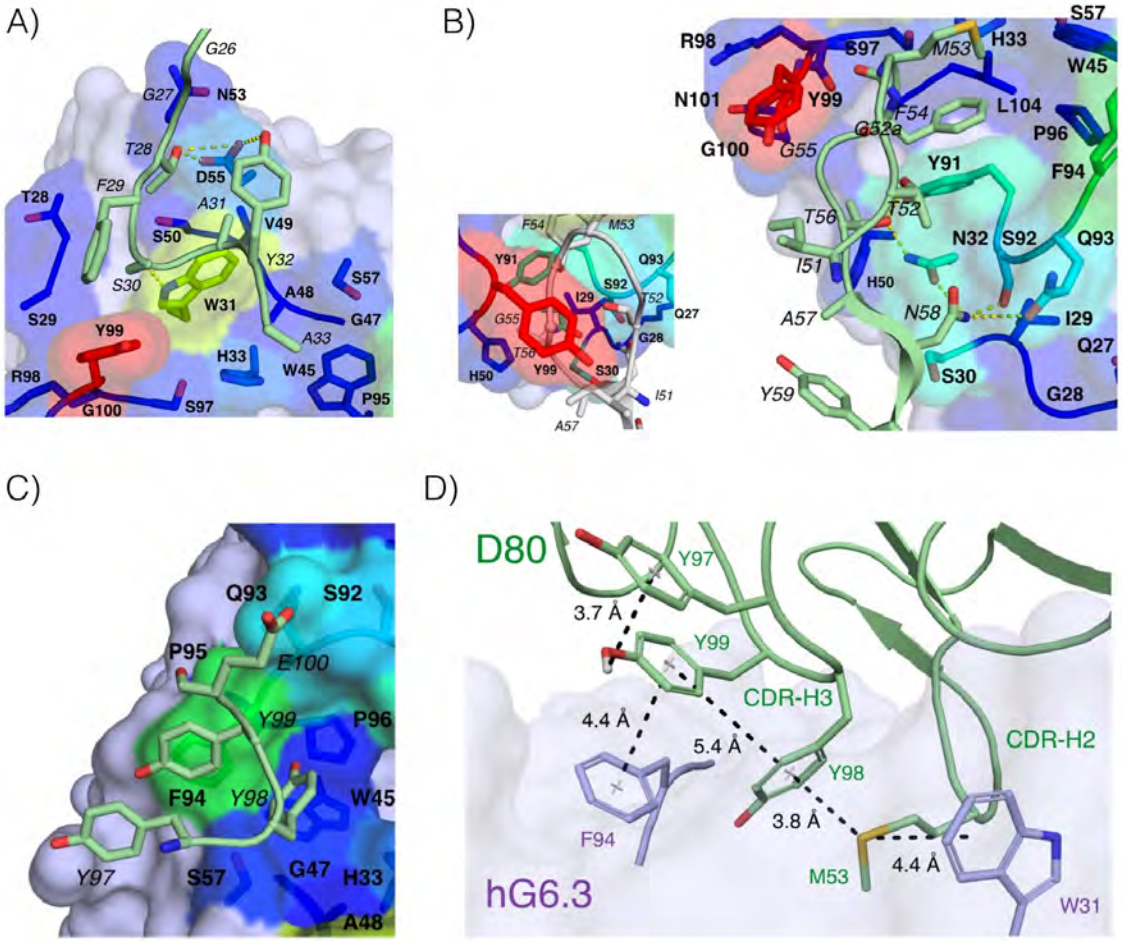


Figure 5.8 Hydrogen bond and pi-stacking interactions between D80 and G6.

Figure 5.8 Hydrogen bond and pi-stacking interactions between D80 and G6. A) Hydrogen bonding interactions between the CDR1 loop of D80 (shown in green sticks) and G6 (shown in purple). B) Main figure: Hydrogen bonding interactions between the CDR2 loop of D80 (shown in green sticks) and G6 (shown in purple). B) Left inset figure: $Y99_H$ of G6 is involved in a pi-stacking interaction with the backbone carbon of G55 in D80. C) Hydrogen bonding interactions between the CDR3 loop of D80 (shown in green sticks) and G6 (shown in purple). D) An extensive pi-stacking interaction network that involves the CDR2 and CDR3 loops of D80 (shown in green sticks) and $F94_L$ and $W31_H$ of G6 (shown in purple sticks).

5.4 Discussion

The crystal structure of D80 bound to G6 highlights some structural features that may be required for an sBnAb to bind to both G6 and HA. In addition to D80, the sBnAbs 70-1F02 and CR6331 also bind to both G6 and HA. One feature that these three Abs have in common is that they all have shorter CDR3 loops compared to the other eight sBnAbs that were tested for binding to G6. Hence, a long CDR3 loop may preclude binding to G6. In addition, these Abs have a triple tyrosine motif in the CDR3 loop, and the CDR2 loop contains flexible residues such as alanine and glycine in positions 52a and 57, whereas many other Abs that do not bind G6 contain proline at these positions. The triple tyrosine motif in D80 is involved in an extensive pi-stacking interaction network that may pre-organize the binding epitope and promote tight binding to G6 and HA, and related tyrosine pi-stacking interactions may also be involved in 70-1F02 and CR6331 binding to both G6 and HA.

The *IGHV1-69* germline gene encodes two hydrophobic residues on CDR2 that are important for binding to hydrophobic pockets, such as the stem region on HA and the pocket in G6, and residue F54 has also been shown to be critical for binding to G6 (236). As sBnAbs mature and accumulate mutations, the CDR2 loops often acquire prolines or other rigid residues, which may abrogate binding to G6.

Theoretically, G6 could be used as a “primogen” immunogen during influenza vaccination to promote proliferation of naïve *IGHV1-69* B cells and

memory *IGHV1-69* sBnAb B cells that could then produce sBnAbs (12). sBnAbs are not often elicited naturally in humans because the globular head of HA is immunodominant, even though stem epitopes have been shown to be accessible to sBnAbs (194, 235, 252-254). The germline precursors of *IGHV1-69* sBnAbs, such as CR6261 (which does not bind G6 as the mature antibody), do not bind soluble HA but do engage HA when the antibody is expressed as a cell surface IgM and trigger B-cell receptor associated tyrosine kinase signaling (12, 255). G6 could play a role as a potential substrate for affinity maturation of *IGHV1-69* B cells and promote proliferation of germline precursors of *IGHV1-69* B cells that could then mature and produce sBnAbs (255). G6 may also help to G6 binds best to *IGHV1-69* antibodies that are in the germline configuration.

Some other strategies that have been used to develop therapeutics with sBnAbs include priming with a DNA vaccine followed by boosting with a seasonal vaccine and the development of “headless” HA immunogens that have been designed to promote an antibody response to the HA stem (12, 256-261). There is a report of a successful anti-idiotypic antibody vaccine that was developed in cats for feline infectious peritonitis virus (262). In that case, the anti-idiotypic antibody was used as an “image” of the antigen that could induce immune-mediated responses similar to those produced by the original antigen, and anti-idiotypic antibodies have also been used in the development of other vaccines (262-265).

In conclusion, we elucidated the determinants of the G6 idiotype and solved the crystal structure of D80, a G6-binding sBnAb, both alone and in complex with the anti-idiotypic Ab G6. This work provides key insights that may guide the design of future therapeutics or vaccines that utilize BnAbs and anti-idiotypic antibodies such as G6.

5.5 Methods

5.5.1 Panning the naïve human scFv-phagemid library against beads coupled with G6

Magnetic M-280 Tosylactivated dynabeads were coupled with an isotype control mouse Ab and with mG6 according to the manufacturer's protocol. The phagemid library mixture was subtracted of non-specific and mouse-IgG binders by performing two rounds of subtractions with the isotype control beads after which the phagemid suspension was added to G6 coupled beads. Following six washes with TBS-Tween-20, the phagemids were eluted with 100 mM of TEA and neutralized with PBS. The eluted phagemid suspension was used to infect TG1 cells from which single clones were send to sequencing.

5.5.2 B cell cross-linking assay and binding kinetics assay

B cell receptor cross-linking assays were performed as described in the study of Andrew McGuire, et al (266). Biacore assays were performed using CM5 chips that were immobilized with anti-mouse Ab according to manufacturer's protocol. G6 was captured at a level of ~320 RUs and scFvs or IgG were flowed

over the chip. The single cycle kinetics function was employed with association set for 120 sec and dissociation for 800 sec.

5.5.3 Protein expression, purification, and crystallization

For crystallographic purposes, Fab fragments were expressed in insect cells using a baculovirus expression system. Fab fragments were purified using a polyhistidine tag with nickel-affinity chromatography and subsequent size exclusion chromatography. For unbound D80 crystallization, the D80 Fab fragment crystals were grown at 20 degrees Celsius by hanging-drop vapor diffusion in a solution using 20% PEG 4000, 0.2 M CaCl2, and 0.1 M Tris pH 8.4. D80 crystallized in a P 21 21 21 space group with 2 molecules in the asymmetric unit and unit cell dimensions of a = 89.0, b = 91.6, and c = 106.8 Å. For D80-G6 complex crystallization, the D80-G6 Fab fragment complex crystals were grown at 20 degrees Celsius by hanging-drop vapor diffusion in a solution using 0.1 M NH4SO4 and 12% PEG 3350. The D80-G6 complex was crystallized in a C 1 2 1 space group with 1 molecule in the asymmetric unit and unit cell dimensions of a = 143.2, b = 50.8, and c = 139.7 Å.

5.5.4 Diffraction data collection and structure solution

Diffraction data was collected at cryogenic temperatures using synchrotron radiation at the GM/CA-CAT 23-ID-B beam line at the Argonne National Laboratory (Advanced Photon Source, Chicago, IL). Data was indexed and scaled using xia2 and HKL3000 (267-273). Molecular replacement solutions

were determined using M-RAGE in PHENIX (274, 275). Cycles of rigid body, restrained, and TLS refinement with subsequent model building were performed using phenix.refine in PHENIX and Coot, respectively (274, 276).

5.5.5 Structural data analysis

Molecular graphics images and elbow angles were calculated using PyMOL (116). Double difference plot calculations and van der Waals calculations were performed using in house scripts as described previously (94, 249). Pocket volume analysis was performed using the POVME pocket volume analysis tool (250, 251). Hydrogen bonds were detected using the PyMOL polar contacts tool, where h_bond_cutoff_center was set to 3.6 Å and h_bond_cutoff_edge was set to 3.2 Å (116). Pi-stacking interactions or pi-atom interactions were detected in Maestro and defined using the default values as follows: A face-to-face pi stacking interaction is defined as an angle between ring planes of less than 30 degrees and a distance between ring centroids of less than 4.4 Å. An edge-to-face pi stacking interaction is defined as an angle between ring planes that is between 60 and 120 degrees, and a distance between ring centroids that is less than 5.5 Å (108). A pi-atom interaction is defined as a distance between an atom and a ring centroid of less than 6.6 Å, and the angle between the ring plane and the line between the cation center and the ring centroid does not deviate from the perpendicular by more than 30 degrees.

CHAPTER 6

6 DISCUSSION

6.1 Impact of uncertainty in coordinates on molecular dynamics simulation analysis

One concern with the molecular dynamics simulations examined in this thesis is that the precision of the simulations has not been addressed, specifically how the uncertainty in the crystal structure coordinates relates to the precision of the quantities calculated from the simulations. To begin to address this concern, Table 6.1 presents some known parameters from the simulations that can provide initial insight into this uncertainty, including some crystallographic statistics from the starting models and the correlation of the average alpha carbon root mean squared fluctuation (RMSF) from the simulations to the average alpha carbon temperature factors from the crystal structures.

Figure 2.2 also shows how the average RMSF and average crystallographic temperature factors compare in both the N1 and N2 NA molecular dynamics simulations. An important question for the energies calculated from molecular dynamics trajectories is what energy differences would be considered significant, how would this significance be determined, and do energies calculated from the observed structural differences taking into account the expected coordinate uncertainty exceed this value in a significant way. For this thesis, the van der Waals energy calculation was performed every 200

picoseconds and averaged over 400 nanoseconds. To calculate the standard error of the average for each residue in each simulation, the data was “chunked” using block averaging and significant differences were calculated. Increasing the sampling of the calculation from 200 picoseconds to every 10 picoseconds would also further validate these calculations. However, these calculations do not specifically address how the coordinate uncertainty in one static crystal structure would impact the van der Waals energy calculation and how this error might be propagated throughout a molecular dynamics simulation. Thoroughly addressing this specific question especially for the van der Waals energy calculation would be important for validating this work and also validating future and past work that examines differences in van der Waals energy calculations from molecular dynamics simulations.

Table 6.1 Starting crystal structures for molecular dynamics simulations

Influenza NA Subtype	Ligand	Protein Starting Structure (PDB ID)	Monomers in Asymmetric Unit	Resolution (Å)	R-Value Work/ R-Value Free	ESU Based On R / R Free / Maximum Likelihood (Å)	Average Alpha Carbon B Factor	Average Alpha Carbon RMSF (Å)	CC
N1	Zanamivir	3B7E	2	1.45	0.14/0.16	0.05 / 0.05 / 0.03	9.7	0.5	0.6
N2	α-2,3 Substrate	4GZW	4	2.45	0.20/0.25	0.54 / 0.28 / NULL	26.4	0.5	0.5
N2	α-2,6 Substrate	4GZX	4	2.45	0.21/0.27	1.19 / 0.35 / NULL	26.7	0.5	0.5
N2	Oseltamivir	4GZP	1	2.3	0.21/0.28	0.38 / 0.27 / 0.23	37.0	0.5	0.5

6.2 Assessing equilibration of molecular dynamics simulations

Additional concerns about the molecular dynamics simulations examined in this thesis are concerns that the simulations are equilibrated and that quantities calculated from the trajectories should be calculated from the equilibrated aspects of the trajectory only. Traditionally, equilibration is based on the convergence of a root mean squared deviation (RMSD) plot that is visually inspected, and equilibration is determined intuitively, but this method is subjective and controversial (277). To begin to address the concern of equilibration in the molecular dynamics simulations examined in this thesis, the total potential energy during each simulation was graphed for all simulations in Figures 6.1 and 6.2 to show that the total potential energy of each system is equilibrated over time. In addition, Figure 6.3 shows the overall RMSD of each chain from each trajectory over the course of the simulation. The overall RMSD is approximately 1 Å for each chain.

The total potential energy for all systems remains constant throughout the simulation, and based on the current RMSD plots, aside from Figure 6.3A and 6.3B, most of the plots demonstrate a relatively low overall RMSD around 1 Å that remains relatively constant, so this information provides initial insight into how the simulations are behaving. To further evaluate and confirm that the system has equilibrated, especially for the simulations analyzed in Figure 6.3A and 6.3B, it would be important to investigate what portions of the protein are influencing the increase in RMSD from 1 to 1.5 Å and confirm that these regions

of the protein would not have an impact on the results presented in the thesis. In addition, recalculating the values presented in the thesis by using only equilibrated portions of the trajectory instead of the full trajectories and examining how each van der Waals energy for each residue varies over the course of each trajectory would provide further validation, but we do not anticipate that these additional calculations would impact the final results.

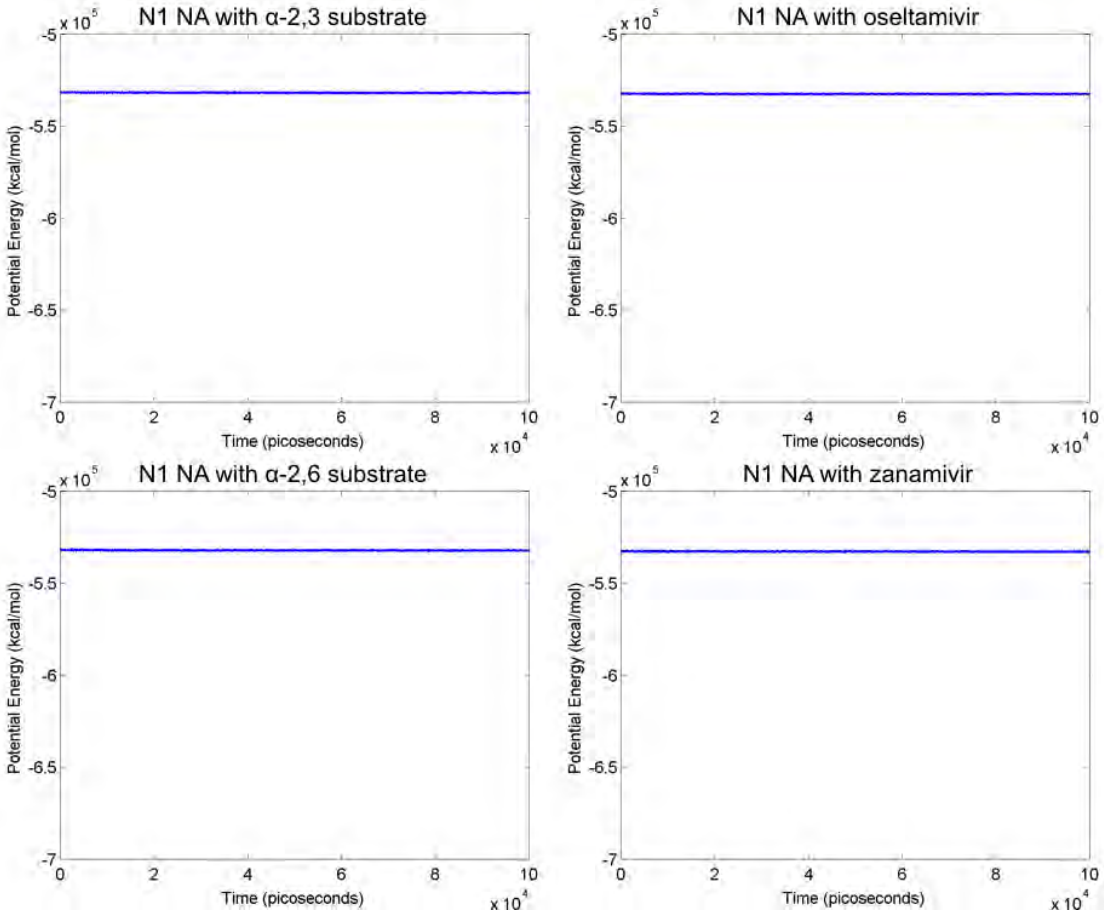


Figure 6.1 Potential Energy Over the Trajectory for N1 NA Molecular Dynamics Simulations.

Figure 6.1 Potential Energy Over the Trajectory for N1 NA Molecular Dynamics Simulations.

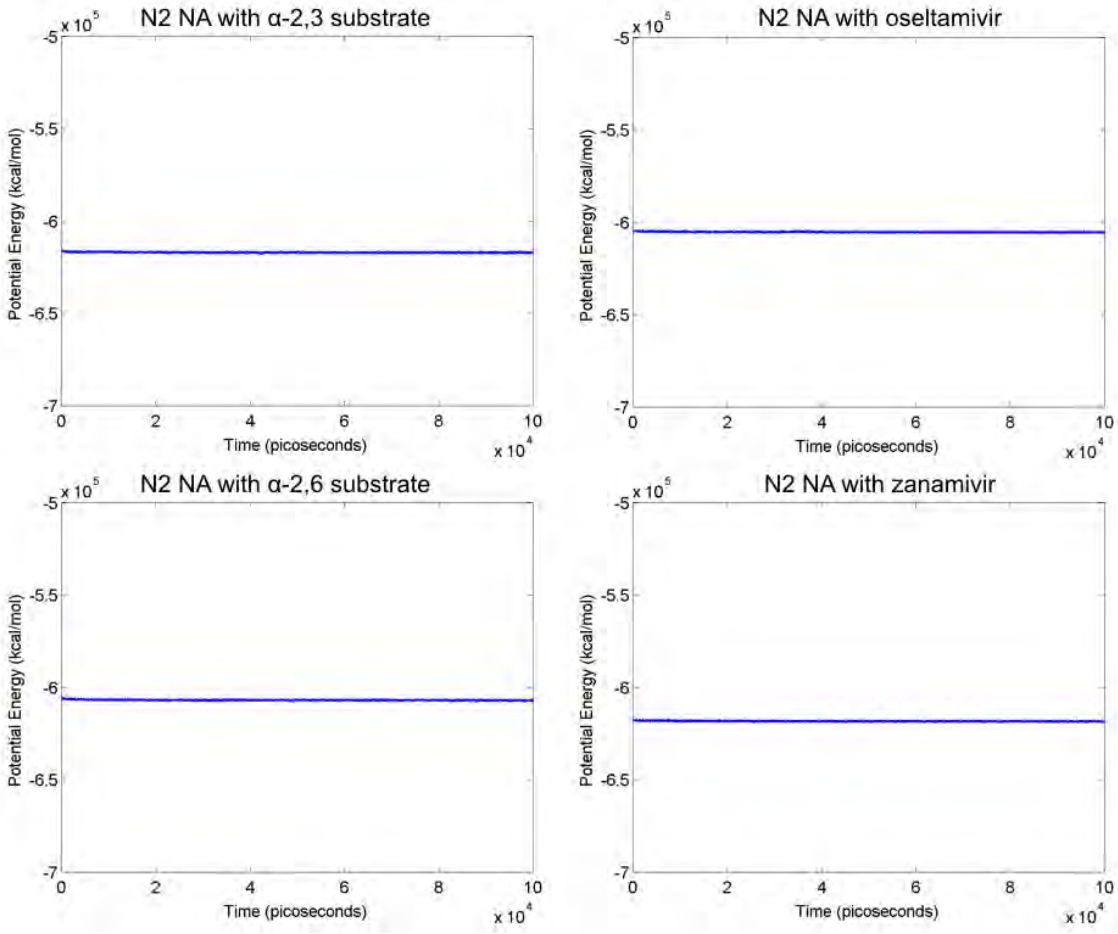


Figure 6.2 Potential Energy Over the Trajectory for N2 NA Molecular Dynamics Simulations.

TO abbreviation = Truncated Octahedron (shape of water box)

Figure 6.2 Potential Energy Over the Trajectory for N2 NA Molecular Dynamics Simulations.

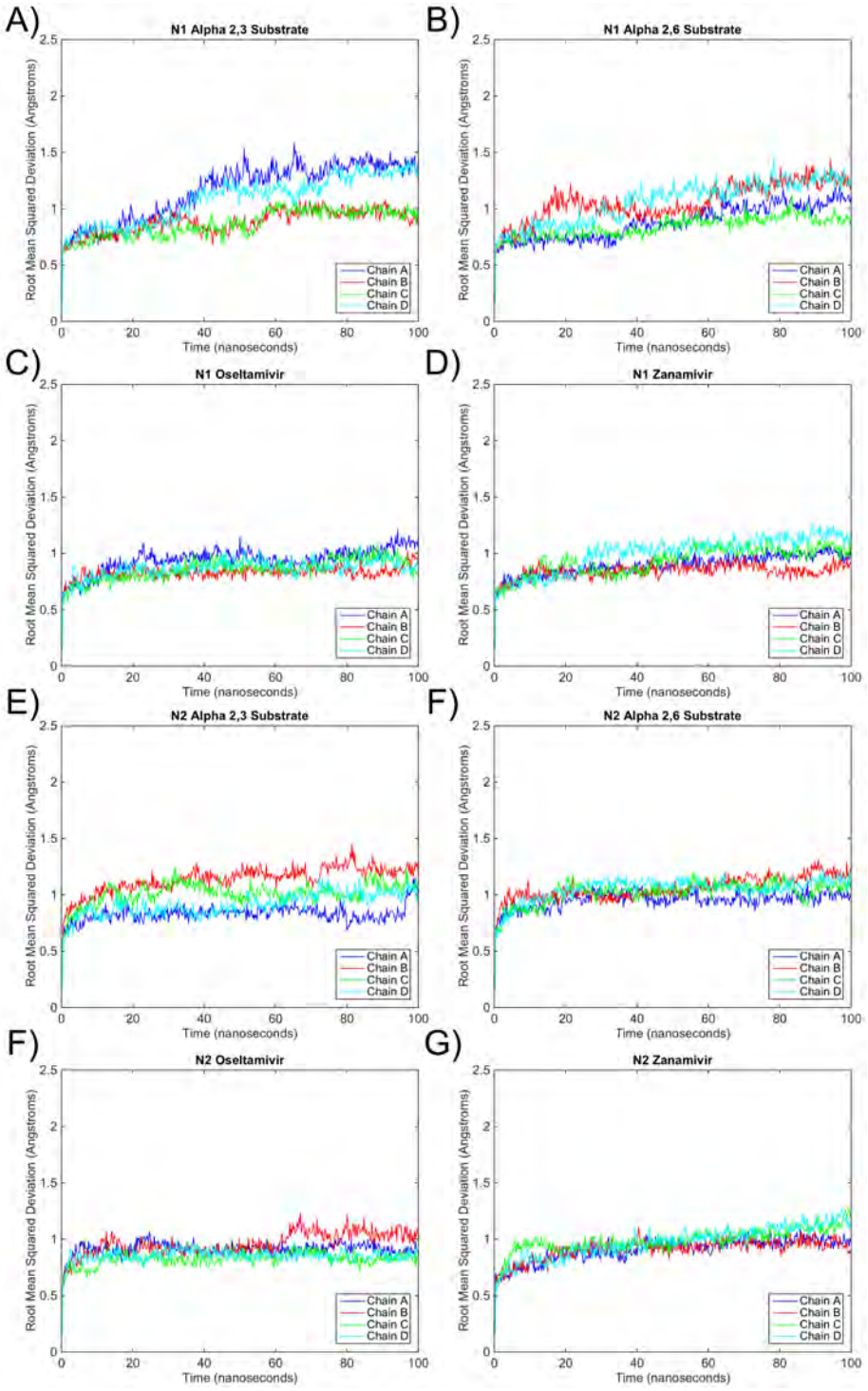


Figure 6.3 Per Chain Root Mean Squared Deviation for Molecular Dynamics Simulations.

Figure 6.3 Per Chain Root Mean Squared Deviation for Molecular Dynamics Simulations.

6.3 Dynamic substrate envelopes compared to static substrate envelopes

Another question about the dynamic substrate envelope is whether dynamic substrate envelopes are better than static substrate envelopes for NA. Previous work by Ozen, et al has shown that the dynamic substrate envelope was better than that static substrate envelope for HIV protease and HCV NS3/4A protease (94, 95). Local and global dynamics are important for function for many enzymes, so incorporating additional information from protein dynamics can provide a more accurate representation of the volume occupied by the substrate in the active site. However, in both HIV protease and HCV NS3/4A protease, the dynamic substrate envelope reflects many of the same characteristics of the static substrate envelope, and we expect that this pattern would be maintained with NA. However, the static substrate envelope represents a step-function like assessment of interactions in the active site whereas the dynamic substrate envelope represents a probabilistic distribution function, which inherently contains more information. However, to thoroughly confirm that the dynamic substrate envelope reflects many of the essential characteristics of the static substrate envelope, the static substrate envelopes for each system can be determined quantitatively to compare interactions directly with the results from the dynamic substrate envelopes.

6.4 Generalization of N1 and N2 NA sequences examined in molecular dynamics simulations compared to all N1 and N2 NA sequences.

The sequences in the crystal structures of N1 and N2 that were examined in this study are representative of N1 and N2 sequences in general. We chose prototypic N1 and N2 sequences for this study based on three criteria: 1) existence of high quality high resolution crystal structures for MD simulations 2) presence of a “typical” 150-loop in the active site based on previous reports for N1 and N2, and 3) high percent identity to N1 and N2 consensus sequences based on multiple sequence alignments (72, 106). All of the crystal structures used in this study are of the globular head domain. Alignments were performed using the multiple sequence alignment tools available on the Influenza Research Database (www.fludb.org) and accessed on February 18, 2016. The strain of N2 NA used is A/Tanzania/205/2010 H3N2 NA. This strain has 94% sequence identity and 96% sequence similarity to a consensus sequence determined from an alignment of 8,745 complete and unique sequences. The strain of N1 NA used is A/Brevig Mission/1/1918 H1N1 NA. This strain has 92% sequence identity and 96% sequence similarity to a consensus sequence determined from an alignment of 7,370 complete and unique N1 NA sequences. The high sequence similarity and identity of the crystal structures compared to the consensus sequence from the multiple sequence alignments shows that the crystal structures examined in the study are representative of N1 and N2 sequences in general.

6.5 Role of electrostatics in analysis of drug resistance residues in influenza neuraminidase

To strengthen the analysis of the molecular dynamics simulations of influenza neuraminidase, the role of electrostatics for the drug resistance residues that were examined should be determined. The key questions to answer are whether charge-charge interactions might play a role in differential binding of substrates (typically uncharged) versus inhibitors (typically charged). To address this concern more quantitatively, the Coulombic contribution to interaction energies can be calculated using pre-existing tools available from Schrodinger that were developed to analyze Desmond trajectories, including the Simulation Interaction Analysis tool and additional analysis scripts. Using these tools, the Coulombic contribution to interaction energies can be directly compared to van der Waals interaction energies, and both quantities can be calculated using the same tools. The Salt Bridges Plugin in VMD can also be used to identify salt bridges that occur over the course of a trajectory and measure the duration of the interactions.

Electrostatics play an important role in the high binding affinity of neuraminidase inhibitors compared to substrates (278, 279). Von Itzstein, et al used structure based drug design to develop the neuraminidase inhibitor zanamivir, and they discovered that there is a negatively charged pocket in the neuraminidase active site lined by residue E119, which is one of the drug resistance residues discussed in this study (278). The binding affinity of the transition state analog of sialic acid, DANA (2,3-didehydro-2-deoxy-*N*-

acetylneuraminic acid), to neuraminidase is 4 μM, but when the C4 hydroxyl on DANA was replaced with an amine, they discovered the binding affinity improved to 0.04 μM. When this moiety was replaced with guanidinium, the binding affinity improved to between 0.08-0.7 nM (278, 279). Crystal structures of these inhibitors were solved in complex with influenza A/Tokyo/3/67 neuraminidase, and they showed that the amine forms a salt bridge with E119, and one of the terminal nitrogen atoms on the guanidinium is 3.5 Å from the side chain carboxylate of E119 and likely also within the electrostatic influence of E119 (278). Furthermore, E119 does not form these interactions with sialic acid, which was also noted previously by von Itzstein, et al (278). To address electrostatics in the analysis described in this study of comparing substrate and inhibitor interactions in both N1 and N2 subtypes, the electrostatic contribution to interactions should be analyzed more quantitatively and compared to the calculated van der Waals interactions to address the contribution of both van der Waals and Coulombic energy to the differences in molecular interactions seen between substrates and inhibitors in N1 and N2 neuraminidase.

6.6 Incorporating the substrate envelope hypothesis and protein dynamics into drug design

Clinically, a common strategy for treating diseases and decreasing susceptibility to drug resistance is to use combination drug therapy to create a high genetic barrier to the development of resistance. With this strategy, different drugs inhibit distinct targets simultaneously, decreasing the likelihood of (both or)

all targets evolving to select resistance mutations at the same time. However, developing individual drugs to avoid drug resistance is still important so that robust drugs are available when needed. Drug development can take over 12 years and cost over $300 million dollars to bring a single drug from the research laboratory to a patient, so investing effort to design drugs that are less susceptible to drug resistance from the beginning can be very valuable (280). Currently, there are clinical situations where some diseases have become refractory to nearly all treatments as a result of drug resistance mutations, such as carbapenem-resistant enterobacteriaceae (CRE) and vancomycin resistant enterococcus (VRE), which have become resistant to “drugs of last resort” (281, 282).

Structure based drug design (SBDD) should incorporate strategies for reducing susceptibility to drug resistance, such as the substrate envelope hypothesis. In addition, while current SBDD incorporates ligand flexibility, aspects of target flexibility are not as well integrated. A common strategy in SBDD is to design relatively rigid ligands with increased van der Waals contacts and hydrogen bonds in the binding site, which was used in the design of HIV protease inhibitors, but these conformationally constrained ligands may not be able to adapt to drug resistance mutations and dynamic changes in the binding site (46). Including aspects of protein dynamics and the substrate envelope hypothesis into SBDD would support the development of more effective and robust drugs that are less susceptible to drug resistance (283-285).

Some general principles for SBDD have been developed based on lessons learned through evaluation of the substrate envelope hypothesis in HIV-1 and HCV proteases and the results presented in this thesis on influenza NA (43, 44, 46, 176, 286). To minimize susceptibility to drug resistance, drugs should be designed to remain within the substrate envelope and refrain from making extensive contacts with residues outside of the substrate envelope that are not important for substrate binding. In addition, taking advantage of unused regions of the substrate envelope to increase inhibitor contacts can improve potency. It may be necessary for a drug to protrude from the substrate envelope to improve the pharmacokinetic properties of a drug or to facilitate chemical synthesis. If a drug must protrude outside of the substrate envelope, the protrusions should be either in the direction of solvent or evolutionarily conserved residues if possible, such as catalytic residues or structural framework residues, to reduce the potential for the development of drug resistance mutations.

Inhibitors are often designed to form multiple intermolecular hydrogen bonds to increase potency, and these hydrogen bonds should also be optimized to target evolutionarily conserved residues (46, 176, 286). To minimize susceptibility to drug resistance mutations, inhibitors can be designed to establish hydrogen bonds with backbone atoms, since these hydrogen bonds may be impacted less by mutations compared to hydrogen bonds with side chain atoms. Inhibitors can also be designed to fill interactions that are made by

conserved water molecules that form hydrogen bonds with conserved side chain or backbone atoms in the binding site (176).

In addition, inhibitors should be designed to maintain a balance between flexibility and potency (46, 176, 286). Natural substrates are often more flexible than inhibitors, which can allow them to accommodate changes in the active site, such as drug resistance mutations or changes due to different protein conformations. Inhibitors that mimic the flexibility of natural substrates may decrease the susceptibility to drug resistance, but inhibitors that are too flexible may not be very potent, so a balance between inhibitor flexibility and potency must be maintained. For example, in influenza NA, the glycerol moiety on zanamivir matches the flexibility of the glycerol moiety on sialic acid, and zanamivir is not susceptible to drug resistance mutations at this location. However, oseltamivir has a pentyl-ether hydrophobic group at this location, and it is highly susceptible to the drug resistance mutation H274Y at this location. In HCV protease inhibitors, macrocycles, which are chains of carbons that join two distant moieties in an inhibitor together, can also be used to position inhibitors in an optimal conformation in the active site to enhance inhibitor potency while also allowing the inhibitor enough flexibility to accommodate drug resistance mutations (176, 287).

Protein dynamics can be incorporated into *dynamic substrate envelopes* for drug targets, which would provide additional guidance for developing drugs that are less affected by drug resistance (46). These calculations are a

systematic way of providing insight into the flexibility of substrates and the conformations of drug targets that are most prevalent and accessible to inhibitors. This analysis provides a more accurate representation of the substrate envelope compared to static crystal structures because it defines a probabilistic consensus volume distribution of which portions of the binding site are most occupied. This volume distribution can be compared to the volume distribution of inhibitors inside and outside of the dynamic substrate envelope and also the volume remaining in the envelope. These volumes can then be optimized to mimic natural substrates and minimize unnecessary protrusions from the dynamic substrate envelope (46).

6.7 Understanding additional drug resistance mutations in influenza neuraminidase N1 and N2 subtypes

In influenza NA, *dynamic substrate envelopes* explain different patterns of drug resistance in N1 and N2 subtypes for residues that directly contact inhibitors in the active site, such as E119 in N2 and I222 and S246 in N1. In addition, dynamic substrate envelopes provide insight into the main drug resistance mutation H274Y in N1 because this residue is directly in contact with E276 in the active site, which is makes van der Waals contacts specifically with oseltamivir. Results from these MD simulations show how many of the residues in the active site that are correlated with reduced susceptibility to NA inhibitors are more important for inhibitor binding compared to substrate binding. However, the substrate envelope hypothesis is not as effective for explaining mechanisms of

drug resistance for mutations that are not directly in the active site, multiple drug resistant variants, secondary permissive mutations, and mutations that involve changes in electrostatic interactions.

For instance, R292K is a drug resistance mutation in N2 that affects zanamivir, oseltamivir, and peramivir. This residue is important for both inhibitor and substrate contacts, and the currently understood mechanism of resistance is that R292K prevents rotation of residue E276, which leads to drug resistance in N2 (84). However, E276 seems to be important for both substrate and inhibitor binding in N2, and interactions with E276 are not specific for oseltamivir unlike in the N1 subtype. In addition, both inhibitors and substrates form hydrogen bonds with R292 in N2. Understanding the drug resistance mechanism of this residue and other more complex drug resistance mutations would require more in depth study, including molecular dynamics simulations of NA drug resistant variants to understand how the effects of specific mutations are propagated to the active site.

The existing crystal structures of NA are of the globular head domain, but the full neuraminidase enzyme consists of a cytoplasmic N-terminal tail, a transmembrane domain of approximately 30 amino acids, and a protein stalk connecting the transmembrane and globular head domains, which has approximately 40 amino acids. These domains of the enzyme have not yet been visualized at atomic resolution, but at least one drug resistance mutation has occurred in the stalk region, N70S in N1 (84). Although the globular head domain

is still active without the transmembrane or stalk domains, mutations in these domains, such as N70S, have been shown to impact NA activity and cause resistance. The globular head domain has also coevolved with the transmembrane domain (288). Therefore, further investigation into the structures of these domains in NA are needed to have a complete understanding of NA and how mutations in these additional domains may impact substrate cleavage, inhibitor binding, and the development of drug resistance mutations.

6.8 Understanding molecular recognition and antibody neutralization escape in influenza broadly neutralizing antibodies

Broadly neutralizing antibodies (BnAbs) against influenza have invigorated the influenza field with their potential to develop into universal therapies and vaccines that are protective against a broad spectrum of strains and subtypes (12). However, antibody neutralization escape mutations have emerged, and increasing our understanding of molecular recognition and the development of resistance mutations in these BnAbs is important for designing antibodies with reduced susceptibility to resistance. For instance, antibody neutralization escape mutations in stem directed BnAbs have emerged directly in the binding epitope and have been reported for C179, CR8020, and CR6261 (12, 58, 61, 64-67). Through our investigation of influenza evolution during viral passaging experiments in the presence of F10 as a selective pressure, we discovered that antibody neutralization escape mutations can occur outside of the binding epitope in locations that may shift the protein dynamics of HA and impact the

conformational changes in HA that occur during fusion. While several structural studies have been reported of BnAbs alone and in complex with HA (11, 12, 58, 60-62, 192), few studies have characterized antibody resistant HA variants. Further investigations are also needed to understand mechanisms of resistance at a molecular level, including both direct mechanisms, which occur at the binding epitope, and indirect mechanisms, which occur outside of the binding epitope.

In addition to using x-ray crystallography to study structures of BnAbs in complex with HA, cryo-electron microscopy (cryo-EM) is a technique that is gaining popularity and would also be useful in studying these structures. The main advantage of cryo-EM is that a molecule or system of interest does not need to be crystallized, but some limitations of cryo-EM are that systems need to be at least ~200 kDa or greater in size, and visualizing structures at resolutions approaching 2 Å, which is required for drug design, have only recently been accomplished (289). However, the size of HA in complex with 3 Fab fragments is approximately 330 kDa, and negative-stain EM has been used to study HA in complex with Fab fragments, so cryo-EM should be beneficial in the future in conjunction with x-ray crystallography (290).

Stem-directed BnAbs all bind to the conserved stem region of HA but in slightly different regions, and consequently these BnAbs have slightly different binding epitopes and neutralization profiles. Increasing our understanding of the different binding epitopes of BnAbs on the stem of HA, such as which residues

are most important for binding in each epitope, which residues are most conserved, where resistance mutations are most likely to develop, and how different resistance mutations would impact antibody binding, would be beneficial for engineering antibodies that are less susceptible to resistance and also are as broadly neutralizing as possible.

For instance, both F10 and D80 neutralize group 1 HA, but they are not effective in neutralizing group 2 HA or influenza B HA. However, antibodies have been discovered that are more broadly neutralizing. FI6v3 is a universal influenza A antibody, CR8033 and CR8071 are both universal influenza B antibodies, and CR9114 neutralizes all influenza A and B HA (11, 12, 291). Like in D80 and F10, the heavy chain of CR9114 is derived from the V_H1-69 germline gene, and CR9114 likely shares a similar epitope to F10 and CR6261, but subtle differences in the binding site on HA compared to D80 and F10 allow CR9114 to also bind influenza A group 2 HA and influenza B HA (11, 12). Increased flexibility in the CDR2 of the heavy chain allows CR9114 to accommodate differences in binding epitopes between group 1 HA and other HAs. Although CR9114 has been well studied, there are only three high-resolution structures of CR9114 in complex with HA. To have a better understanding of molecular recognition in CR9114 and other BnAbs and to identify potential resistance mutations, x-ray crystallography or cryo-EM should be attempted to determine additional structures of BnAbs in complex with different HA variants and potential antibody resistant variants. If resistance mutations develop, these structures

would inform the design of better antibodies that neutralize all influenza HA while avoiding resistance. Targeting regions that are evolutionarily conserved and important for function in HA may prevent resistance, but we have shown that escape mutations can still emerge in conserved regions of HA. In addition, we have shown that protein dynamics can play an important role in mechanisms of escape. Incorporating this knowledge into future antibody design may be challenging.

6.9 Impact of crystal contacts on D80 Fab fragment antigen binding site.

There are two molecules in the asymmetric unit for the D80 Fab fragment crystal structure, and on examination, the crystal contacts do not appear to impact the conclusion that the pre-epitope is pre-organized. For one molecule in the asymmetric unit, the crystal contacts are with the C-terminal portion of the constant domain of the light chain of D80, which is on the opposite end of the D80 Fab fragment from the antigen binding site on the heavy chain variable domain of D80. For the second molecule in the asymmetric unit, many of the crystal contacts are also with the C-terminal portion of the constant domain of the light chain of D80. There are also additional contacts with the beta-sheet body of the heavy chain in the variable domain of on D80, but there are no contacts with the antigen binding site on D80. To more clearly illustrate the crystal contacts in the structure of D80, a list of the exact crystal contacts can also be delineated and additional illustrative figures can be created.

6.10 Understanding binding of D80 to HA

Although the structure of D80 bound to HA has not yet been determined, there is some binding data available about how D80 binds to HA. In addition, the crystal structures of many different stem-directed broadly neutralizing antibody (BnAb) fragments bound to HA have been determined. In particular, a single chain Fv fragment of the BnAb F10, which is derived from the same germline gene VH1-69 as D80, has been crystallized in complex with H5 HA, and this crystal structure shows that F10 binds a highly conserved epitope on the stem region of HA, as shown in Chapter IV. Binding of both D80 and F10 to HA have been tested together in several experiments (both antibodies were discovered in the same laboratory), and these experiments show that D80 and F10 share an overlapping epitope on the stem region of HA. The binding epitope on HA is in a pocket on the HA stem formed by the HA_2 protomer fusion peptide with residues from the HA_1 protomer on one side and the αA helix of HA_2 on the other side. In one experiment, binding of various stem directed BnAbs was compared using a competition ELISA (enzyme-linked immunosorbent assay) between phage BnAbs and soluble BnAbs on plates coated with H5 HA (194). The phage BnAbs (10^{12} pfu) were mixed with 5 μg/mL of soluble BnAbs and added to the HA coated plates, which were then washed, and then followed up by HRP-anti-M13 (horseradish peroxidase conjugated to anti-M13 monoclonal antibody) for detecting phage-displayed antibodies by ELISA. The results of this experiment, shown in Figure 6.4, suggest that these BnAbs share an overlapping epitope. In

another experiment, binding of three stem directed BnAbs, F10, A66, and D8, which shares the same heavy chain as D80 but a different light chain, was tested against 14 different HA variants with mutations in the binding epitope (194). All of these antibodies bound similarly to the variants, and these results, shown in Figure 6.5D, also suggest that these antibodies share an overlapping epitope.

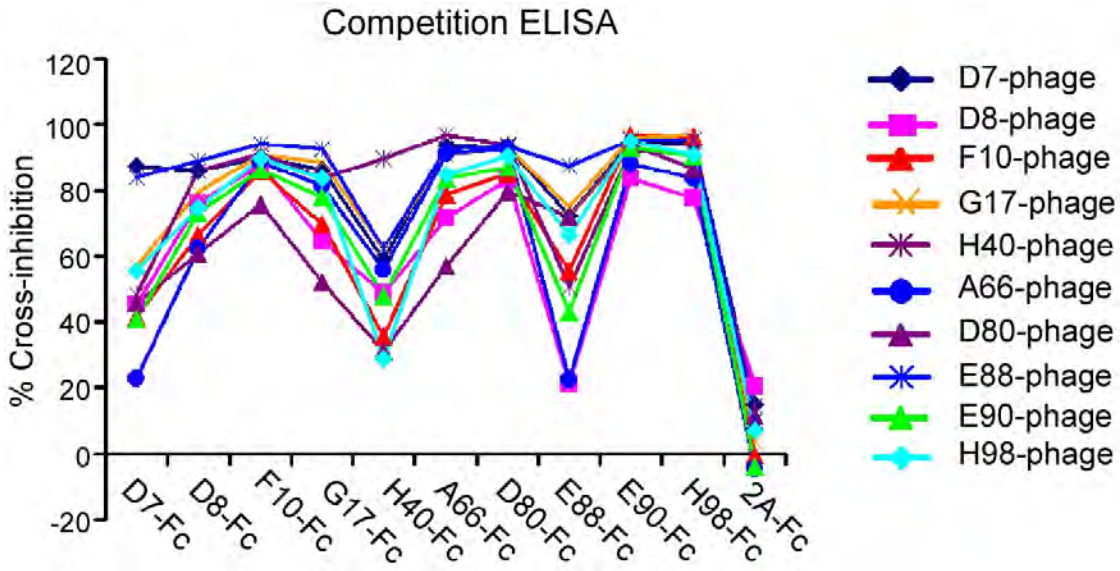


Figure 6.4 Competition ELISA assay between phage-bound broadly neutralizing antibodies and soluble broadly neutralizing antibodies.

Figure 6.4 Competition ELISA assay between phage-bound broadly neutralizing antibodies and soluble broadly neutralizing antibodies. Adapted from Figure S3 of Sui, et al (194).

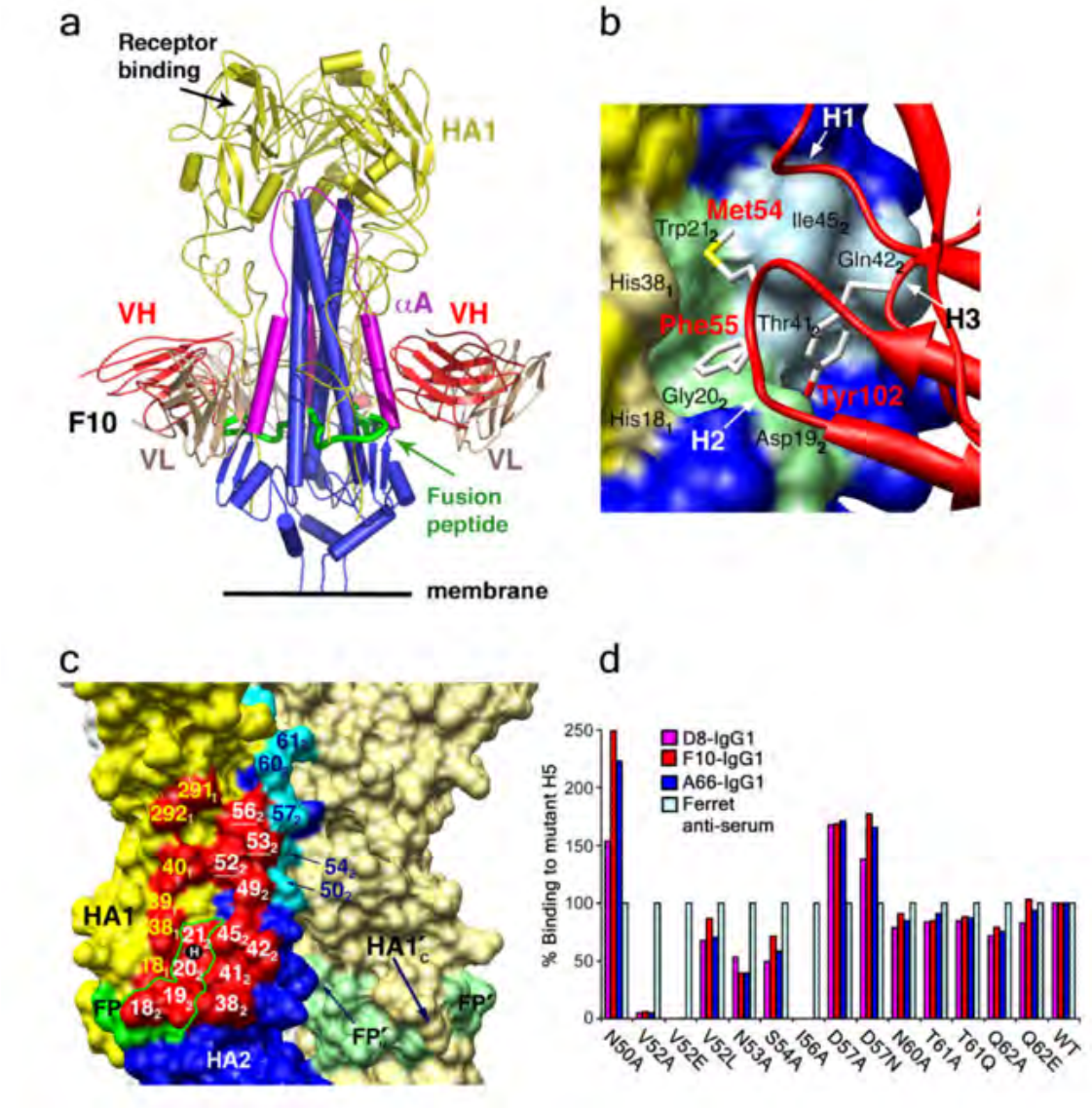


Figure 6.5 Analysis of crystal structure of H5 HA and F10 single chain Fv fragment and HA binding epitope.

Figure 6.5 Analysis of crystal structure of H5 HA and F10 single chain Fv fragment and HA binding epitope. Adapted from Figure 4 of Sui, et al (194). A) Structure of the H5 HA trimer bound to a single chain Fv fragment of F10. HA_1, HA_2, the αA of HA_2, the fusion peptide, and V_H and V_L of F10 are shown in yellow, blue, magenta, red, and gold, respectively. B) The binding epitope of F10 on H5 HA. H5 HA is shown in surface representation. C) Surface of the conserved stem region on HA where F10 binds. HA_1 and HA_2 are in yellow and blue. Residues involved in binding are in red, and residues in the epitope that do not impact binding are in cyan. D) Binding of three stem-directed broadly neutralizing antibodies, D8, F10, and A66 to variants of H5 HA with mutations in the binding epitope.

To also understand if the crystal structure of D80 bound to G6 may provide some insight into how D80 binds to HA, the binding epitopes of D80 on G6 and F10 on HA were compared. Although the epitope on HA is a shallow binding surface while the epitope on G6 is a deep pocket, one important similarity between the two epitopes is that there is a tryptophan in both G6 and HA ($W31_H$ in G6 and W_{21} in the fusion peptide of HA_2) that make extensive contacts with M54 of the antibody. However, there are significant differences in the CDR loop sequences of D80 compared to F10, and these differences may affect how D80 binds to HA compared to F10. For instance, F10 has two prolines in the heavy chain CDR2 that may limit the conformation of the CDR2 loop, and the CDR3 loop in F10 is also four residues longer than the CDR3 loop in D80, so understanding how the binding epitope may accommodate this longer loop and how the different CDR2 loops may fit in the binding site are important for understanding exactly how D80 binds to HA and how it compares to D80 binding to G6. Furthermore, several stem-directed BnAbs are derived from the VH1-69 germline gene, and while the un-mutated germline gene retains binding to G6, most BnAbs lose binding affinity to G6 as they gain binding affinity to HA, such as F10. Only three BnAbs are known to retain binding to both G6 and HA: D80, CR6331, and 70-1F02 (Figure 5.3). Although all of this information provides some initial insight into how D80 may bind to HA, further experiments are needed to clearly understand how D80 binds to HA, how it compares to other known

crystal structures of VH1-69 BnAbs bound to HA, and how this information relates to D80 binding to G6.

6.11 Using broadly neutralizing antibodies in influenza therapy

Research involving influenza BnAbs has mostly focused on various strategies for incorporating BnAbs into influenza therapy and inducing BnAbs through vaccines (12, 292). Priming with a DNA vaccine and boosting using a seasonal vaccine has successfully induced stem-directed BnAbs in ferrets and mice (12, 260). One challenge with eliciting stem-directed BnAbs through vaccines is that the stem region of HA is immunosubdominant, so these antibodies are not as easily induced as antibodies against the head region of HA. Some potential reasons why the head region is immunodominant is that the HA stem is not as accessible as the HA head region on the surface of the influenza virus, pre-existing immunological memory can bias the antibody response towards more head region antibodies, and HA stem reactive B cells are rare and can be polyreactive, limiting their development (293). However, further study is needed to definitively confirm these hypotheses in humans. HA immunogens that are “headless” have been designed to promote an antibody response to the stem. Recently, a vaccine using this strategy was developed that induces BnAbs against influenza A group 1 viruses, is protective in mice, and reduces fever in nonhuman primates (290). To overcome the influence of the immunodominant head region, HA immunogens have been designed with glycan shielding on the

head region (294). HA is glycosylated by host enzymes, and these glycans do not produce an immune response because they are recognized as “self” structures. Glycan shielding can be used to reduce an immune response and has been shown to reduce the immune response to the HIV envelope protein (295).

Structure-based rational immunogen design can also be used to focus an antibody response. Antibodyomics is an emerging field of study where sequencing transcripts from B cells are analyzed to understand how antibodies mature in response to immunogens (296). Many BnAbs such as D80, F10, and CR9114 are derived from the V_H1-69 germline gene, and each of these antibodies developed with a typical level of affinity maturation, so understanding how these antibodies evolved would inform immunogen design. In addition, it may be possible to focus an antibody response by using an immunogen that can bind specifically to the germline gene, such as humanized G6 (12, 239). Although we solved the structure of humanized G6 bound to the BnAb D80, the structure of D80 bound to HA is still unknown. Understanding how D80 and other antibodies are able to bind both G6 and HA would be important for understanding the mechanism of how G6 may be an effective immunogen. Some laboratories are currently using an iterative approach for immunogen design that includes structure based drug design and computational techniques for vaccine design in HIV-1 and other viruses, and this type of approach may also be useful in the design of immunogens that produce a broad response to influenza infection.

6.12 Concluding remarks

In conclusion, we used the substrate envelope hypothesis to provide insight into different patterns of drug resistance in influenza NA and HCV NS3/4A protease, and we used structural analyses to understand mechanisms of potential antibody neutralization escape mutations in influenza against the BnAb F10. In addition, we solved crystal structures of the Fab fragment of BnAb D80 alone and in complex with the Fab fragment of the anti-idiotypic antibody G6, which binds antibodies from the V_H1-69 germline gene. I hope that this work will provide helpful guidelines for the design of future antiviral therapies that are more effective and less susceptible to resistance and antibody neutralization escape mutations.

APPENDIX CHAPTER I

7 APPENDIX CHAPTER I

7.1 Molecular Dynamics Protocol

7.1.1 Protein Preparation Wizard Protocol

This utility processes the structures by assigning bond orders, adding hydrogens, creating zero-order bonds to metals, creating disulfide bonds, and filling in missing side chains using Prime. Next, tautomerization states were optimized using Epik, and hydrogen bond networks and protonation states were determined and optimized using PROPKA pH 7.0, with the options of using exhaustive sampling of water orientations and minimizing hydrogens of altered species. Finally, hydrogens were minimized using the Impact Refinement Module and the OPLS2005 force field.

7.1.2 Molecular Dynamics Minimization Protocol

For the initial minimization step, solute heavy atoms were restrained using a 1000 kcal mol^{-1} $Å^{2}$ force constant and a hybrid method of steepest descent for up to 10 steps and of the limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm for up to 2000 steps, with a convergence threshold of 50 kcal mol^{-1} $Å^{2}$. Next, the system was minimized in 7 stages with a harmonic restraint on all backbone atoms that was gradually reduced from 1000 to 1 kcal mol^{-1} $Å^{2}$ with 5000 steps for each stage using the hybrid steepest descent LBFGS method

(250 steps steepest descent and 4750 steps LBFGS). Lastly, a minimization without any restraints was performed, for a total of over ~40,000 minimization steps.

7.1.3 Molecular Dynamics Protocol

For the first short MD stage, a 10 ps MD simulation was performed in an NVT ensemble using a Berendsen thermostat at 10 K with a force constant of 50 kcal mol^{1} $Å^{2}$ on heavy atoms and velocity resampling every 1 ps. MD steps were integrated using 1 fs timesteps for bonded and non-bonded near interactions (van der Waals, short-range electrostatic interactions) within a 9 Å cutoff and 3 fs timesteps for long-range interactions (electrostatic interactions). A fast temperature relaxation constant of 0.1 was used. For the second and third short equilibration MD stages, a 10 ps MD simulation was performed in an NPT ensemble using a Berendsen thermostat and barostat at 10 K and 300 K, respectively. A pressure of 1 atm was used, the harmonic restraint was retained, and the velocity was resampled every 1 ps. A fast temperature relaxation constant of 0.1 and a slow pressure relaxation constant of 50.0 were used. MD steps were integrated using 2 fs timesteps for bonded and non-bonded interactions within a 9 Å cutoff and 6 fs timesteps for long-range interactions. For the final short equilibration stage, a 10 ps MD simulation was performed in an NPT ensemble using a Berendsen thermostat and barostat at 300 K with a pressure of 1 atm. A fast temperature relaxation constant of 0.1 and a normal

pressure relaxation constant of 2.0 were used. The harmonic restraint was decreased to 10 kcal mol ^{1} Å ^{2} on heavy atoms.

For the production stage, MD simulations were performed for 100 ns and 1 atm in the NPT ensemble using a Nose-Hoover thermostat and a Martyna-Tuckerman-Klein (MTK) barostat. Long-range electrostatics were calculated using the Particle Mesh Ewald method with a cutoff radius of 9 Å. The M-SHAKE algorithm was used to implement constraints that eliminate the highest frequency vibrational motions so that longer timesteps can be used (2 fs instead of 1 fs). Timesteps are scheduled using a parameter called RESPA (reference system propagator algorithm). MD steps were integrated using 2 fs timesteps for bonded and non-bonded near interactions (van der Waals, short-range electrostatic interactions) within a 9 Å cutoff and 6 fs timesteps for long-range interactions (electrostatic interactions). For each system, the trajectories of each monomer were concatenated to provide 400 ns of sampling.

APPENDIX CHAPTER II

8 APPENDIX CHAPTER II

8.1 Determination of D80 Fab Fragment Resolution Cutoff

We worked on solving the crystal structure of the D80 Fab fragment for approximately a year and a half. As a result of some difficulty optimizing cryoprotectant conditions for the D80 crystals, even though we had sent more than a dozen crystals to the synchrotron on two separate occasions, we were only able to collect one decent quality diffraction dataset for these crystals. The square edge of the data collected was to 2.7 Å, but the corners of the frames contained data out to 2.52 Å. Overall, I chose to deposit the structure of D80 that was processed and refined using data to 2.59 Å instead of 2.7 Å.

Although x-ray crystallography methodology has improved over time, specific criteria for the selection of resolution cutoff for a crystallographic data set is still debated and the relationship between data quality and model quality is not well understood. Traditionally, R_{merge} values, completeness, multiplicity, and signal to noise ratios are used to determine data quality and resolution cutoff. R_{merge} values indicate how well multiple measurements of a reflection throughout a dataset agree. The traditional criteria were that the data should have an overall R_{merge} value of less than 0.1, high resolution R_{merge} values less than 0.8 and a high resolution signal to noise ratio (I/σ) around 2. However, these types of

cutoffs can be too conservative and useful data can be discarded, as has been shown previously (297).

Karplus, et al instead recommend the use of Pearson's correlation coefficient (CC) to more accurately assess data quality and model agreement on the same scale and to decrease the omission of useful data that can be used to create better models (297). $CC^{1/2}$ is the correlation coefficient between two random halves of the unmerged data. Another measurement, CC*, estimates the correlation of the averaged dataset with the estimated true signal without noise by assuming that errors in the two half datasets are random and of similar size. In addition, the values CC_{work} and CC_{free} are the correlations of the experimental intensities with the intensities calculated from the refined model. A CC_{work} greater than CC* indicates overfitting and a CC_{free} less than CC* (observed at low resolution) implies that overfitting has not occurred and the model does not account for all of the signal from the data. At high resolution, a CC_{free} close to CC* indicates that the model is limited by data quality.

Using these criteria and by examining refinements of the data with different resolution limits, following the method described by Karplus, et al, I evaluated three different models of the D80 Fab fragment (297). I processed and refined the data at 2.7, 2.59, and 2.52 Å resolution, and I calculated crystallographic statistics, including overall R_{work} and R_{free} values at lower resolution limits than what was used during refinement. To perform these calculations, I used the commands in Phenix described on page 5 of the

supplemental material of Karplus, et al (297). Table 8.1 shows the overall R_{work}/R_{free} for the D80 Fab fragment dataset, analogous to Table S2 from Karplus, et al (297). Table 8.2 shows a similar representation for $CC^{1/2}$, CC*, CC_{work}, and CC_{free}.

Overall, I chose to deposit the structure of D80 that was processed and refined using data to 2.59 Å instead of 2.7 Å that I originally proposed in the thesis. Tables 8.3, 8.4, and 8.5 show the results of the paired refinement. For the refinement at 2.7 Å, an unmerged data file was not created at the time of data processing, and I did not re-process the data to get the unmerged data file at 2.7 Å, so some crystallographic statistics are missing from Table 8.2. Using the data to 2.59 Å improved the model, but the data out to 2.52 Å did not seem to improve the model.

Table 8.1 Overall R_{work}/R_{free} for D80 Refinement

	High-resolution limit for refinement (Å)		
High-resolution limit for R value calculation (Å)	2.7 Å	2.59 Å	2.52 Å
2.7 Å	0.2194 / 0.2632	0.2163 / 0.2592	0.2342 / 0.2612
2.59 Å	-	0.2212 / 0.2629	0.2403 / 0.2680
2.52 Å	-	-	0.2447 / 0.2730

Table 8.2 CC^*, $CC^{1/2}$, CC_{work}, and CC_{free} for D80 Refinement

	High-resolution limit for refinement (Å)		
High-resolution limit for R value calculation (Å)	2.7 Å	2.59 Å	2.52 Å
2.7 Å	-	0.996 (0.868) 0.999 (0.964) 0.919 (0.782) 0.950 (0.812)	0.993 (0.74) 0.998 (0.922) 0.916 (0.687) 0.877 (0.657)
2.59 Å	-	0.996 (0.746) 0.999 (0.925) 0.921 (0.661) 0.950 (0.584)	0.993 (0.543) 0.998 (0.839) 0.918 (0.604) 0.877 (0.538)
2.52 Å	-	-	0.992 (0.429) 0.998 (0.775) 0.919 (0.516) 0.877 (0.344)

Table 8.3 High resolution limit for refinement set at 2.7 Å
Original Structure Presented in Thesis; No Unmerged Data File;
Data Indexed and Scaled Using HKL3000 and Structure Refined Using Phenix

Structure	D80 Fab Fragment
Resolution Range	46.12 – 2.70 (2.80-2.70)
Space Group	P 21 21 21
Unit Cell	89.06 91.58 106.76 90 90 90
Total Reflections	116269
Unique Reflections	23313 (2316)
Multiplicity	5.0 (4.9)
Completeness (%)	0.95 (0.92)
Mean I/Sigma (I)	17.7 (3.8)
Wilson B-Factor	47.05
R-merge	0.083
Reflections Used in Refinement	23312 (2316)
Reflections Used for R-Free	1126 (113)
R-work	0.2194 (0.2819)
R-free	0.2632 (0.3689)
Number of Non-Hydrogen Atoms	6395
Macromolecules	6330
Ligands	N/A
Protein Residues	851
RMS(Bonds)	0.016
RMS(Angles)	1.43
Ramachandran Favored (%)	97
Ramachandran Allowed (%)	2.5
Ramachandran Outliers (%)	0.24
Rotamer Outliers (%)	3.6
Clash score	6.72
Average B-Factor	48.50
Macromolecules	48.64
Ligands	N/A
Solvent	35.40

Table 8.4 High resolution limit for refinement set at 2.59 Å
Data indexed and scaled using HKL3000 and structure refined using Phenix

Limit for R Value Calculation (Å)	2.59	2.7
PDB Code	5JQD	N/A
Resolution Range	46.12 – 2.59 (2.68 – 2.59)	46.12 – 2.7 (2.797 – 2.7)
Space Group	P 21 21 21	P 21 21 21
Unit Cell	89.06 91.58 106.76 90 90 90	89.06 91.58 106.76 90 90 90
Total Reflections	132762 (9946)	120917 (12012)
Unique Reflections	26405 (2602)	23382 (2343)
Multiplicity	5.0 (3.8)	5.2 (5.1)
Completeness (%)	0.95 (0.95)	0.95 (0.97)
Mean I/Sigma (I)	11.16 (2.42)	12.27 (3.62)
Wilson B-Factor	46.81	47.06
R-merge	0.09 (0.52)	0.09 (0.42)
R-meas	0.10 (0.61)	0.10 (0.46)
CC1/2	0.996 (0.746)	0.996 (0.868)
CC*	0.999 (0.925)	0.999 (0.964)
Reflections Used in Refinement	26404 (2602)	23381 (2343)
Reflections Used for R-Free	1273 (127)	1129 (114)
R-work	0.2152 (0.2891)	0.2163 (0.2863)
R-free	0.2438 (0.2981)	0.2592 (0.3465)
CC(work)	0.93 (0.75)	0.919 (0.782)
CC(free)	0.96 (0.69)	0.950 (0.812)
Number of Non-Hydrogen Atoms	6431	6403
Macromolecules	6338	6338
Ligands	0	0
Protein Residues	851	851
RMS(Bonds)	0.005	0.003
RMS(Angles)	0.76	0.74
Ramachandran Favored (%)	97	97
Ramachandran Allowed (%)	3.2	3.4
Ramachandran Outliers (%)	0.24	0.12
Rotamer Outliers (%)	3.5	2.5
Clash score	7.11	2.51
Average B-Factor	48.57	50.28
Macromolecules	48.70	50.38
Ligands	N/A	N/A
Solvent	40.05	40.76

Table 8.5 High resolution limit for refinement set at 2.52 Å
Data indexed and scaled using XIA2 and structure refined using Phenix

Limit for R Value Calculation (Å)	2.52	2.59	2.7
Resolution Range	69.53 – 2.52 (2.61 – 2.52)	69.53 – 2.59 (2.683 – 2.59)	69.53 – 2.7 (2.797 – 2.7)
Space Group	P 21 21 21	P 21 21 21	P 21 21 21
Unit Cell	89.11 91.59 106.81 90 90 90	89.11 91.59 106.81 90 90 90	89.11 91.59 106.81 90 90 90
Total Reflections	137016 (8633)	130503 (9576)	118980 (11567)
Unique Reflections	28583 (2727)	26470 (2610)	23399 (2330)
Multiplicity	4.8 (3.2)	4.9 (3.7)	5.1 (5.0)
Completeness (%)	0.94 (0.91)	0.94 (0.95)	0.95 (0.97)
Mean I/Sigma (I)	10.14 (2.45)	10.76 (3.04)	11.76 (4.33)
Wilson B-Factor	29.64	31.09	33.13
R-merge	0.1314 (0.6204)	0.127 (0.553)	0.1186 (0.464)
R-meas	0.1473 (0.7332)	0.142 (0.6418)	0.1322 (0.5189)
CC1/2	0.992 (0.429)	0.993 (0.543)	0.993 (0.74)
CC*	0.998 (0.775)	0.998 (0.839)	0.998 (0.922)
Reflections Used in Refinement	28332 (2622)	26300 (2547)	23302 (2299)
Reflections Used for R-Free	1981 (189)	1838 (179)	1629 (166)
R-work	0.2447 (0.3523)	0.2403 (0.3360)	0.2342 (0.3032)
R-free	0..2730 (0.3914)	0.2680 (0.3695)	0.2612 (0.3268)
CC(work)	0.919 (0.516)	0.918 (0.604)	0.916 (0.687)
CC(free)	0.877 (0.344)	0.877 (0.538)	0.877 (0.657)
Number of Non-Hydrogen Atoms	6525	6525	6525
Macromolecules	6285	6285	6285
Ligands	N/A	N/A	N/A
Protein Residues	849	849	849
RMS(Bonds)	0.002	0.004	0.002
RMS(Angles)	0.66	0.98	0.66
Ramachandran Favored (%)	94	94	94
Ramachandran Allowed (%)	5.2	5.2	5.2
Ramachandran Outliers (%)	0.72	0.72	0.72
Rotamer Outliers (%)	1.5	1.5	1.5
Clash score	7.02	7.02	7.02
Average B-Factor	38.71	38.71	38.71
Macromolecules	39.02	39.02	39.02
Ligands	N/A	N/A	N/A
Solvent	30.65	30.65	30.65

APPENDIX CHAPTER III

9 APPENDIX CHAPTER III

9.1 Hepatitis C NS3/4A Protease Inhibitor Ki Determination Assay: Data and Modeling

Using an assay described elsewhere for determining the inhibitor constant (Ki) for HCV NS3/4A protease, I determined the Ki values of 7 novel HCV protease inhibitors synthesized in the Schiffer laboratory with wild-type HCV NS3/4A protease genotype 1a (175). These inhibitors included AH-58, AH-79, AH-82, EK-31, JA-30, JA-43, and JA-44. Djade Soumana expressed and purified the protein, and I performed the assay and analyzed the data. The results of this work is shown in Figure 9.1. In addition, I used the Schrodinger Software Suite to model additional 6 novel inhibitors synthesized in the Schiffer laboratory bound to HCV NS3/4A protease, and I performed 10 nanosecond molecular dynamics simulations of these models using Desmond. These inhibitors included AH-54, AH-58, JA-43, AH-82, AH-87, and an isopropyl analog of AH-87. The results of this work is shown in Figure 9.2. The chemical structures of these inhibitors is shown in Figure 9.3.

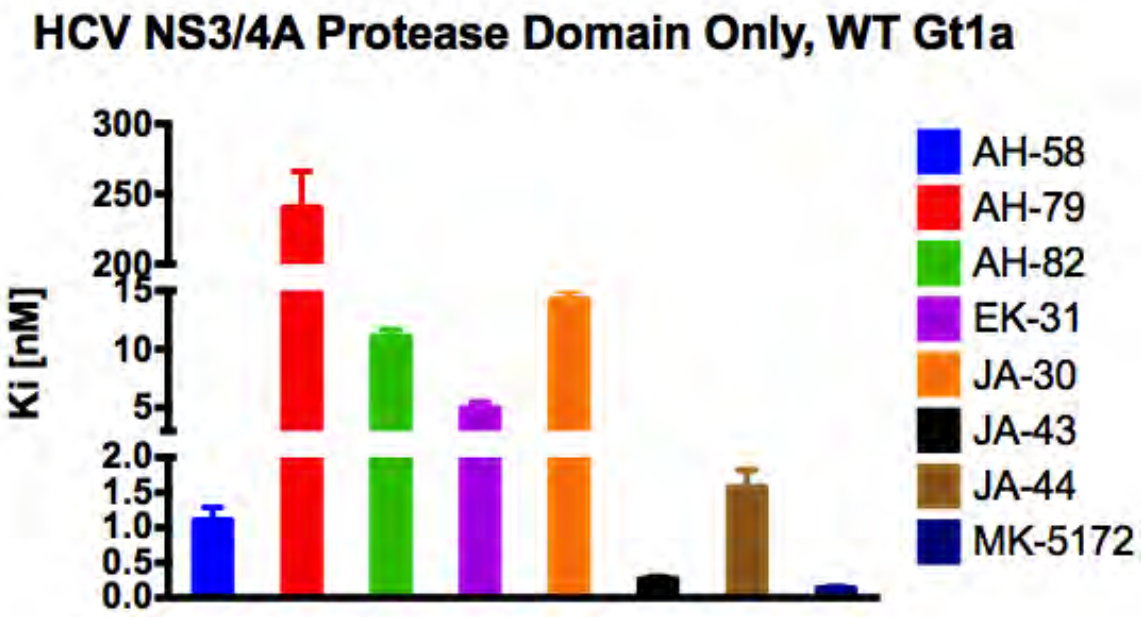


Figure 9.1 Ki values of seven novel HCV NS3/4A protease inhibitors with wild-type genotype 1a HCV NS3/4A protease.

Figure 9.1 Ki values of seven novel HCV NS3/4A protease inhibitors with wild-type genotype 1a HCV NS3/4A protease.

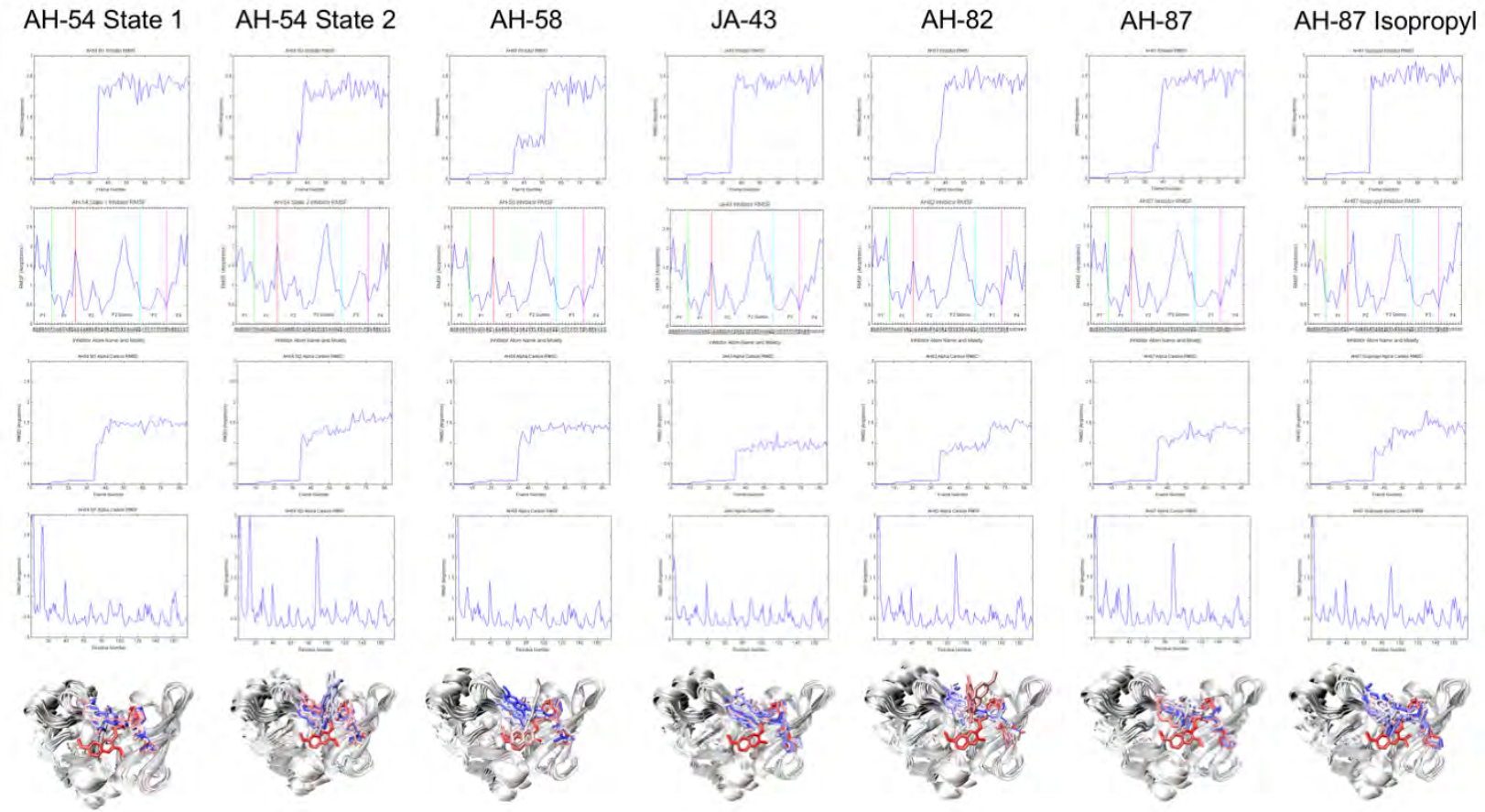


Figure 9.2 Molecular dynamics simulation results for models of HCV NS3/4A protease bound to novel inhibitors.

Figure 9.2 Molecular dynamics simulation results for models of HCV NS3/4A protease bound to novel inhibitors. Each column of results corresponds to the analysis from each model of HCV NS3/4A protease bound to a different inhibitor. From left to right, the inhibitors that were analyzed were AH-54 with a protonated amide nitrogen adjacent to the P1' sulfonamide (AH-54 State 1), AH-54 with a deprotonated amide nitrogen adjacent to the P1' sulfonamide (AH-54 State 2), AH-58, JA-43, AH-82, AH-87, and an isopropyl analog of AH-87. The top row shows the alpha carbon RMSD over the course of the minimization, equilibration, and molecular dynamics simulation of each model. The second row from the top shows the average ligand RMSF over the course of the molecular dynamics simulation. The third row from the top shows the ligand RMSD over the course of the minimization, equilibration, and molecular dynamics simulation of each model. The fourth row from the top shows the average alpha carbon RMSF over the course of the molecular dynamics simulation, and the bottom row shows snapshots of the model over the course of the molecular dynamics simulation.

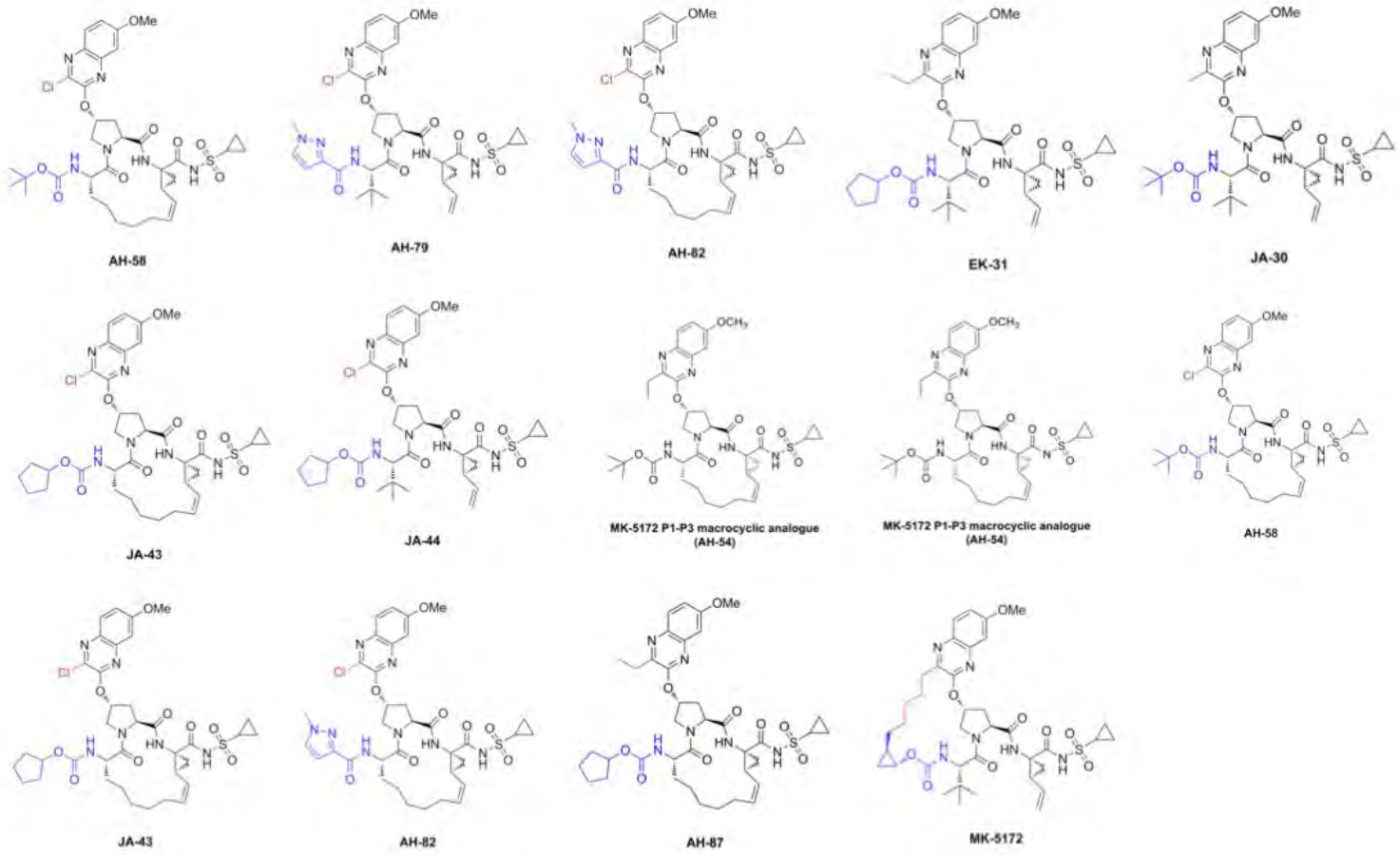


Figure 9.3 Chemical structures of novel HCV NS3/4A protease inhibitors.

Figure 9.3 Chemical structures of novel HCV NS3/4A protease inhibitors. Aspects of the inhibitors in the P2' and P4 regions that are variable are shown in red and blue, respectively.

APPENDIX CHAPTER IV

10 APPENDIX CHAPTER IV

10.1 Influenza Neuraminidase Baculovirus Insect Cell Expression and Purification

Using an insect cell expression system with baculovirus expression technology, I successfully expressed and purified N1 NA. The NA construct was prepared based on published methods (106, 107, 298, 299). cDNA corresponding to the NA ectodomain, residues 82 to 467, was cloned into baculovirus transfer vector pFastBac1 (Bac-to-Bac^{TM} kit from Invitrogen) with a GP67 signal peptide, a 6X-histidine tag, a tetramerization domain, and a thrombin cleavage site at the N-terminal end of NA. Suspension cultures of insect cell Sf9 and Hi5 cells were cultured in Sf-900 II SFM serum free media (GIBCO) and HyQ SFX-Insect media (HyClone), respectively. Transfection was performed using Sf9 cells, and expression was performed using Hi5 cells. Recombinant baculovirus was prepared according to the manufacturer's protocol (Invitrogen). Protein was expressed in Hi5 suspension cultures for 3 days at 28°C and 120 rpm with a MOI of 5, and preparation was performed on a 1 L scale. Hi5 cells were removed by centrifugation, and soluble NA was recovered from cell supernatant by Ni^{2+} affinity chromatography. The eluted protein was analyzed using SDS-PAGE (Figure 10.1) and a Western blot using an anti-His tag antibody

(Figure 10.2). The activity of this protein was also assessed using a cleavage assay described in Appendix V.

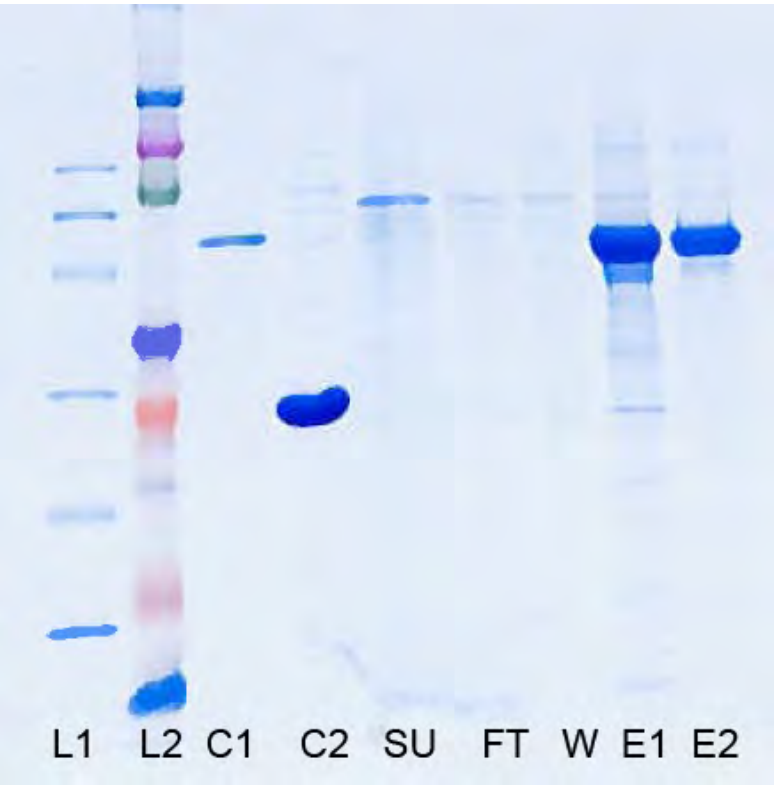


Figure 10.1 Expression and purification of N1 NA using a nickel-NTA column analyzed using SDS-PAGE.

Figure 10.1 Expression and purification of N1 NA using a nickel-NTA column analyzed using SDS-PAGE. The abbreviations represent low molecular weight ladder (L1), Kaleidoscope ladder (L2), control NA from the Biodefense and Emerging Infections Research Resources Repository (C1), control HCV NS3/4A protease (C2), virus supernatant (SU), flow through (FT), wash (W), elution followed by TCA precipitation (E1), and elution follow by dialysis and concentration (C2).

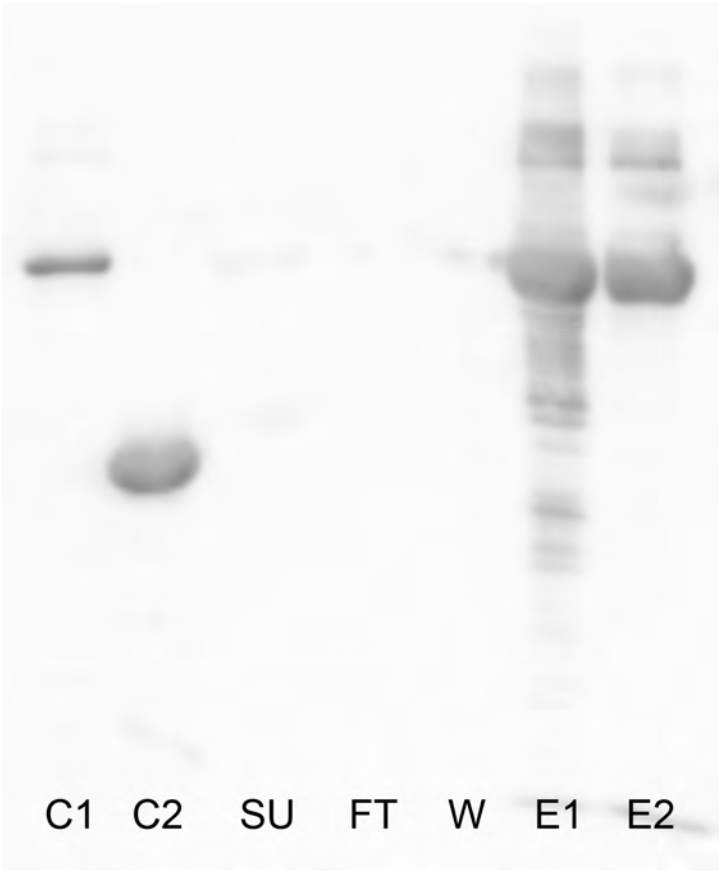


Figure 10.2 Expression and purification of N1 NA using a nickel-NTA column analyzed by Western blot using an anti-His-tag antibody.

Figure 10.2 Expression and purification of N1 NA using a nickel-NTA column analyzed by Western blot using an anti-His-tag antibody. The abbreviations represent low molecular weight ladder (L1), Kaleidoscope ladder (L2), control NA from the Biodefense and Emerging Infections Research Resources Repository (C1), control HCV NS3/4A protease (C2), virus supernatant (SU), flow through (FT), wash (W), elution followed by TCA precipitation (E1), and elution follow by dialysis and concentration (C2).

APPENDIX CHAPTER V

11 APPENDIX CHAPTER V

11.1 Influenza Neuraminidase 1D NMR Cleavage Assay

With assistance from Akbar Ali in the Schiffer lab, I implemented a previously published one-dimensional (1D) proton NMR assay performed over time to detect the formation of free sialic acid during a neuraminidase cleavage reaction (300, 301). Proton NMR can be used to monitor sialic acid release from glycans since the protons on free sialic acid have different chemical shifts compared to the protons on sialic acid bound to glycans. Figure 11.1 shows a previously published example of this experiment using sialyllactose, a trisaccharide (301).

To understand the results that would be expected from this experiment, the changes in the proton NMR signal must be described. First, the signal from the full substrate can be measured at time zero before cleavage occurs. The signals from the protons on the sialic acid of the substrate are between 2.7 and 2.8 ppm, and that signal corresponds to the signal from the equatorial hydrogen on C3 of sialic acid. There is also a strong peak around 2 ppm that corresponds to the signal from the methyl hydrogens on the N-acetyl group of sialic acid. Finally, there is a signal around 1.7 ppm that corresponds to the signal from the axial hydrogen on C3 of sialic acid.

Then, neuraminidase can be added to the substrate solution, and how these proton signals evolve over time can be observed. Over the course of the experiment, the signal from the N-acetyl group shifts and the signal splits. In addition, signals from the equatorial and axial C3 hydrogens of free sialic acid emerge at 2.2 ppm and 1.8 ppm, respectively, and become stronger over time.

I was able to replicate this experiment with commercial bacterial neuraminidase from Clostridium perfringens produced by New England Biolabs. Two 1D proton NMR scans are shown in Figure 11.2A, one at time zero with substrate only and one at 36 hours after neuraminidase was added to the substrate solution and the substrate was allowed to incubate and be cleaved by neuraminidase for 36 hours. This assay was performed with a pentasaccharide LST-c substrate. With this compound, the N-acetyl peak is split initially, and it splits more over the course of the reaction. In addition, the peaks from the free sialic acid hydrogens at 2.2 ppm and 1.8 ppm emerge and are very clear signals.

Figure 11.2B shows a second analogous experiment where two scans were taken at time zero and time 90 minutes, and a full glycoprotein was used as the substrate, α-1 acid glycoprotein. The signal is not as prominent as with the pentasaccharide or the trisaccharide, but over time the signal from the free sialic acid emerges.

Finally, this experiment was performed with neuraminidase that was expressed and purified in the laboratory and was described in Appendix IV, and the results of this experiment are shown in Figure 11.3B and compared to the

results from Figure 11.2B (Figures 11.2B and 11.3A are equivalent). The cleavage results using the in-house neuraminidase are not as clear as the results from the experiments with the commercial bacterial neuraminidase, but the signal from the N-acetyl peak appears to split, and this data suggests that the protein shown in Appendix IV is folded and active.

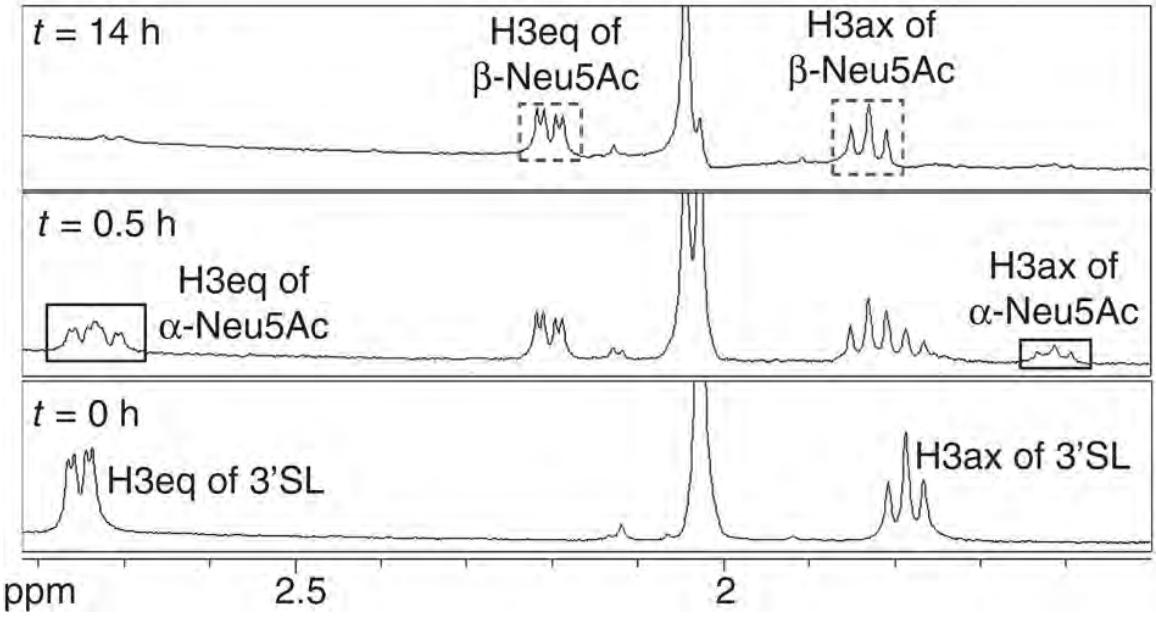


Figure 11.1 A previously published example of using proton NMR to detect neuraminidase substrate cleavage using sialyllactose, a trisaccharide (301).

Figure 11.1 A previously published example of using proton NMR to detect neuraminidase substrate cleavage using sialyllactose, a trisaccharide (300, 301)

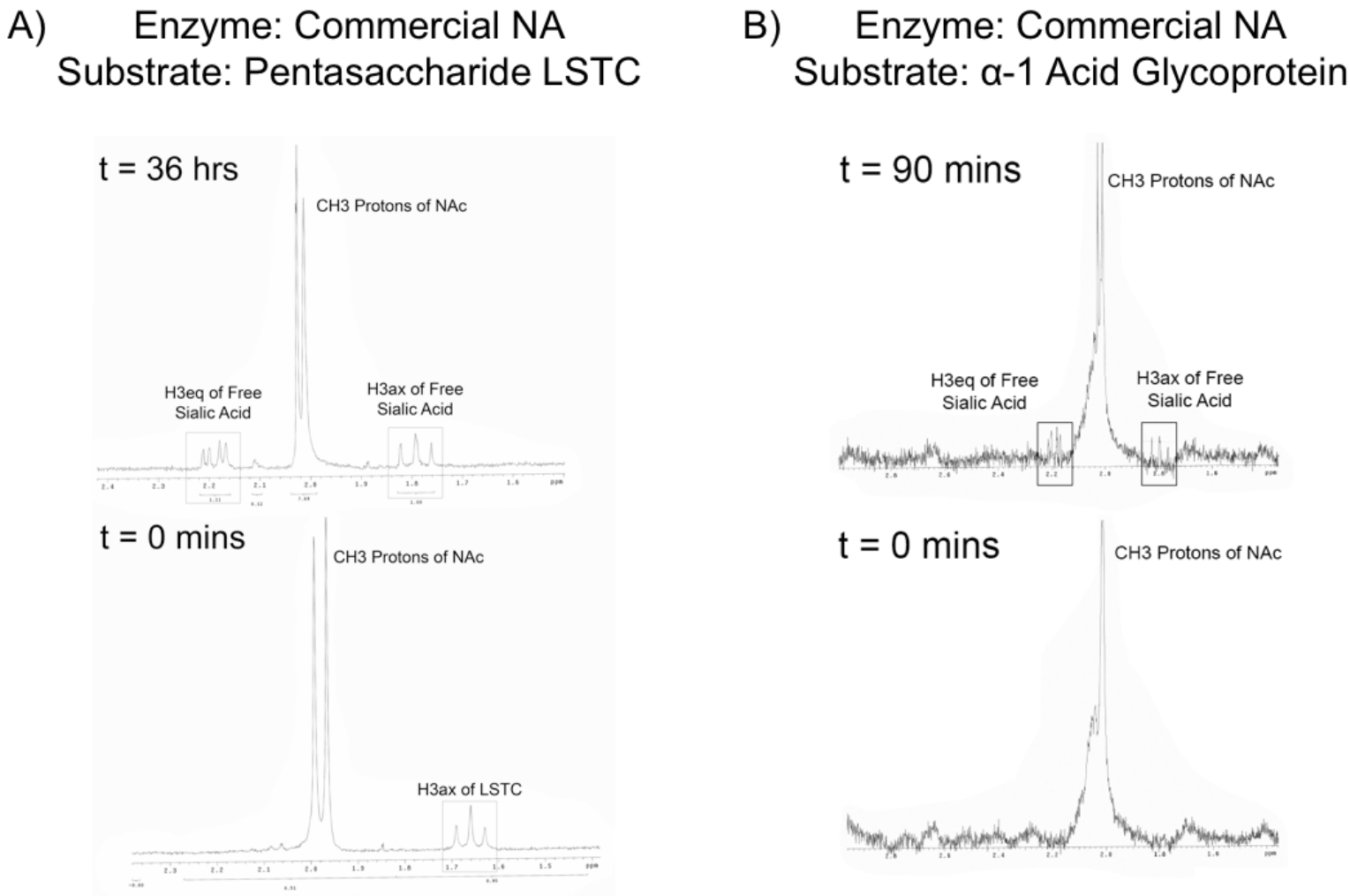


Figure 11.2 1D proton NMR neuraminidase cleavage assay performed with commercially available neuraminidase (NA) from New England Biolabs.

Figure 11.2 1D proton NMR neuraminidase cleavage assay performed with commercially available neuraminidase (NA) from New England Biolabs. A) This experiment was performed with the pentasaccharide substrate LST-c. B) This experiment was performed with the glycoprotein substrate α-1 acid glycoprotein.

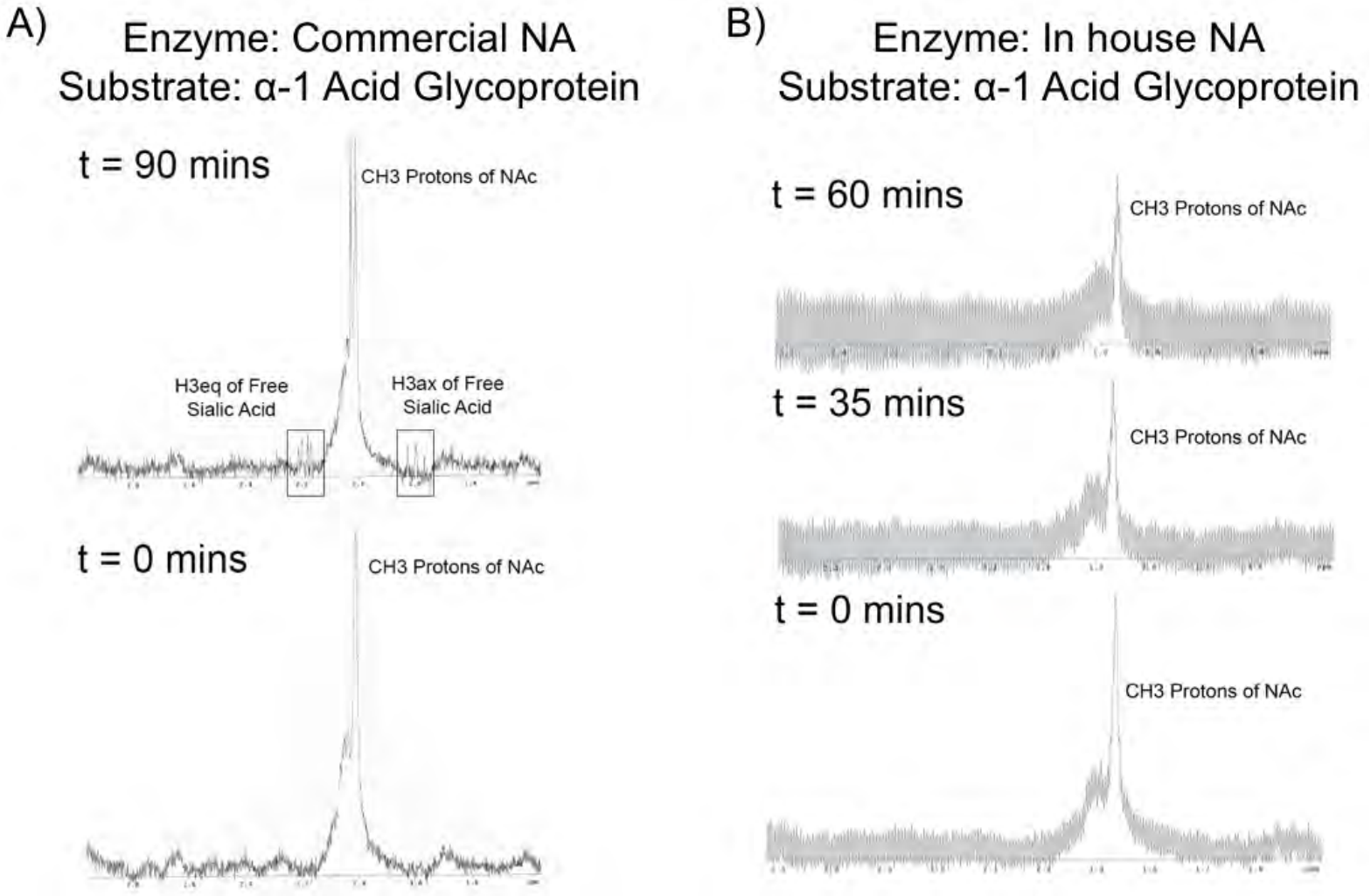


Figure 11.3 1D proton NMR neuraminidase cleavage assay performed with the substrate α-1 acid glycoprotein and two different neuraminidases (NA).

Figure 11.3 1D proton NMR neuraminidase cleavage assay performed with the substrate α-1 acid glycoprotein and two different neuraminidases (NA). A) This experiment was performed with commercially available neuraminidase from New England Biolabs. B) This experiment was performed with influenza neuraminidase produced in house and described in Appendix IV.

12 BIBLIOGRAPHY

1. **Geier MR, Geier DA.** 2002. The state of polio vaccination in the world: the case for continuing routine vaccination. Toxicol Mech Methods **12:**221-228.
2. **Riedel S.** 2005. Edward Jenner and the history of smallpox and vaccination. Proc (Bayl Univ Med Cent) **18:**21-25.
3. **Miller E.** 2002. MMR vaccine: review of benefits and risks. J Infect **44:**1-6.
4. **Osterholm MT, Kelley NS, Sommer A, Belongia EA.** 2012. Efficacy and effectiveness of influenza vaccines: a systematic review and meta-analysis. Lancet Infect Dis **12:**36-44.
5. **Moscona A.** 2005. Neuraminidase inhibitors for influenza. N Engl J Med **353:**1363-1373.
6. **Brooks M.** August 1, 2014 2014. Top 100 Most Prescribed, Top-Selling Drugs. Medscape.
7. **Wikipedia Contributors.** 2016. Antiviral Drug, 2 March 2016 03:05 UTC ed. Wikipedia, The Free Encyclopedia.
8. **Richman DD.** 2006. Antiviral drug resistance. Antiviral Res **71:**117-121.
9. **von Itzstein M.** 2007. The war against influenza: discovery and development of sialidase inhibitors. Nat Rev Drug Discov **6:**967-974.
10. **American Society of Health-System Pharmacists, Inc.** 2013. Ribavirin, *on* American Society of Heatlh-System Pharmacists, Inc.

https://www.nlm.nih.gov/medlineplus/druginfo/meds/a605018.html - why. Accessed March 16, 2016.

11. **Dreyfus C, Laursen NS, Kwaks T, Zuijdgeest D, Khayat R, Ekiert DC, Lee JH, Metlagel Z, Bujny MV, Jongeneelen M, van der Vlugt R, Lamrani M, Korse HJ, Geelen E, Sahin O, Sieuwerts M, Brakenhoff JP, Vogels R, Li OT, Poon LL, Peiris M, Koudstaal W, Ward AB, Wilson IA, Goudsmit J, Friesen RH.** 2012. Highly conserved protective epitopes on influenza B viruses. Science **337:**1343-1348.
12. **Laursen NS, Wilson IA.** 2013. Broadly neutralizing antibodies against influenza viruses. Antiviral Res **98:**476-483.
13. **Doyle TM, Hashem AM, Li C, Van Domselaar G, Larocque L, Wang J, Smith D, Cyr T, Farnsworth A, He R, Hurt AC, Brown EG, Li X.** 2013. Universal anti-neuraminidase antibody inhibiting all influenza A subtypes. Antiviral Res **100:**567-574.
14. **Scheel TK, Rice CM.** 2013. Understanding the hepatitis C virus life cycle paves the way for highly effective therapies. Nat Med **19:**837-849.
15. **Martin P.** February 15, 2001 2001. Update on Hepatitis C Treatment. Medscape Gastroenterology. **3**(1)**:**
16. **Chopra SP, P.J.** 2015. Overview of management of chronic hepatitis C virus infection. UpToDate.
17. **Lok AS.** 2015. Overview of the management of hepatitis B and case examples. UpToDate.

18. **Wikipedia Contributors.** 2016. Interferon, 5 March 2016 14:16 UTC ed. Wikipedia, The Free Encyclopedia.

19. **Kelleher TBA, N.H.** 2014. Management of the side effects of peginterferon and ribavirin used for treatment of chronic hepatitis C virus infection. UpToDate.

20. **Kamali A, Holodniy M.** 2013. Influenza treatment and prophylaxis with neuraminidase inhibitors: a review. Infect Drug Resist **6:**187-198.

21. **Horscroft NJ, Pryde DC, Bright H.** 2012. Antiviral applications of Toll-like receptor agonists. J Antimicrob Chemother **67:**789-801.

22. **Graziani AL.** 2014. Pharmacology of non-nucleoside reverse transcriptase inhibitors. UpToDate.

23. **Flexner C.** 1998. HIV-protease inhibitors. N Engl J Med **338:**1281-1292.

24. **Tsantrizos YS.** 2004. The design of a potent inhibitor of the hepatitis C virus NS3 protease: BILN 2061--from the NMR tube to the clinic. Biopolymers **76:**309-323.

25. **Jiang Y, Andrews SW, Condroski KR, Buckman B, Serebryany V, Wenglowsky S, Kennedy AL, Madduru MR, Wang B, Lyon M, Doherty GA, Woodard BT, Lemieux C, Geck Do M, Zhang H, Ballard J, Vigers G, Brandhuber BJ, Stengel P, Josey JA, Beigelman L, Blatt L, Seiwert SD.** 2014. Discovery of danoprevir (ITMN-191/R7227), a highly selective and potent inhibitor of hepatitis C virus (HCV) NS3/4A protease. J Med Chem **57:**1753-1769.

26. **Scola PM, Sun LQ, Wang AX, Chen J, Sin N, Venables BL, Sit SY, Chen Y, Cocuzza A, Bilder DM, D'Andrea SV, Zheng B, Hewawasam P, Tu Y, Friborg J, Falk P, Hernandez D, Levine S, Chen C, Yu F, Sheaffer AK, Zhai G, Barry D, Knipe JO, Han YH, Schartman R, Donoso M, Mosure K, Sinz MW, Zvyaga T, Good AC, Rajamani R, Kish K, Tredup J, Klei HE, Gao Q, Mueller L, Colonno RJ, Grasela DM, Adams SP, Loy J, Levesque PC, Sun H, Shi H, Sun L, Warner W, Li D, Zhu J, Meanwell NA, McPhee F.** 2014. The discovery of asunaprevir (BMS-650032), an orally efficacious NS3 protease inhibitor for the treatment of hepatitis C virus infection. J Med Chem **57:**1730-1752.

27. **Surleraux DL, Tahri A, Verschueren WG, Pille GM, de Kock HA, Jonckers TH, Peeters A, De Meyer S, Azijn H, Pauwels R, de Bethune MP, King NM, Prabu-Jeyabalan M, Schiffer CA, Wigerinck PB.** 2005. Discovery and selection of TMC114, a next generation HIV-1 protease inhibitor. J Med Chem **48:**1813-1822.

28. **Lin C, Kwong AD, Perni RB.** 2006. Discovery and development of VX-950, a novel, covalent, and reversible inhibitor of hepatitis C virus NS3.4A serine protease. Infect Disord Drug Targets **6:**3-16.

29. **Lew W, Chen X, Kim CU.** 2000. Discovery and development of GS 4104 (oseltamivir): an orally active influenza neuraminidase inhibitor. Curr Med Chem **7:**663-672.

30. **Sidwell RW, Smee DF.** 2002. Peramivir (BCX-1812, RWJ-270201): potential new therapy for influenza. Expert Opin Investig Drugs **11:**859-869.

31. **Woods JM, Bethell RC, Coates JA, Healy N, Hiscox SA, Pearson BA, Ryan DM, Ticehurst J, Tilling J, Walcott SM, et al.** 1993. 4-Guanidino-2,4-dideoxy-2,3-dehydro-N-acetylneuraminic acid is a highly effective inhibitor both of the sialidase (neuraminidase) and of growth of a wide range of influenza A and B viruses in vitro. Antimicrob Agents Chemother **37:**1473-1479.

32. **Strasfeld L, Chou S.** 2010. Antiviral drug resistance: mechanisms and clinical implications. Infect Dis Clin North Am **24:**413-437.

33. **King NM, Prabu-Jeyabalan M, Nalivaika EA, Schiffer CA.** 2004. Combating susceptibility to drug resistance: lessons from HIV-1 protease. Chem Biol **11:**1333-1338.

34. **Ragland DA, Nalivaika EA, Nalam MN, Prachanronarong KL, Cao H, Bandaranayake RM, Cai Y, Kurt-Yilmaz N, Schiffer CA.** 2014. Drug resistance conferred by mutations outside the active site through alterations in the dynamic and structural ensemble of HIV-1 protease. J Am Chem Soc **136:**11956-11963.

35. **Kolli M, Ozen A, Kurt-Yilmaz N, Schiffer CA.** 2014. HIV-1 protease-substrate coevolution in nelfinavir resistance. J Virol **88:**7145-7154.

36. **Ozen A, Lin KH, Kurt Yilmaz N, Schiffer CA.** 2014. Structural basis and distal effects of Gag substrate coevolution in drug resistance to HIV-1 protease. Proc Natl Acad Sci U S A **111:**15993-15998.

37. **Wagner R, Matrosovich M, Klenk HD.** 2002. Functional balance between haemagglutinin and neuraminidase in influenza virus infections. Rev Med Virol **12:**159-166.

38. **Ozer N, Ozen A, Schiffer CA, Haliloglu T.** 2015. Drug-resistant HIV-1 protease regains functional dynamics through cleavage site coevolution. Evol Appl **8:**185-198.

39. **Kuntzen T, Timm J, Berical A, Lennon N, Berlin AM, Young SK, Lee B, Heckerman D, Carlson J, Reyor LL, Kleyman M, McMahon CM, Birch C, Schulze Zur Wiesch J, Ledlie T, Koehrsen M, Kodira C, Roberts AD, Lauer GM, Rosen HR, Bihl F, Cerny A, Spengler U, Liu Z, Kim AY, Xing Y, Schneidewind A, Madey MA, Fleckenstein JF, Park VM, Galagan JE, Nusbaum C, Walker BD, Lake-Bakaar GV, Daar ES, Jacobson IM, Gomperts ED, Edlin BR, Donfield SM, Chung RT, Talal AH, Marion T, Birren BW, Henn MR, Allen TM.** 2008. Naturally occurring dominant resistance mutations to hepatitis C virus protease and polymerase inhibitors in treatment-naive patients. Hepatology **48:**1769-1778.

40. **L'Huillier AG, Abed Y, Petty TJ, Cordey S, Thomas Y, Bouhy X, Schibler M, Simon A, Chalandon Y, van Delden C, Zdobnov E,**

Boquete-Suter P, Boivin G, Kaiser L. 2015. E119D Neuraminidase Mutation Conferring Pan-Resistance to Neuraminidase Inhibitors in an A(H1N1)pdm09 Isolate From a Stem-Cell Transplant Recipient. J Infect Dis **212:**1726-1734.

41. **Nguyen HT, Fry AM, Gubareva LV.** 2012. Neuraminidase inhibitor resistance in influenza viruses and laboratory testing methods. Antivir Ther **17:**159-173.

42. **Cai Y, Myint W, Paulsen JL, Schiffer CA, Ishima R, Kurt Yilmaz N.** 2014. Drug Resistance Mutations Alter Dynamics of Inhibitor-Bound HIV-1 Protease. J Chem Theory Comput **10:**3438-3448.

43. **Romano KP, Ali A, Royer WE, Schiffer CA.** 2010. Drug resistance against HCV NS3/4A inhibitors is defined by the balance of substrate recognition versus inhibitor binding. Proc Natl Acad Sci U S A **107:**20986-20991.

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

45. **Kairys V, Gilson MK, Lather V, Schiffer CA, Fernandes MX.** 2009. Toward the design of mutation-resistant enzyme inhibitors: further evaluation of the substrate envelope hypothesis. Chem Biol Drug Des **74:**234-245.

46. **Ozen A.** 2013. Structure and Dynamics of Viral Substrate Recognition and Drug ResistanceUniversity of Massachusetts Medical School.

47. **Tuske S, Sarafianos SG, Clark AD, Jr., Ding J, Naeger LK, White KL, Miller MD, Gibbs CS, Boyer PL, Clark P, Wang G, Gaffney BL, Jones RA, Jerina DM, Hughes SH, Arnold E.** 2004. Structures of HIV-1 RT-DNA complexes before and after incorporation of the anti-AIDS drug tenofovir. Nat Struct Mol Biol **11:**469-474.

48. **Zhu X, McBride R, Nycholat CM, Yu W, Paulson JC, Wilson IA.** 2012. Influenza virus neuraminidases with reduced enzymatic activity that avidly bind sialic Acid receptors. J Virol **86:**13371-13383.

49. **Ali A, Bandaranayake RM, Cai Y, King NM, Kolli M, Mittal S, Murzycki JF, Nalam MN, Nalivaika EA, Ozen A, Prabu-Jeyabalan MM, Thayer K, Schiffer CA.** 2010. Molecular Basis for Drug Resistance in HIV-1 Protease. Viruses **2:**2509-2535.

50. **Le TB, V.** 2015. First Aid for the USMLE Step 1 2015. 25th Anniversary Edition. McGraw-Hill Medical, New York.

51. **Boivin S, Cusack S, Ruigrok RW, Hart DJ.** 2010. Influenza A virus polymerase: structural insights into replication and host adaptation mechanisms. J Biol Chem **285:**28411-28417.

52. **Drake JW.** 1993. Rates of spontaneous mutation among RNA viruses. Proc Natl Acad Sci U S A **90:**4171-4175.

53. **Carrat F, Flahault A.** 2007. Influenza vaccine: the challenge of antigenic drift. Vaccine **25:**6852-6862.

54. **Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, Balish A, Sessions WM, Xu X, Skepner E, Deyde V, Okomo-Adhiambo M, Gubareva L, Barnes J, Smith CB, Emery SL, Hillman MJ, Rivailler P, Smagala J, de Graaf M, Burke DF, Fouchier RA, Pappas C, Alpuche-Aranda CM, Lopez-Gatell H, Olivera H, Lopez I, Myers CA, Faix D, Blair PJ, Yu C, Keene KM, Dotson PD, Jr., Boxrud D, Sambol AR, Abid SH, St George K, Bannerman T, Moore AL, Stringer DJ, Blevins P, Demmler-Harrison GJ, Ginsberg M, Kriner P, Waterman S, Smole S, Guevara HF, Belongia EA, Clark PA, Beatrice ST, Donis R, et al.** 2009. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. Science **325:**197-201.

55. **Wikipedia Contributors.** 2016. List of therapeutic monoclonal antibodies, 1 March 2016 00:39 UTC ed. Wikipedia, The Free Encyclopedia.

56. **Keller MA, Stiehm ER.** 2000. Passive immunity in prevention and treatment of infectious diseases. Clin Microbiol Rev **13:**602-614.

57. **Zhao X, Sullender WM.** 2005. In vivo selection of respiratory syncytial viruses resistant to palivizumab. J Virol **79:**3962-3968.

58. **Dreyfus C, Ekiert DC, Wilson IA.** 2013. Structure of a classical broadly neutralizing stem antibody in complex with a pandemic H2 influenza virus hemagglutinin. J Virol **87:**7149-7154.

59. **Burton DR, Poignard P, Stanfield RL, Wilson IA.** 2012. Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. Science **337:**183-186.

60. **Ekiert DC, Bhabha G, Elsliger MA, Friesen RH, Jongeneelen M, Throsby M, Goudsmit J, Wilson IA.** 2009. Antibody recognition of a highly conserved influenza virus epitope. Science **324:**246-251.

61. **Ekiert DC, Friesen RH, Bhabha G, Kwaks T, Jongeneelen M, Yu W, Ophorst C, Cox F, Korse HJ, Brandenburg B, Vogels R, Brakenhoff JP, Kompier R, Koldijk MH, Cornelissen LA, Poon LL, Peiris M, Koudstaal W, Wilson IA, Goudsmit J.** 2011. A highly conserved neutralizing epitope on group 2 influenza A viruses. Science **333:**843-850.

62. **Ekiert DC, Kashyap AK, Steel J, Rubrum A, Bhabha G, Khayat R, Lee JH, Dillon MA, O'Neil RE, Faynboym AM, Horowitz M, Horowitz L, Ward AB, Palese P, Webby R, Lerner RA, Bhatt RR, Wilson IA.** 2012. Cross-neutralization of influenza A viruses mediated by a single antibody loop. Nature **489:**526-532.

63. **Friesen RH, Lee PS, Stoop EJ, Hoffman RM, Ekiert DC, Bhabha G, Yu W, Juraszek J, Koudstaal W, Jongeneelen M, Korse HJ, Ophorst C, Brinkman-van der Linden EC, Throsby M, Kwakkenbos MJ, Bakker AQ, Beaumont T, Spits H, Kwaks T, Vogels R, Ward AB, Goudsmit J, Wilson IA.** 2014. A common solution to group 2 influenza virus neutralization. Proc Natl Acad Sci U S A **111:**445-450.

64. **Tharakaraman K, Subramanian V, Cain D, Sasisekharan V, Sasisekharan R.** 2014. Broadly neutralizing influenza hemagglutinin stem-specific antibody CR8020 targets residues that are prone to escape due to host selection pressure. Cell Host Microbe **15:**644-651.

65. **Okuno Y, Isegawa Y, Sasao F, Ueda S.** 1993. A common neutralizing epitope conserved between the hemagglutinins of influenza A virus H1 and H2 strains. J Virol **67:**2552-2558.

66. **Han T, Marasco WA.** 2011. Structural basis of influenza virus neutralization. Ann N Y Acad Sci **1217:**178-190.

67. **Throsby M, van den Brink E, Jongeneelen M, Poon LL, Alard P, Cornelissen L, Bakker A, Cox F, van Deventer E, Guan Y, Cinatl J, ter Meulen J, Lasters I, Carsetti R, Peiris M, de Kruif J, Goudsmit J.** 2008. Heterosubtypic neutralizing monoclonal antibodies cross-protective against H5N1 and H1N1 recovered from human IgM+ memory B cells. PLoS One **3:**e3942.

68. **Levantino M, Yorke BA, Monteiro DC, Cammarata M, Pearson AR.** 2015. Using synchrotrons and XFELs for time-resolved X-ray crystallography and solution scattering experiments on biomolecules. Curr Opin Struct Biol **35:**41-48.

69. **Foulkes-Murzycki JE, Scott WR, Schiffer CA.** 2007. Hydrophobic sliding: a possible mechanism for drug resistance in human immunodeficiency virus type 1 protease. Structure **15:**225-233.

70. **Mittal S, Cai Y, Nalam MN, Bolon DN, Schiffer CA.** 2012. Hydrophobic core flexibility modulates enzyme activity in HIV-1 protease. J Am Chem Soc **134:**4163-4168.

71. **Galiano L, Ding F, Veloro AM, Blackburn ME, Simmerling C, Fanucci GE.** 2009. Drug pressure selected mutations in HIV-1 protease alter flap conformations. J Am Chem Soc **131:**430-431.

72. **Amaro RE, Swift RV, Votapka L, Li WW, Walker RC, Bush RM.** 2011. Mechanism of 150-cavity formation in influenza neuraminidase. Nat Commun **2:**388.

73. **Russell RJ, Haire LF, Stevens DJ, Collins PJ, Lin YP, Blackburn GM, Hay AJ, Gamblin SJ, Skehel JJ.** 2006. The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. Nature **443:**45-49.

74. **Reeves JD, Lee FH, Miamidian JL, Jabara CB, Juntilla MM, Doms RW.** 2005. Enfuvirtide resistance mutations: impact on human immunodeficiency virus envelope function, entry inhibitor sensitivity, and virus neutralization. J Virol **79:**4991-4999.

75. **Miller MD, Hazuda DJ.** 2004. HIV resistance to the fusion inhibitor enfuvirtide: mechanisms and clinical implications. Drug Resist Updat **7:**89-95.

76. **Kwong PD, Doyle ML, Casper DJ, Cicala C, Leavitt SA, Majeed S, Steenbeke TD, Venturi M, Chaiken I, Fung M, Katinger H, Parren PW,**

Robinson J, Van Ryk D, Wang L, Burton DR, Freire E, Wyatt R, Sodroski J, Hendrickson WA, Arthos J. 2002. HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. Nature **420:**678-682.

77. **Guttman M, Cupo A, Julien JP, Sanders RW, Wilson IA, Moore JP, Lee KK.** 2015. Antibody potency relates to the ability to recognize the closed, pre-fusion form of HIV Env. Nat Commun **6:**6144.

78. **Mirsaliotis A, Nurkiyanova K, Lamb D, Kuo CW, Brighty DW.** 2007. Resistance to neutralization by antibodies targeting the coiled coil of fusion-active envelope is a common feature of retroviruses. J Biol Chem **282:**36724-36735.

79. **Molinari NA, Ortega-Sanchez IR, Messonnier ML, Thompson WW, Wortley PM, Weintraub E, Bridges CB.** 2007. The annual impact of seasonal influenza in the US: measuring disease burden and costs. Vaccine **25:**5086-5096.

80. **Noll H, Aoyagi T, Orlando J.** 1962. The structural relationship of sialidase to the influenza virus surface. Virology **18:**154-157.

81. **Palese P, Tobita K, Ueda M, Compans RW.** 1974. Characterization of temperature sensitive influenza virus mutants defective in neuraminidase. Virology **61:**397-410.

82. **Air GM.** 2012. Influenza neuraminidase. Influenza Other Respir Viruses **6:**245-256.

83. **Yang X, Steukers L, Forier K, Xiong R, Braeckmans K, Van Reeth K, Nauwynck H.** 2014. A beneficiary role for neuraminidase in influenza virus penetration through the respiratory mucus. PLoS One **9:**e110026.

84. **McKimm-Breschkin JL.** 2013. Influenza neuraminidase inhibitors: antiviral action and mechanisms of resistance. Influenza Other Respir Viruses **7 Suppl 1:**25-36.

85. **Meijer A, Rebelo-de-Andrade H, Correia V, Besselaar T, Drager-Dayal R, Fry A, Gregory V, Gubareva L, Kageyama T, Lackenby A, Lo J, Odagiri T, Pereyaslov D, Siqueira MM, Takashita E, Tashiro M, Wang D, Wong S, Zhang W, Daniels RS, Hurt AC.** 2014. Global update on the susceptibility of human influenza viruses to neuraminidase inhibitors, 2012-2013. Antiviral Res **110:**31-41.

86. **Takashita E, Meijer A, Lackenby A, Gubareva L, Rebelo-de-Andrade H, Besselaar T, Fry A, Gregory V, Leang SK, Huang W, Lo J, Pereyaslov D, Siqueira MM, Wang D, Mak GC, Zhang W, Daniels RS, Hurt AC, Tashiro M.** 2015. Global update on the susceptibility of human influenza viruses to neuraminidase inhibitors, 2013-2014. Antiviral Res **117:**27-38.

87. **Abed Y, Baz M, Boivin G.** 2006. Impact of neuraminidase mutations conferring influenza resistance to neuraminidase inhibitors in the N1 and N2 genetic backgrounds. Antivir Ther **11:**971-976.

88. **Orozovic G, Orozovic K, Lennerstrand J, Olsen B.** 2011. Detection of resistance mutations to antivirals oseltamivir and zanamivir in avian influenza A viruses isolated from wild birds. PLoS One **6:**e16028.

89. **Samson M, Pizzorno A, Abed Y, Boivin G.** 2013. Influenza virus resistance to neuraminidase inhibitors. Antiviral Res **98:**174-185.

90. **Rameix-Welti M-A, Munier S, Naffakh N.** 2012. Resistance Development to Influenza Virus Sialidase Inhibitors, p 153-174. *In* Itzstein M (ed), Influenza Virus Sialidase - A Drug Discovery Target doi:10.1007/978-3-7643-8927-7 8. Springer Basel, Basel.

91. **Hurt AC, Holien JK, Parker M, Kelso A, Barr IG.** 2009. Zanamivir-resistant influenza viruses with a novel neuraminidase mutation. J Virol **83:**10366-10373.

92. **Eshaghi A, Shalhoub S, Rosenfeld P, Li A, Higgins RR, Stogios PJ, Savchenko A, Bastien N, Li Y, Rotstein C, Gubbay JB.** 2014. Multiple influenza A (H3N2) mutations conferring resistance to neuraminidase inhibitors in a bone marrow transplant recipient. Antimicrob Agents Chemother **58:**7188-7197.

93. **McKimm-Breschkin JL, Barrett S, Pudjiatmoko, Azhar M, Wong FY, Selleck P, Mohr PG, McGrane J, Kim M.** 2013. I222 Neuraminidase mutations further reduce oseltamivir susceptibility of Indonesian Clade 2.1 highly pathogenic Avian Influenza A(H5N1) viruses. PLoS One **8:**e66105.

94. **Ozen A, Haliloglu T, Schiffer CA.** 2011. Dynamics of preferential substrate recognition in HIV-1 protease: redefining the substrate envelope. J Mol Biol **410:**726-744.

95. **Ozen A, Sherman W, Schiffer CA.** 2013. Improving the Resistance Profile of Hepatitis C NS3/4A Inhibitors: Dynamic Substrate Envelope Guided Design. J Chem Theory Comput **9:**5693-5705.

96. **Li Y, Cao H, Dao N, Luo Z, Yu H, Chen Y, Xing Z, Baumgarth N, Cardona C, Chen X.** 2011. High-throughput neuraminidase substrate specificity study of human and avian influenza A viruses. Virology **415:**12-19.

97. **Jiang L, Liu P, Bank C, Renzette N, Prachanronarong K, Yilmaz LS, Caffrey DR, Zeldovich KB, Schiffer CA, Kowalik TF, Jensen JD, Finberg RW, Wang JP, Bolon DN.** 2016. A Balance between Inhibitor Binding and Substrate Processing Confers Influenza Drug Resistance. J Mol Biol **428:**538-553.

98. **Collins PJ, Haire LF, Lin YP, Liu J, Russell RJ, Walker PA, Skehel JJ, Martin SR, Hay AJ, Gamblin SJ.** 2008. Crystal structures of oseltamivir-resistant influenza virus neuraminidase mutants. Nature **453:**1258-1261.

99. **Yen HL, Hoffmann E, Taylor G, Scholtissek C, Monto AS, Webster RG, Govorkova EA.** 2006. Importance of neuraminidase active-site residues to the neuraminidase inhibitor resistance of influenza viruses. J Virol **80:**8787-8795.

100. **McKimm-Breschkin JL, Williams J, Barrett S, Jachno K, McDonald M, Mohr PG, Saito T, Tashiro M.** 2013. Reduced susceptibility to all neuraminidase inhibitors of influenza H1N1 viruses with haemagglutinin mutations and mutations in non-conserved residues of the neuraminidase. J Antimicrob Chemother **68:**2210-2221.

101. **Varghese JN.** 1999. Development of neuraminidase inhibitors as anti-influenza virus drugs. Drug Development Research **46:**176-196.

102. **Chan J, Bennet AJ.** 2012. Enzymology of Influenza Virus Sialidase, p 47-66. *In* Itzstein M (ed), Influenza Virus Sialidase - A Drug Discovery Target doi:10.1007/978-3-7643-8927-7 3. Springer Basel, Basel.

103. **Varki A, Schauer R.** 2009. Sialic Acids. *In* Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (ed), Essentials of Glycobiology, 2nd ed, Cold Spring Harbor (NY).

104. **Bate C, Nolan W, Williams A.** 2016. Sialic Acid on the Glycosylphosphatidylinositol Anchor Regulates PrP-mediated Cell Signaling and Prion Formation. J Biol Chem **291:**160-170.

105. **Wikipedia Contributors.** 2016. Glycophosphatidylinositol, 26 January 2016 08:42 UTC ed. Wikipedia, The Free Encyclopedia.

106. **Wang M, Qi J, Liu Y, Vavricka CJ, Wu Y, Li Q, Gao GF.** 2011. Influenza A virus N5 neuraminidase has an extended 150-cavity. J Virol **85:**8431-8435.

107. **Xu X, Zhu X, Dwek RA, Stevens J, Wilson IA.** 2008. Structural characterization of the 1918 influenza virus H1N1 neuraminidase. J Virol **82:**10493-10501.

108. **Schrodinger, LLC.** 2015. Maestro, Version 10.2, New York, NY.

109. **Schrodinger, LLC.** 2015. Prime, Version 4.0, New York, NY.

110. **Schrodinger, LLC.** 2015. Schrodinger Suite 2015-2 Protein Preparation Wizard, New York, NY.

111. **Bowers KJ, Chow, E., Xu, H., Dror, R.O., Eastwood, M.P., Gregersen, B.A., Klepeis, J.L., Kolossvary, I., Moraes, M.A., Sacerdoti, F.D., Salmon, J.K., Yibing, S., Shaw, D.E. .** Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters, p. *In* (ed),

112. **Banks JL, Beard HS, Cao Y, Cho AE, Damm W, Farid R, Felts AK, Halgren TA, Mainz DT, Maple JR, Murphy R, Philipp DM, Repasky MP, Zhang LY, Berne BJ, Friesner RA, Gallicchio E, Levy RM.** 2005. Integrated Modeling Program, Applied Chemical Theory (IMPACT). J Comput Chem **26:**1752-1780.

113. **Schrodinger, LLC.** 2015. Desmond 4.2 User Manual, New York, NY.

114. **Humphrey W, Dalke A, Schulten K.** 1996. VMD: visual molecular dynamics. J Mol Graph **14:**33-38, 27-38.

115. **VMD Mailing List.** 2008. VMD B Factor Script, http://www.ks.uiuc.edu/Research/vmd/mailing_list/vmd-l/12594.html.

116. **Schrodinger, LLC.** 2010. The PyMOL Molecular Graphics System, Version 1.3r1.

117. **Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A.** 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics **28:**1647-1649.

118. **The MathWorks, Inc.** 2014. MATLAB and Toolboxes Release 2014b, The MathWorks, Inc., Natick, Massachusetts, United States.

119. **GraphPad Software.** 2016. GraphPad Prism for Mac OSX, Version 6.0, GraphPad Software, San Diego, California, USA.

120. **WHO.** 2014. Hepatitis C Fact Sheet No. 164. http://www.who.int/mediacentre/factsheets/fs164/enl. Accessed April 23, 2014.

121. **Luu L.** December 27, 2015 2015. Liver Transplants. Medscape.

122. **Simmonds P, Bukh J, Combet C, Deleage G, Enomoto N, Feinstone S, Halfon P, Inchauspe G, Kuiken C, Maertens G, Mizokami M, Murphy DG, Okamoto H, Pawlotsky JM, Penin F, Sablon E, Shin IT, Stuyver LJ, Thiel HJ, Viazov S, Weiner AJ, Widell A.** 2005. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. Hepatology **42:**962-973.

123. **Au JS, Pockros PJ.** 2014. Novel therapeutic approaches for hepatitis C. Clin Pharmacol Ther **95:**78-88.

124. **Jensen SB, Serre SB, Humes DG, Ramirez S, Li YP, Bukh J, Gottwein JM.** 2015. Substitutions at NS3 Residue 155, 156, or 168 of Hepatitis C Virus Genotypes 2 to 6 Induce Complex Patterns of Protease Inhibitor Resistance. Antimicrob Agents Chemother **59:**7426-7436.

125. **Bukh J, Miller RH, Purcell RH.** 1995. Biology and genetic heterogeneity of hepatitis C virus. Clin Exp Rheumatol **13 Suppl 13:**S3-7.

126. **Bukh J, Miller RH, Purcell RH.** 1995. Genetic heterogeneity of hepatitis C virus: quasispecies and genotypes. Semin Liver Dis **15:**41-63.

127. **Pockros PJDB, A.M.; Bloom, A.B.** 2014. Direct acting antivirals for the treatment of hepatitis C virus infection, *on* UpToDate. http://www.uptodate.com/contents/direct-acting-antivirals-for-the-treatment-of-hepatitis-c-virus-infection. Accessed

128. **Silverman E.** August 12, 2014 2014. From Riches to Rags: Vertex Discontinues Incivek as Sales Evaporate, p *In* The Wall Street Journal. http://blogs.wsj.com/pharmalot/2014/08/12/from-riches-to-rags-vertex-discontinues-incivek-as-sales-evaporate/.

129. **Nalam MN, Ali A, Altman MD, Reddy GS, Chellappan S, Kairys V, Ozen A, Cao H, Gilson MK, Tidor B, Rana TM, Schiffer CA.** 2010. Evaluating the substrate-envelope hypothesis: structural analysis of novel

HIV-1 protease inhibitors designed to be robust against drug resistance. J Virol **84:**5368-5378.

130. **Chen Z, Benureau Y, Rijnbrand R, Yi J, Wang T, Warter L, Lanford RE, Weinman SA, Lemon SM, Martin A, Li K.** 2007. GB virus B disrupts RIG-I signaling by NS3/4A-mediated cleavage of the adaptor protein MAVS. J Virol **81:**964-976.

131. **Heim MH.** 2013. Innate immunity and HCV. J Hepatol **58:**564-574.

132. **Li XD, Sun L, Seth RB, Pineda G, Chen ZJ.** 2005. Hepatitis C virus protease NS3/4A cleaves mitochondrial antiviral signaling protein off the mitochondria to evade innate immunity. Proc Natl Acad Sci U S A **102:**17717-17722.

133. **Li K, Foy E, Ferreon JC, Nakamura M, Ferreon AC, Ikeda M, Ray SC, Gale M, Jr., Lemon SM.** 2005. Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF. Proc Natl Acad Sci U S A **102:**2992-2997.

134. **FDA.** 2016. Drugs @ FDA. Federal Drug Administration.

135. **National Institute of Health.** 2016. ClinicalTrials.gov.

136. **Sheng XC, Casarez A, Cai R, Clarke MO, Chen X, Cho A, Delaney WEt, Doerffler E, Ji M, Mertzman M, Pakdaman R, Pyun HJ, Rowe T, Wu Q, Xu J, Kim CU.** 2012. Discovery of GS-9256: a novel phosphinic acid derived inhibitor of the hepatitis C virus NS3/4A protease with potent clinical activity. Bioorg Med Chem Lett **22:**1394-1396.

137. **Vachon M-LD, Douglas T.** 2011. The Era of Direct-acting Antivirals Has Begun. Seminars in Liver Disease. **31**(4)**:**399-409.

138. **Huang MP, S.; Patel, D.; et al.** 2010. ACH-2684: HCV NS3 protease inhibitor with potent activity against multiple genotypes and known resistant variants. Hepatology **52:**1204A.

139. **Poordad F, Dieterich D.** 2012. Treating hepatitis C: current standard of care and emerging direct-acting antiviral agents. J Viral Hepat **19:**449-464.

140. **Mathias A.** Clinical Pharmacology of DAA's for HCV: What's New and What's in the Pipeline, p. *In* (ed),

141. **Steinkuhler C, Biasiol G, Brunetti M, Urbani A, Koch U, Cortese R, Pessi A, De Francesco R.** 1998. Product inhibition of the hepatitis C virus NS3 protease. Biochemistry **37:**8899-8905.

142. **Llinas-Brunet M, Bailey M, Fazal G, Goulet S, Halmos T, Laplante S, Maurice R, Poirier M, Poupart MA, Thibeault D, Wernic D, Lamarre D.** 1998. Peptide-based inhibitors of the hepatitis C virus serine protease. Bioorg Med Chem Lett **8:**1713-1718.

143. **De Francesco R, Migliaccio G.** 2005. Challenges and successes in developing new therapies for hepatitis C. Nature **436:**953-960.

144. **Perni RB, Almquist SJ, Byrn RA, Chandorkar G, Chaturvedi PR, Courtney LF, Decker CJ, Dinehart K, Gates CA, Harbeson SL, Heiser A, Kalkeri G, Kolaczkowski E, Lin K, Luong YP, Rao BG, Taylor WP,**

Thomson JA, Tung RD, Wei Y, Kwong AD, Lin C. 2006. Preclinical profile of VX-950, a potent, selective, and orally bioavailable inhibitor of hepatitis C virus NS3-4A serine protease. Antimicrob Agents Chemother **50:**899-909.

145. **Kwong AD, Kauffman RS, Hurter P, Mueller P.** 2011. Discovery and development of telaprevir: an NS3-4A protease inhibitor for treating genotype 1 chronic hepatitis C virus. Nat Biotechnol **29:**993-1003.

146. **Malcolm BA, Liu R, Lahser F, Agrawal S, Belanger B, Butkiewicz N, Chase R, Gheyas F, Hart A, Hesk D, Ingravallo P, Jiang C, Kong R, Lu J, Pichardo J, Prongay A, Skelton A, Tong X, Venkatraman S, Xia E, Girijavallabhan V, Njoroge FG.** 2006. SCH 503034, a mechanism-based inhibitor of hepatitis C virus NS3 protease, suppresses polyprotein maturation and enhances the antiviral activity of alpha interferon in replicon cells. Antimicrob Agents Chemother **50:**1013-1020.

147. **Idrees S, Ashfaq UA.** 2013. HCV infection and NS-3 serine protease inhibitors. Virology & Mycology **2013**.

148. **Schaefer EA, Chung RT.** 2012. Anti-hepatitis C virus drugs in development. Gastroenterology **142:**1340-1350 e1341.

149. **Arasappan A, Bennett F, Bogen SL, Venkatraman S, Blackman M, Chen KX, Hendrata S, Huang Y, Huelgas RM, Nair L, Padilla AI, Pan W, Pike R, Pinto P, Ruan S, Sannigrahi M, Velazquez F, Vibulbhan B, Wu W, Yang W, Saksena AK, Girijavallabhan V, Shih NY, Kong J,**

Meng T, Jin Y, Wong J, McNamara P, Prongay A, Madison V, Piwinski JJ, Cheng KC, Morrison R, Malcolm B, Tong X, Ralston R, Njoroge FG. 2010. Discovery of Narlaprevir (SCH 900518): A Potent, Second Generation HCV NS3 Serine Protease Inhibitor. ACS Med Chem Lett **1:**64-69.

150. **Llinas-Brunet M, Bailey MD, Goudreau N, Bhardwaj PK, Bordeleau J, Bos M, Bousquet Y, Cordingley MG, Duan J, Forgione P, Garneau M, Ghiro E, Gorys V, Goulet S, Halmos T, Kawai SH, Naud J, Poupart MA, White PW.** 2010. Discovery of a potent and selective noncovalent linear inhibitor of the hepatitis C virus NS3 protease (BI 201335). J Med Chem **53:**6466-6476.

151. **White PW, Llinas-Brunet M, Amad M, Bethell RC, Bolger G, Cordingley MG, Duan J, Garneau M, Lagace L, Thibeault D, Kukolj G.** 2010. Preclinical characterization of BI 201335, a C-terminal carboxylic acid inhibitor of the hepatitis C virus NS3-NS4A protease. Antimicrob Agents Chemother **54:**4611-4618.

152. **McPhee F, Sheaffer AK, Friborg J, Hernandez D, Falk P, Zhai G, Levine S, Chaniewski S, Yu F, Barry D, Chen C, Lee MS, Mosure K, Sun LQ, Sinz M, Meanwell NA, Colonno RJ, Knipe J, Scola P.** 2012. Preclinical Profile and Characterization of the Hepatitis C Virus NS3 Protease Inhibitor Asunaprevir (BMS-650032). Antimicrob Agents Chemother **56:**5387-5396.

153. **Agarwal A, Zhang B, Olek E, Robison H, Robarge L, Deshpande M.** 2012. Rapid and sharp decline in HCV upon monotherapy with NS3 protease inhibitor, ACH-1625. Antivir Ther **17:**1533-1539.

154. **Sheng XC, Appleby T, Butler T, Cai R, Chen X, Cho A, Clarke MO, Cottell J, Delaney WEt, Doerffler E, Link J, Ji M, Pakdaman R, Pyun HJ, Wu Q, Xu J, Kim CU.** 2012. Discovery of GS-9451: an acid inhibitor of the hepatitis C virus NS3/4A protease. Bioorg Med Chem Lett **22:**2629-2634.

155. **De Clercq E.** 2014. Current race in the development of DAAs (direct-acting antivirals) against HCV. Biochem Pharmacol **89:**441-452.

156. **McPhee F, Hernandez D, Yu F, Ueland J, Monikowski A, Carifa A, Falk P, Wang C, Fridell R, Eley T, Zhou N, Gardiner D.** 2013. Resistance analysis of hepatitis C virus genotype 1 prior treatment null responders receiving daclatasvir and asunaprevir. Hepatology **58:**902-911.

157. **McCauley JA, McIntyre CJ, Rudd MT, Nguyen KT, Romano JJ, Butcher JW, Gilbert KF, Bush KJ, Holloway MK, Swestock J, Wan BL, Carroll SS, DiMuzio JM, Graham DJ, Ludmerer SW, Mao SS, Stahlhut MW, Fandozzi CM, Trainor N, Olsen DB, Vacca JP, Liverton NJ.** 2010. Discovery of vaniprevir (MK-7009), a macrocyclic hepatitis C virus NS3/4a protease inhibitor. J Med Chem **53:**2443-2463.

158. **Harper S, McCauley JA, Rudd MT, Ferrara M, DiFilippo M, Crescenzi B, Koch U, Petrocchi A, Holloway MK, Butcher JW, Romano JJ, Bush**

KJ, Gilbert KF, McIntyre CJ, Nguyen KT, Nizi E, Carroll SS, Ludmerer SW, Burlein C, DiMuzio JM, Graham DJ, McHale CM, Stahlhut MW, Olsen DB, Monteagudo E, Cianetti S, Giuliano C, Pucci V, Trainor N, Fandozzi CM, Rowley M, Coleman PJ, Vacca JP, Summa V, Liverton NJ. 2012. Discovery of MK-5172, a Macrocyclic Hepatitis C Virus NS3/4a Protease Inhibitor. ACS Med Chem Lett **3:**332-336.

159. **Rosenquist A, Samuelsson B, Johansson PO, Cummings MD, Lenz O, Raboisson P, Simmen K, Vendeville S, de Kock H, Nilsson M, Horvath A, Kalmeijer R, de la Rosa G, Beumont-Mauviel M.** 2014. Discovery and development of simeprevir (TMC435), a HCV NS3/4A protease inhibitor. J Med Chem **57:**1673-1693.

160. **He Y, King MS, Kempf DJ, Lu L, Lim HB, Krishnan P, Kati W, Middleton T, Molla A.** 2008. Relative replication capacity and selective advantage profiles of protease inhibitor-resistant hepatitis C virus (HCV) NS3 protease mutants in the HCV genotype 1b replicon system. Antimicrob Agents Chemother **52:**1101-1110.

161. **Lange CM, Sarrazin C, Zeuzem S.** 2010. Review article: specifically targeted anti-viral therapy for hepatitis C - a new era in therapy. Aliment Pharmacol Ther **32:**14-28.

162. **Liverton NJ, Carroll SS, Dimuzio J, Fandozzi C, Graham DJ, Hazuda D, Holloway MK, Ludmerer SW, McCauley JA, McIntyre CJ, Olsen DB, Rudd MT, Stahlhut M, Vacca JP.** 2010. MK-7009, a potent and selective

inhibitor of hepatitis C virus NS3/4A protease. Antimicrob Agents Chemother **54:**305-311.

163. **Seiwert SD, Andrews SW, Jiang Y, Serebryany V, Tan H, Kossen K, Rajagopalan PT, Misialek S, Stevens SK, Stoycheva A, Hong J, Lim SR, Qin X, Rieger R, Condroski KR, Zhang H, Do MG, Lemieux C, Hingorani GP, Hartley DP, Josey JA, Pan L, Beigelman L, Blatt LM.** 2008. Preclinical characteristics of the hepatitis C virus NS3/4A protease inhibitor ITMN-191 (R7227). Antimicrob Agents Chemother **52:**4432-4441.

164. **Summa V, Ludmerer SW, McCauley JA, Fandozzi C, Burlein C, Claudio G, Coleman PJ, Dimuzio JM, Ferrara M, Di Filippo M, Gates AT, Graham DJ, Harper S, Hazuda DJ, Huang Q, McHale C, Monteagudo E, Pucci V, Rowley M, Rudd MT, Soriano A, Stahlhut MW, Vacca JP, Olsen DB, Liverton NJ, Carroll SS.** 2012. MK-5172, a selective inhibitor of hepatitis C virus NS3/4a protease with broad activity across genotypes and resistant variants. Antimicrob Agents Chemother **56:**4161-4167.

165. **Poordad F, Lawitz E, Kowdley KV, Cohen DE, Podsadecki T, Siggelkow S, Heckaman M, Larsen L, Menon R, Koev G, Tripathi R, Pilot-Matias T, Bernstein B.** 2013. Exploratory study of oral combination antiviral therapy for hepatitis C. N Engl J Med **368:**45-53.

166. **Lin TI, Lenz O, Fanning G, Verbinnen T, Delouvroy F, Scholliers A, Vermeiren K, Rosenquist A, Edlund M, Samuelsson B, Vrang L, de**

Kock H, Wigerinck P, Raboisson P, Simmen K. 2009. In vitro activity and preclinical profile of TMC435350, a potent hepatitis C virus protease inhibitor. Antimicrob Agents Chemother **53:**1377-1385.

167. **Levin JNTP-M, T.; Lu, L.; Reisch, T.; Dekhtyar, T,; Krishnan, P.; Beyer, J.; Tripathi, R.; Pithawalla, R.; Asatryan, A.; Campbell, A.; Kort, J.; Collins, C.** 2014. A Next Generation HCV DAA Combination: Potent, Pangenotypic Inhibitors ABT-493 and ABT-530 With High Barriers to Resistance, abstr 65th Annual Meeting of the American Association for the Study of Liver Diseases, Boston, Massachusetts, United States, November 7-11, 2014.

168. **Lontok E, Harrington P, Howe A, Kieffer T, Lennerstrand J, Lenz O, McPhee F, Mo H, Parkin N, Pilot-Matias T, Miller V.** 2015. Hepatitis C virus drug resistance-associated substitutions: State of the art summary. Hepatology **62:**1623-1632.

169. **Kieffer TL, Sarrazin C, Miller JS, Welker MW, Forestier N, Reesink HW, Kwong AD, Zeuzem S.** 2007. Telaprevir and pegylated interferon-alpha-2a inhibit wild-type and resistant genotype 1 hepatitis C virus replication in patients. Hepatology **46:**631-639.

170. **Lin C, Gates CA, Rao BG, Brennan DL, Fulghum JR, Luong YP, Frantz JD, Lin K, Ma S, Wei YY, Perni RB, Kwong AD.** 2005. In vitro studies of cross-resistance mutations against two hepatitis C virus serine

protease inhibitors, VX-950 and BILN 2061. J Biol Chem **280:**36784-36791.

171. **Sarrazin C, Rouzier R, Wagner F, Forestier N, Larrey D, Gupta SK, Hussain M, Shah A, Cutler D, Zhang J, Zeuzem S.** 2007. SCH 503034, a novel hepatitis C virus protease inhibitor, plus pegylated interferon alpha-2b for genotype 1 nonresponders. Gastroenterology **132:**1270-1278.

172. **Tong X, Arasappan A, Bennett F, Chase R, Feld B, Guo Z, Hart A, Madison V, Malcolm B, Pichardo J, Prongay A, Ralston R, Skelton A, Xia E, Zhang R, Njoroge FG.** 2010. Preclinical characterization of the antiviral activity of SCH 900518 (narlaprevir), a novel mechanism-based inhibitor of hepatitis C virus NS3 protease. Antimicrob Agents Chemother **54:**2365-2370.

173. **Tong X, Bogen S, Chase R, Girijavallabhan V, Guo Z, Njoroge FG, Prongay A, Saksena A, Skelton A, Xia E, Ralston R.** 2008. Characterization of resistance mutations against HCV ketoamide protease inhibitors. Antiviral Res **77:**177-185.

174. **Tong X, Chase R, Skelton A, Chen T, Wright-Minogue J, Malcolm BA.** 2006. Identification and analysis of fitness of resistance mutations against the HCV protease inhibitor SCH 503034. Antiviral Res **70:**28-38.

175. **Ali A, Aydin C, Gildemeister R, Romano KP, Cao H, Ozen A, Soumana D, Newton A, Petropoulos CJ, Huang W, Schiffer CA.** 2013. Evaluating

the role of macrocycles in the susceptibility of hepatitis C virus NS3/4A protease inhibitors to drug resistance. ACS Chem Biol **8:**1469-1478.

176. **Soumana D.** 2015. Hepatitis C Virus: Structural Insights Into Protease Inhibitor Efficacy and Drug ResistanceUniversity of Massachusetts Medical School.

177. **Rudd MT, Butcher JW, Nguyen KT, McIntyre CJ, Romano JJ, Gilbert KF, Bush KJ, Liverton NJ, Holloway MK, Harper S, Ferrara M, DiFilippo M, Summa V, Swestock J, Fritzen J, Carroll SS, Burlein C, DiMuzio JM, Gates A, Graham DJ, Huang Q, McClain S, McHale C, Stahlhut MW, Black S, Chase R, Soriano A, Fandozzi CM, Taylor A, Trainor N, Olsen DB, Coleman PJ, Ludmerer SW, McCauley JA.** 2015. P2-quinazolinones and bis-macrocycles as new templates for next-generation hepatitis C virus NS3/4a protease inhibitors: discovery of MK-2748 and MK-6325. ChemMedChem **10:**727-735.

178. **Shah U, Jayne C, Chackalamannil S, Velazquez F, Guo Z, Buevich A, Howe JA, Chase R, Soriano A, Agrawal S, Rudd MT, McCauley JA, Liverton NJ, Romano J, Bush K, Coleman PJ, Grise-Bard C, Brochu MC, Charron S, Aulakh V, Bachand B, Beaulieu P, Zaghdane H, Bhat S, Han Y, Vacca JP, Davies IW, Weber AE, Venkatraman S.** 2014. Novel Quinoline-Based P2-P4 Macrocyclic Derivatives As Pan-Genotypic HCV NS3/4a Protease Inhibitors. ACS Med Chem Lett **5:**264-269.

179. **Romano KP, Ali A, Aydin C, Soumana D, Ozen A, Deveau LM, Silver C, Cao H, Newton A, Petropoulos CJ, Huang W, Schiffer CA.** 2012. The molecular basis of drug resistance against hepatitis C virus NS3/4A protease inhibitors. PLoS Pathog **8:**e1002832.

180. **Lemke CT, Goudreau N, Zhao S, Hucke O, Thibeault D, Llinas-Brunet M, White PW.** 2011. Combined X-ray, NMR, and kinetic analyses reveal uncommon binding characteristics of the hepatitis C virus NS3-NS4A protease inhibitor BI 201335. J Biol Chem **286:**11434-11443.

181. **Cummings MD, Lindberg J, Lin TI, de Kock H, Lenz O, Lilja E, Fellander S, Baraznenok V, Nystrom S, Nilsson M, Vrang L, Edlund M, Rosenquist A, Samuelsson B, Raboisson P, Simmen K.** 2010. Induced-fit binding of the macrocyclic noncovalent inhibitor TMC435 to its HCV NS3/NS4A protease target. Angew Chem Int Ed Engl **49:**1652-1655.

182. **Soumana DI, Ali A, Schiffer CA.** 2014. Structural analysis of asunaprevir resistance in HCV NS3/4A protease. ACS Chem Biol **9:**2485-2490.

183. **Yao N, Reichert P, Taremi SS, Prosise WW, Weber PC.** 1999. Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase. Structure **7:**1353-1363.

184. **Thompson WW, Shay DK, Weintraub E, Brammer L, Cox N, Anderson LJ, Fukuda K.** 2003. Mortality associated with influenza and respiratory syncytial virus in the United States. JAMA **289:**179-186.

185. **Lambert LC, Fauci AS.** 2010. Influenza vaccines for the future. N Engl J Med **363:**2036-2044.

186. **Palese P, Tumpey TM, Garcia-Sastre A.** 2006. What can we learn from reconstructing the extinct 1918 pandemic influenza virus? Immunity **24:**121-124.

187. **Palese P, Shaw M.** 2007. Orthomyxoviridae: the viruses and their replication. *In* Knipe D, Howley P (ed), Fields Virology. Lippincott Williams & Wilkins, Philadelphia.

188. **Tong S, Li Y, Rivailler P, Conrardy C, Castillo DA, Chen LM, Recuenco S, Ellison JA, Davis CT, York IA, Turmelle AS, Moran D, Rogers S, Shi M, Tao Y, Weil MR, Tang K, Rowe LA, Sammons S, Xu X, Frace M, Lindblade KA, Cox NJ, Anderson LJ, Rupprecht CE, Donis RO.** 2012. A distinct lineage of influenza A virus from bats. Proc Natl Acad Sci U S A **109:**4269-4274.

189. **Russell RJ, Kerry PS, Stevens DJ, Steinhauer DA, Martin SR, Gamblin SJ, Skehel JJ.** 2008. Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion. Proc Natl Acad Sci U S A **105:**17736-17741.

190. **Wilson IA, Skehel JJ, Wiley DC.** 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. Nature **289:**366-373.

191. **Corti D, Lanzavecchia A.** 2013. Broadly neutralizing antiviral antibodies. Annu Rev Immunol **31:**705-742.

192. **Whittle JR, Zhang R, Khurana S, King LR, Manischewitz J, Golding H, Dormitzer PR, Haynes BF, Walter EB, Moody MA, Kepler TB, Liao HX, Harrison SC.** 2011. Broadly neutralizing human antibody that recognizes the receptor-binding pocket of influenza virus hemagglutinin. Proc Natl Acad Sci U S A **108:**14216-14221.

193. **Yoshida R, Igarashi M, Ozaki H, Kishida N, Tomabechi D, Kida H, Ito K, Takada A.** 2009. Cross-protective potential of a novel monoclonal antibody directed against antigenic site B of the hemagglutinin of influenza A viruses. PLoS Pathog **5:**e1000350.

194. **Sui J, Hwang WC, Perez S, Wei G, Aird D, Chen LM, Santelli E, Stec B, Cadwell G, Ali M, Wan H, Murakami A, Yammanuru A, Han T, Cox NJ, Bankston LA, Donis RO, Liddington RC, Marasco WA.** 2009. Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. Nat Struct Mol Biol **16:**265-273.

195. **Sui J, Li W, Murakami A, Tamin A, Matthews LJ, Wong SK, Moore MJ, Tallarico AS, Olurinde M, Choe H, Anderson LJ, Bellini WJ, Farzan M, Marasco WA.** 2004. Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association. Proc Natl Acad Sci U S A **101:**2536-2541.

196. **Foll M, Poh YP, Renzette N, Ferrer-Admetlla A, Bank C, Shim H, Malaspinas AS, Ewing G, Liu P, Wegmann D, Caffrey DR, Zeldovich KB, Bolon DN, Wang JP, Kowalik TF, Schiffer CA, Finberg RW, Jensen JD.** 2014. Influenza virus drug resistance: a time-sampled population genetics perspective. PLoS Genet **10:**e1004185.

197. **Renzette N, Caffrey DR, Zeldovich KB, Liu P, Gallagher GR, Aiello D, Porter AJ, Kurt-Jones EA, Bolon DN, Poh YP, Jensen JD, Schiffer CA, Kowalik TF, Finberg RW, Wang JP.** 2014. Evolution of the influenza A virus genome during development of oseltamivir resistance in vitro. J Virol **88:**272-281.

198. **Zeldovich KB, Liu P, Renzette N, Foll M, Pham ST, Venev SV, Gallagher GR, Bolon DN, Kurt-Jones EA, Jensen JD, Caffrey DR, Schiffer CA, Kowalik TF, Wang JP, Finberg RW.** 2015. Positive Selection Drives Preferred Segment Combinations during Influenza Virus Reassortment. Mol Biol Evol **32:**1519-1532.

199. **Foll M, Shim H, Jensen JD.** 2014. WFABC: a Wright-Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. Mol Ecol Resour doi:10.1111/1755-0998.12280.

200. **Sandbulte MR, Westgeest KB, Gao J, Xu X, Klimov AI, Russell CA, Burke DF, Smith DJ, Fouchier RA, Eichelberger MC.** 2011. Discordant antigenic drift of neuraminidase and hemagglutinin in H1N1 and H3N2 influenza viruses. Proc Natl Acad Sci U S A **108:**20748-20753.

201. **Liu J, Stevens DJ, Haire LF, Walker PA, Coombs PJ, Russell RJ, Gamblin SJ, Skehel JJ.** 2009. Structures of receptor complexes formed by hemagglutinins from the Asian Influenza pandemic of 1957. Proc Natl Acad Sci U S A **106:**17175-17180.

202. **Eisen MB, Sabesan S, Skehel JJ, Wiley DC.** 1997. Binding of the influenza A virus to cell-surface receptors: structures of five hemagglutinin-sialyloligosaccharide complexes determined by X-ray crystallography. Virology **232:**19-31.

203. **Weis W, Brown JH, Cusack S, Paulson JC, Skehel JJ, Wiley DC.** 1988. Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. Nature **333:**426-431.

204. **Tumpey TM, Maines TR, Van Hoeven N, Glaser L, Solorzano A, Pappas C, Cox NJ, Swayne DE, Palese P, Katz JM, Garcia-Sastre A.** 2007. A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. Science **315:**655-659.

205. **Bullough PA, Hughson FM, Skehel JJ, Wiley DC.** 1994. Structure of influenza haemagglutinin at the pH of membrane fusion. Nature **371:**37-43.

206. **Wilschat JC, McElhaney JE, Palache AM.** 2006. Influenza Rapid Reference, 2 ed. Mosby-Elsevier Science, London.

207. **Daniels RS, Downie JC, Hay AJ, Knossow M, Skehel JJ, Wang ML, Wiley DC.** 1985. Fusion mutants of the influenza virus hemagglutinin glycoprotein. Cell **40:**431-439.

208. **Ivanovic T, Choi JL, Whelan SP, van Oijen AM, Harrison SC.** 2013. Influenza-virus membrane fusion by cooperative fold-back of stochastically induced hemagglutinin intermediates. Elife **2:**e00333.

209. **Xu R, Wilson IA.** 2011. Structural characterization of an early fusion intermediate of influenza virus hemagglutinin. J Virol **85:**5172-5182.

210. **Hensley SE, Das SR, Gibbs JS, Bailey AL, Schmidt LM, Bennink JR, Yewdell JW.** 2011. Influenza A virus hemagglutinin antibody escape promotes neuraminidase antigenic variation and drug resistance. PLoS One **6:**e15190.

211. **Air GM, Els MC, Brown LE, Laver WG, Webster RG.** 1985. Location of antigenic sites on the three-dimensional structure of the influenza N2 virus neuraminidase. Virology **145:**237-248.

212. **Air GM, Laver WG, Webster RG, Els MC, Luo M.** 1989. Antibody recognition of the influenza virus neuraminidase. Cold Spring Harb Symp Quant Biol **54 Pt 1:**247-255.

213. **Webster RG, Air GM, Metzger DW, Colman PM, Varghese JN, Baker AT, Laver WG.** 1987. Antigenic structure and variation in an influenza virus N9 neuraminidase. J Virol **61:**2910-2916.

214. **Parmley JL, Hurst LD.** 2007. How do synonymous mutations affect fitness? Bioessays **29:**515-519.

215. **Cuevas JM, Domingo-Calap P, Sanjuan R.** 2012. The fitness effects of synonymous mutations in DNA and RNA viruses. Mol Biol Evol **29:**17-20.

216. **Sauna ZE, Kimchi-Sarfaty C.** 2011. Understanding the contribution of synonymous mutations to human disease. Nat Rev Genet **12:**683-691.

217. **Mair CM, Ludwig K, Herrmann A, Sieben C.** 2014. Receptor binding and pH stability - how influenza A virus hemagglutinin affects host-specific virus infection. Biochim Biophys Acta **1838:**1153-1168.

218. **Matrosovich M, Tuzikov A, Bovin N, Gambaryan A, Klimov A, Castrucci MR, Donatelli I, Kawaoka Y.** 2000. Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals. J Virol **74:**8502-8512.

219. **Maines TR, Chen LM, Van Hoeven N, Tumpey TM, Blixt O, Belser JA, Gustin KM, Pearce MB, Pappas C, Stevens J, Cox NJ, Paulson JC, Raman R, Sasisekharan R, Katz JM, Donis RO.** 2011. Effect of receptor binding domain mutations on receptor binding and transmissibility of avian influenza H5N1 viruses. Virology **413:**139-147.

220. **Chen LM, Blixt O, Stevens J, Lipatov AS, Davis CT, Collins BE, Cox NJ, Paulson JC, Donis RO.** 2012. In vitro evolution of H5N1 avian

influenza virus toward human-type receptor specificity. Virology **422:**105-113.

221. **Stevens J, Blixt O, Tumpey TM, Taubenberger JK, Paulson JC, Wilson IA.** 2006. Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. Science **312:**404-410.

222. **Hensley SE, Das SR, Bailey AL, Schmidt LM, Hickman HD, Jayaraman A, Viswanathan K, Raman R, Sasisekharan R, Bennink JR, Yewdell JW.** 2009. Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. Science **326:**734-736.

223. **Yewdell JW, Caton AJ, Gerhard W.** 1986. Selection of influenza A virus adsorptive mutants by growth in the presence of a mixture of monoclonal antihemagglutinin antibodies. J Virol **57:**623-628.

224. **Thoennes S, Li ZN, Lee BJ, Langley WA, Skehel JJ, Russell RJ, Steinhauer DA.** 2008. Analysis of residues near the fusion peptide in the influenza hemagglutinin structure for roles in triggering membrane fusion. Virology **370:**403-414.

225. **Reed ML, Yen HL, DuBois RM, Bridges OA, Salomon R, Webster RG, Russell CJ.** 2009. Amino acid residues in the fusion peptide pocket regulate the pH of activation of the H5N1 influenza virus hemagglutinin protein. J Virol **83:**3568-3580.

226. **Steinhauer DA, Wharton SA, Skehel JJ, Wiley DC.** 1995. Studies of the membrane fusion activities of fusion peptide mutants of influenza virus hemagglutinin. J Virol **69:**6643-6651.

227. **Gething MJ, Doms RW, York D, White J.** 1986. Studies on the mechanism of membrane fusion: site-specific mutagenesis of the hemagglutinin of influenza virus. J Cell Biol **102:**11-23.

228. **DuBois RM, Zaraket H, Reddivari M, Heath RJ, White SW, Russell CJ.** 2011. Acid stability of the hemagglutinin protein regulates H5N1 influenza virus pathogenicity. PLoS Pathog **7:**e1002398.

229. **Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Zaslavsky L, Tatusova T, Ostell J, Lipman D.** 2008. The influenza virus resource at the National Center for Biotechnology Information. J Virol **82:**596-601.

230. **Roca AI, Abajian AC, Vigerust DJ.** 2013. ProfileGrids solve the large alignment visualization problem: influenza hemagglutinin example. F1000Research doi:10.3410/f1000research.2-2.v1.

231. **Centers for Disease C, Prevention.** 2009. Outbreak of swine-origin influenza A (H1N1) virus infection - Mexico, March-April 2009. MMWR Morb Mortal Wkly Rep **58:**467-470.

232. **Anonymous.** 2009. New influenza A (H1N1) virus: global epidemiological situation, June 2009. Wkly Epidemiol Rec **84:**249-257.

233. **Hendricks GL, Weirich KL, Viswanathan K, Li J, Shriver ZH, Ashour J, Ploegh HL, Kurt-Jones EA, Fygenson DK, Finberg RW, Comolli JC,**

Wang JP. 2013. Sialylneolacto-N-tetraose c (LSTc)-bearing Liposomal Decoys Capture Influenza A Virus. J Biol Chem **288:**8061-8073.

234. **Caffrey DR, Dana PH, Mathur V, Ocano M, Hong EJ, Wang YE, Somaroo S, Caffrey BE, Potluri S, Huang ES.** 2007. PFAAT version 2.0: a tool for editing, annotating, and analyzing multiple sequence alignments. BMC Bioinformatics **8:**381.

235. **Avnir Y, Tallarico AS, Zhu Q, Bennett AS, Connelly G, Sheehan J, Sui J, Fahmy A, Huang CY, Cadwell G, Bankston LA, McGuire AT, Stamatatos L, Wagner G, Liddington RC, Marasco WA.** 2014. Molecular signatures of hemagglutinin stem-directed heterosubtypic human neutralizing antibodies against influenza A viruses. PLoS Pathog **10:**e1004103.

236. **Potter KN, Li Y, Mageed RA, Jefferis R, Capra JD.** 1999. Molecular characterization of the VH1-specific variable region determinants recognized by anti-idiotypic monoclonal antibodies G6 and G8. Scand J Immunol **50:**14-20.

237. **Pan Y, Yuhasz SC, Amzel LM.** 1995. Anti-idiotypic antibodies: biological function and structural studies. FASEB J **9:**43-49.

238. **Kipps TJ, Tomhave E, Pratt LF, Duffy S, Chen PP, Carson DA.** 1989. Developmentally restricted immunoglobulin heavy chain variable region gene expressed at high frequency in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A **86:**5913-5917.

239. **Chang DK, Kurella VB, Biswas S, Avnir Y, Sui J, Wang X, Sun J, Wang Y, Panditrao M, Peterson E, Tallarico A, Fernandes S, Goodall M, Zhu Q, Brown JR, Jefferis R, Marasco WA.** 2016. Humanized mouse G6 anti-idiotypic monoclonal antibody has therapeutic potential against IGHV1-69 germline gene-based B-CLL. MAbs doi:10.1080/19420862.2016.1159365**:**0.

240. **Kipps TJ, Robbins BA, Tefferi A, Meisenholder G, Banks PM, Carson DA.** 1990. CD5-positive B-cell malignancies frequently express cross-reactive idiotypes associated with IgM autoantibodies. Am J Pathol **136:**809-816.

241. **Martin T, Duffy SF, Carson DA, Kipps TJ.** 1992. Evidence for somatic selection of natural autoantibodies. J Exp Med **175:**983-991.

242. **Johnson TA, Rassenti LZ, Kipps TJ.** 1997. Ig VH1 genes expressed in B cell chronic lymphocytic leukemia exhibit distinctive molecular features. J Immunol **158:**235-246.

243. **Kipps TJ, Robbins BA, Carson DA.** 1990. Uniform high frequency expression of autoantibody-associated crossreactive idiotypes in the primary B cell follicles of human fetal spleen. J Exp Med **171:**189-196.

244. **Pratt LF, Szubin R, Carson DA, Kipps TJ.** 1991. Molecular characterization of a supratypic cross-reactive idiotype associated with IgM autoantibodies. J Immunol **147:**2041-2046.

245. **Charles ED, Orloff MI, Dustin LB.** 2011. A flow cytometry-based strategy to identify and express IgM from VH1-69+ clonal peripheral B cells. J Immunol Methods **363:**210-220.

246. **Carbonari M, Caprini E, Tedesco T, Mazzetta F, Tocco V, Casato M, Russo G, Fiorilli M.** 2005. Hepatitis C virus drives the unconstrained monoclonal expansion of VH1-69-expressing memory B cells in type II cryoglobulinemia: a model of infection-driven lymphomagenesis. J Immunol **174:**6532-6539.

247. **Kipps TJ, Duffy SF.** 1991. Relationship of the CD5 B cell to human tonsillar lymphocytes that express autoantibody-associated cross-reactive idiotypes. J Clin Invest **87:**2087-2096.

248. **Stanfield RL, Zemla A, Wilson IA, Rupp B.** 2006. Antibody elbow angles are influenced by their light chain class. J Mol Biol **357:**1566-1574.

249. **Prabu-Jeyabalan M, Nalivaika EA, Romano K, Schiffer CA.** 2006. Mechanism of substrate recognition by drug-resistant human immunodeficiency virus type 1 protease variants revealed by a novel structural intermediate. J Virol **80:**3607-3616.

250. **Durrant JD, de Oliveira CA, McCammon JA.** 2011. POVME: an algorithm for measuring binding-pocket volumes. J Mol Graph Model **29:**773-776.

251. **Durrant JD, Votapka L, Sorensen J, Amaro RE.** 2014. POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. J Chem Theory Comput **10:**5047-5056.

252. **Corti D, Suguitan AL, Jr., Pinna D, Silacci C, Fernandez-Rodriguez BM, Vanzetta F, Santos C, Luke CJ, Torres-Velez FJ, Temperton NJ, Weiss RA, Sallusto F, Subbarao K, Lanzavecchia A.** 2010. Heterosubtypic neutralizing antibodies are produced by individuals immunized with a seasonal influenza vaccine. J Clin Invest **120:**1663-1673.

253. **Wrammert J, Koutsonanos D, Li GM, Edupuganti S, Sui J, Morrissey M, McCausland M, Skountzou I, Hornig M, Lipkin WI, Mehta A, Razavi B, Del Rio C, Zheng NY, Lee JH, Huang M, Ali Z, Kaur K, Andrews S, Amara RR, Wang Y, Das SR, O'Donnell CD, Yewdell JW, Subbarao K, Marasco WA, Mulligan MJ, Compans R, Ahmed R, Wilson PC.** 2011. Broadly cross-reactive antibodies dominate the human B cell response against 2009 pandemic H1N1 influenza virus infection. J Exp Med **208:**181-193.

254. **Harris AK, Meyerson JR, Matsuoka Y, Kuybeda O, Moran A, Bliss D, Das SR, Yewdell JW, Sapiro G, Subbarao K, Subramaniam S.** 2013. Structure and accessibility of HA trimers on intact 2009 H1N1 pandemic influenza virus to stem region-specific neutralizing antibodies. Proc Natl Acad Sci U S A **110:**4592-4597.

255. **Lingwood D, McTamney PM, Yassine HM, Whittle JR, Guo X, Boyington JC, Wei CJ, Nabel GJ.** 2012. Structural and genetic basis for development of broadly neutralizing influenza antibodies. Nature **489:**566-570.

256. **Wang TT, Tan GS, Hai R, Pica N, Ngai L, Ekiert DC, Wilson IA, Garcia-Sastre A, Moran TM, Palese P.** 2010. Vaccination with a synthetic peptide from the influenza virus hemagglutinin provides protection against distinct viral subtypes. Proc Natl Acad Sci U S A **107:**18979-18984.

257. **Steel J, Lowen AC, Wang TT, Yondola M, Gao Q, Haye K, Garcia-Sastre A, Palese P.** 2010. Influenza virus vaccine based on the conserved hemagglutinin stalk domain. MBio **1**.

258. **Sagawa H, Ohshima A, Kato I, Okuno Y, Isegawa Y.** 1996. The immunological activity of a deletion mutant of influenza virus haemagglutinin lacking the globular region. J Gen Virol **77 (Pt 7):**1483-1487.

259. **Bommakanti G, Lu X, Citron MP, Najar TA, Heidecker GJ, ter Meulen J, Varadarajan R, Liang X.** 2012. Design of Escherichia coli-expressed stalk domain immunogens of H1N1 hemagglutinin that protect mice from lethal challenge. J Virol **86:**13434-13444.

260. **Wei CJ, Boyington JC, McTamney PM, Kong WP, Pearce MB, Xu L, Andersen H, Rao S, Tumpey TM, Yang ZY, Nabel GJ.** 2010. Induction

of broadly neutralizing H1N1 influenza antibodies by vaccination. Science **329:**1060-1064.

261. **Wei CJ, Yassine HM, McTamney PM, Gall JG, Whittle JR, Boyington JC, Nabel GJ.** 2012. Elicitation of broadly neutralizing influenza antibodies in animals with previous influenza exposure. Sci Transl Med **4:**147ra114.

262. **Ban N, Escobar C, Garcia R, Hasel K, Day J, Greenwood A, McPherson A.** 1994. Crystal structure of an idiotype-anti-idiotype Fab complex. Proc Natl Acad Sci U S A **91:**1604-1608.

263. **Ladjemi MZ.** 2012. Anti-idiotypic antibodies as cancer vaccines: achievements and future improvements. Front Oncol **2:**158.

264. **Poskitt DC, Jean-Francois MJ, Turnbull S, Macdonald L, Yasmeen D.** 1991. Internal image (Ab2 beta) anti-idiotype vaccines. Theoretical and practical aspects. Vaccine **9:**792-796.

265. **Dalgleish AG, Kennedy RC.** 1988. Anti-idiotypic antibodies as immunogens: idiotype-based vaccines. Vaccine **6:**215-220.

266. **McGuire AT, Hoot S, Dreyer AM, Lippy A, Stuart A, Cohen KW, Jardine J, Menis S, Scheid JF, West AP, Schief WR, Stamatatos L.** 2013. Engineering HIV envelope protein to activate germline B cell receptors of broadly neutralizing anti-CD4 binding site antibodies. J Exp Med **210:**655-663.

267. **Winter G, Lobley CM, Prince SM.** 2013. Decision making in xia2. Acta Crystallogr D Biol Crystallogr **69:**1260-1273.

268. **Anonymous.** 1994. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D **50:**760--763.

269. **Evans P.** 2006. Scaling and assessment of data quality. Acta Crystallographica Section D **62:**72--82.

270. **Kabsch W.** 2010. XDS. Acta Crystallographica Section D **66:**125--132.

271. **Sauter NK, Grosse-Kunstleve RW, Adams PD.** 2004. Robust indexing for automatic data collection. Journal of Applied Crystallography **37:**399--409.

272. **Zhang Z, Sauter NK, van den Bedem H, Snell G, Deacon AM.** 2006. Automated diffraction image analysis and spot searching for high-throughput crystal screening. J Appl Cryst **39:**112--119.

273. **Minor W, Cymborowski M, Otwinowski Z, Chruszcz M.** 2006. HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. Acta Crystallogr D Biol Crystallogr **62:**859-866.

274. **Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH.** 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr **66:**213-221.

275. **Bunkoczi G, Echols N, McCoy AJ, Oeffner RD, Adams PD, Read RJ.** 2013. Phaser.MRage: automated molecular replacement. Acta Crystallogr D Biol Crystallogr **69:**2276-2286.

276. **Emsley P, Lohkamp B, Scott WG, Cowtan K.** 2010. Features and development of Coot. Acta Crystallogr D Biol Crystallogr **66:**486-501.

277. **Knapp B, Frantal S, Cibena M, Schreiner W, Bauer P.** 2011. Is an intuitive convergence definition of molecular dynamics simulations solely based on the root mean square deviation possible? J Comput Biol **18:**997-1005.

278. **von Itzstein M, Wu WY, Kok GB, Pegg MS, Dyason JC, Jin B, Van Phan T, Smythe ML, White HF, Oliver SW, et al.** 1993. Rational design of potent sialidase-based inhibitors of influenza virus replication. Nature **363:**418-423.

279. **Smith BJ, Colman PM, Von Itzstein M, Danylec B, Varghese JN.** 2001. Analysis of inhibitor binding in influenza virus neuraminidase. Protein Sci **10:**689-696.

280. **California Biomedical Research Association.** 2008. Fact Sheet: New Drug Development Process, *on* California Biomedical Research Association. http://www.ca-biomed.org/pdf/media-kit/fact-sheets/CBRADrugDevelop.pdf. Accessed March 16, 2016.

281. **McKenna M.** 2013. Antibiotic resistance: the last resort. Nature **499:**394-396.

282. **Ventola CL.** 2015. The antibiotic resistance crisis: part 1: causes and threats. P T **40:**277-283.

283. **Mountain V.** 2003. Astex, Structural Genomix, and Syrrx. I can see clearly now: structural biology and drug discovery. Chem Biol **10:**95-98.

284. **Teague SJ.** 2003. Implications of protein flexibility for drug discovery. Nat Rev Drug Discov **2:**527-541.

285. **Velazquez-Campoy A, Luque I, Freire E.** 2001. The application of thermodynamic methods in drug design. Thermochimica Acta **380:**217-227.

286. **Kurt Yilmaz NS, C.A.** 2016. Improving Viral Protease Inhibitors to Counter Drug Resistance.

287. **Soumana DI, Kurt Yilmaz N, Prachanronarong KL, Aydin C, Ali A, Schiffer CA.** 2015. Structural and Thermodynamic Effects of Macrocyclization in HCV NS3/4A Inhibitor MK-5172. ACS Chem Biol doi:10.1021/acschembio.5b00647.

288. **da Silva DV, Nordholm J, Dou D, Wang H, Rossman JS, Daniels R.** 2015. The influenza virus neuraminidase protein transmembrane and head domains have coevolved. J Virol **89:**1094-1104.

289. **Kuhlbrandt W.** 2014. Cryo-EM enters a new era. Elife **3:**e03678.

290. **Impagliazzo A, Milder F, Kuipers H, Wagner MV, Zhu X, Hoffman RM, van Meersbergen R, Huizingh J, Wanningen P, Verspuij J, de Man M, Ding Z, Apetri A, Kukrer B, Sneekes-Vriese E, Tomkiewicz D, Laursen**

NS, Lee PS, Zakrzewska A, Dekking L, Tolboom J, Tettero L, van Meerten S, Yu W, Koudstaal W, Goudsmit J, Ward AB, Meijberg W, Wilson IA, Radosevic K. 2015. A stable trimeric influenza hemagglutinin stem as a broadly protective immunogen. Science **349:**1301-1306.

291. **Corti D, Voss J, Gamblin SJ, Codoni G, Macagno A, Jarrossay D, Vachieri SG, Pinna D, Minola A, Vanzetta F, Silacci C, Fernandez-Rodriguez BM, Agatic G, Bianchi S, Giacchetto-Sasselli I, Calder L, Sallusto F, Collins P, Haire LF, Temperton N, Langedijk JP, Skehel JJ, Lanzavecchia A.** 2011. A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. Science **333:**850-856.

292. **Karlsson Hedestam GB, Fouchier RA, Phogat S, Burton DR, Sodroski J, Wyatt RT.** 2008. The challenges of eliciting neutralizing antibodies to HIV-1 and to influenza virus. Nat Rev Microbiol **6:**143-155.

293. **Andrews SF, Huang Y, Kaur K, Popova LI, Ho IY, Pauli NT, Henry Dunand CJ, Taylor WM, Lim S, Huang M, Qu X, Lee JH, Salgado-Ferrer M, Krammer F, Palese P, Wrammert J, Ahmed R, Wilson PC.** 2015. Immune history profoundly affects broadly protective B cell responses to influenza. Sci Transl Med **7:**316ra192.

294. **Eggink D, Goff PH, Palese P.** 2014. Guiding the immune response against influenza virus hemagglutinin toward the conserved stalk domain by hyperglycosylation of the globular head domain. J Virol **88:**699-704.

295. **Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, Salazar-Gonzalez JF, Salazar MG, Kilby JM, Saag MS, Komarova NL, Nowak MA, Hahn BH, Kwong PD, Shaw GM.** 2003. Antibody neutralization and escape by HIV-1. Nature **422:**307-312.

296. **Zhu J, Ofek G, Yang Y, Zhang B, Louder MK, Lu G, McKee K, Pancera M, Skinner J, Zhang Z, Parks R, Eudailey J, Lloyd KE, Blinn J, Alam SM, Haynes BF, Simek M, Burton DR, Koff WC, Program NCS, Mullikin JC, Mascola JR, Shapiro L, Kwong PD.** 2013. Mining the antibodyome for HIV-1-neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains. Proc Natl Acad Sci U S A **110:**6470-6475.

297. **Karplus PA, Diederichs K.** 2012. Linking crystallographic model and data quality. Science **336:**1030-1033.

298. **Li Q, Qi J, Zhang W, Vavricka CJ, Shi Y, Wei J, Feng E, Shen J, Chen J, Liu D, He J, Yan J, Liu H, Jiang H, Teng M, Li X, Gao GF.** 2010. The 2009 pandemic H1N1 neuraminidase N1 lacks the 150-cavity in its active site. Nat Struct Mol Biol **17:**1266-1268.

299. **Vavricka CJ, Li Q, Wu Y, Qi J, Wang M, Liu Y, Gao F, Liu J, Feng E, He J, Wang J, Liu H, Jiang H, Gao GF.** 2011. Structural and Functional Analysis of Laninamivir and its Octanoate Prodrug Reveals Group Specific Mechanisms for Influenza NA Inhibition. PLoS Pathog **7:**e1002249.

300. **Barb AW, Glushka JN, Prestegard JH.** 2011. Kinetics of Neuraminidase Action on Glycoproteins by 1D and 2D NMR. J Chem Educ **88:**95-97.

301. **Vavricka CJ, Liu Y, Kiyota H, Sriwilaijaroen N, Qi J, Tanaka K, Wu Y, Li Q, Li Y, Yan J, Suzuki Y, Gao GF.** 2013. Influenza neuraminidase operates via a nucleophilic mechanism and can be targeted by covalent inhibitors. Nat Commun **4:**1491.